Failure of the L^1 pointwise ergodic theorem for $\mathrm{PSL}_2(\mathbb{R})$

Lewis Bowen[*] and Peter Burton[†]

University of Texas at Austin

August 2, 2019

Abstract

Amos Nevo established the pointwise ergodic theorem in L^p for measure-preserving actions of $\mathrm{PSL}_2(\mathbb{R})$ on probability spaces with respect to ball averages and every $p > 1$. This paper shows by explicit example that Nevo's Theorem cannot be extended to $p = 1$.

Keywords:pointwise ergodic theorem, maximal inequality

MSC:37A35

Contents

[*] supported in part by NSF grant DMS-1500389

[†] supported by an R.H. Bing Fellowship

1 Introduction

Birkhoff's ergodic theorem is that if $T : (X, \mu) \to (X, \mu)$ is a measure-preserving transformation of a standard probability space and $f \in L^1(X, \mu)$ then for a.e. $x \in X$, the time-averages $(n+1)^{-1} \sum_{i=0}^{n} f(T^i x)$ converge to the space average $\mathbb{E}[f|\mathcal{I}(T)](x)$ (this is the conditional expectation of f on the sigma-algebra of T-invariant measurable subsets). In particular, if T is ergodic then $(n+1)^{-1} \sum_{i=0}^{n} f(T^i x) \to \int f \, \mathrm{d}\mu$ for a.e. x.

To generalize this result, one can replace the single transformation T with a group G of transformations and the intervals $\{0, \ldots, n\}$ with a sequence of subsets of G or more generally, with a sequence of probability measures on G. To be precise, a sequence $\{\eta_n\}_{n=1}^{\infty}$ of probability measures on an abstract group G is **pointwise ergodic in** L^p if for every measure-preserving action $G \curvearrowright (X, \mu)$ on a standard probability space and for a.e. $x \in X$, the time-averages

$$\int f(gx) \, \mathrm{d}\eta_n(g)$$

converge to the space average $\mathbb{E}[f|\mathcal{I}(G)](x)$ as $n \to \infty$ where $\mathbb{E}[f|\mathcal{I}(G)]$ is the conditional expectation of f on the sigma-algebra of G-invariant measurable subsets. If the measure η_n is uniformly distributed over a ball then the time-averages are called ball-averages.

Pointwise ergodic theorems for amenable groups with respect to averaging over Følner sets were established in a variety of special cases culminating in Lindenstrauss' general theorem [Lin01]. This theorem also holds for L^1-functions. Nevo and co-authors established the first pointwise ergodic theorems for free groups [Nev94a, NS94] and simple Lie groups [Nev94b, Nev97, NS97, MNS00] with respect to ball and sphere averages. See also [Nev06, GN10] for surveys. These results hold in L^p for every $p > 1$. It was open problem whether ball-averages could be pointwise ergodic in L^1 for any non-amenable group.

Terrence Tao showed by explicit example that the pointwise ergodic theorem fails in L^1 for actions of free groups with respect to ball averages [Tao15]. His technique was inspired by Ornstein's counterexample demonstrating the failure of the maximal ergodic theorem in L^1 for iterates P^n of a certain well-chosen self-adjoint Markov operator [Orn69].

This note proves the analogous theorem for $\mathrm{PSL}_2(\mathbb{R})$ in place of free groups. Our approach is based on the geometry of hyperbolic surfaces. In the abstract, there is a lot in common with Tao's approach but the details of the construction are significantly different. It seems likely that our methods will generalize beyond $\mathrm{PSL}_2(\mathbb{R})$.

1.1 The main theorem

To make the result precise, we need to introduce some notation. The **hyperbolic plane** $\mathbb{H}^2$ is a complete, simply-connected Riemannian surface with constant curvature -1. It is unique up to isometry. Its orientation-preserving isometry group is isomorphic to $G := \mathrm{PSL}_2(\mathbb{R})$. Fix a base-point $p_0 \in \mathbb{H}^2$. Let $F_r \subset G$ be the set of all g such that $d_{\mathbb{H}^2}(p_0, gp_0) \le r$.

Given a probability-measure-preserving (pmp) action $G \curvearrowright (X, \mu)$, $r > 0$, a function $f \in L^1(X, \mu)$ and $x \in X$ the **ergodic average** is defined by

$$(\mathsf{A}_r f)(x) = \lambda(F_r)^{-1} \int_{F_r} f(g \cdot x) \, \mathrm{d}\lambda(g)$$

where λ is the Haar measure on G. The **terminal maximal average** is defined by $(\mathsf{M} f)(x) = \sup_{r \ge 1} (\mathsf{A}_r |f|)(x)$. Nevo proved [Nev94b]:

Theorem 1.1 (Nevo)**.** *Let $G \curvearrowright (X, \mu)$ be an ergodic pmp action, $p > 1$ and $f \in L^p(X, \mu)$. Then*

$$\lim_{r \to \infty} (\mathsf{A}_r f)(x) = \int_X f(x) \, \mathrm{d}\mu(x)$$

for μ-almost every $x \in X$.

The main theorem of this paper is that Nevo's Theorem does not extend to $p = 1$:

Theorem 1.2. *There exists an ergodic pmp action $G \curvearrowright (X, \mu)$ and a nonnegative function $f \in L^1(X, \mu)$ such that $(\mathsf{M}f)(x)$ is infinite for almost every $x \in X$. In particular, for almost every $x \in X$ the averages $(\mathsf{A}_r f)(x)$ fail to converge as $r \to \infty$.*

1.2 A rough overview of the construction

Ornstein's counterexample in [Orn69] shows that the maximal ergodic theorem fails in L^1 for powers of a certain self-adjoint operator P^n. The example consists of an L^1-function f with many components f_i, each of which comes with a "time delay" which means that $P^n f_i$ is roughly singular unless n is very large (depending on i). This allows the amplitude of f_i to be slightly smaller than would otherwise be necessary to make $\sup_n P^n f$ large on a set of significant measure.

The example here is similar in spirit although the implementation is based on the geometry of hyperbolic surfaces. The measure space is the tangent space of a hyperbolic surface. Each component function f_i is constant on a neighborhood of a cusp and the time delays are instituted by gluing surfaces together with narrow "bottlenecks".

Here is more detail. For every $\epsilon > 0$, a hyperbolic surface $S = \mathbb{H}^2/\Gamma$ (for some lattice $\Gamma < G$) and a non-negative $f \in L^\infty(S)$ are constructed to satisfy: (1) the L^1-norm of f is bounded by ϵ and (2) there is a subset $V \subset S$ with area(V)/area(S) bounded from below such that for all $x \in V$, there is some radius r so that the r-ball average of f centered at x is ≥ 1. This latter property means: if $\widetilde{x} \in \mathbb{H}^2$ is a point in the inverse image of x under the universal cover $\pi : \mathbb{H}^2 \to S$ and $\widetilde{f} = f \circ \pi$ is the lift of π then the average of $\widetilde{f}$ over the ball of radius r centered at x is at least 1. A small additional argument (which also appears in Tao's paper) finishes the proof.

These pairs (S, f) are constructed inductively. Given a pair (S, f) for some $\epsilon > 0$ (with some additional structure), a new pair $(\widehat{S}, \widehat{f})$ is constructed satisfying roughly the same maximal function lower bounds as (S, f) so that $\|\widehat{f}\|_1 \leq \|f\|_1(1 - \|f\|_1/6)$ (up to a small multiplicative error). By iterating this construction, the L^1-norm of the function can be made arbitrarily close to zero.

The new pair $(\widehat{S}, \widehat{f})$ is constructed from (S, f) as follows. We take two isometric copies of (S, f), deform them by stretching cusps into geodesics and then glue them to a pair of pants with a cusp to obtain $\widehat{S}$. The new surface has two large subsurfaces $S^{(1)}, S^{(2)}$ (each of which is isometric to a large subsurface of S) connected by a long narrow "neck" which is actually a pair of pants with a cusp. There are also two copies of f, denoted $f^{(1)}$ and $f^{(2)}$ supported on $S^{(1)}, S^{(2)}$ respectively. By choosing the neck to be very narrow, a continuity argument shows that the ball averages of each $f^{(i)}$ in $\widehat{S}$ are close to the ball averages of f in S. Theorem 1.1 shows that if $t > 0$ is chosen sufficiently large then for most p in $S^{(2)}$, the radius $(r + t)$-ball averages of $f^{(1)}$ around p are close to its space average $\int f^{(1)}\, d\nu_{\widehat{S}}$ (for every $r > 0$).

Finally, we replace $f^{(2)}$ by "flowing" it for time t into the cusps of $S^{(2)}$ and scaling it by a factor of $e^t[1 - \int f^{(1)}\, d\nu_{\widehat{S}}]$. Let f' be the new function. The radius-$(r+t)$ ball averages of f' are, up to small errors, equal to the radius-r ball averages of $f^{(2)}$ multiplied by $[1 - \int f^{(1)}\, d\nu_{\widehat{S}}]$. So let $\widehat{f} = f^{(1)} + f'$. Then we have controlled the maximal ball averages of $\widehat{f}$ on both $S^{(1)}$ and $S^{(2)}$ and the norm of $\widehat{f}$ is bounded by $\|f\|_1(1 - \|f\|_1/6)$, finishing the argument.

2 Quantitative counterexample

This section reduces Theorem 1.1 to the next lemma (which is similar to [Tao15, Theorem 2.1]).

Lemma 2.1. *There exists a constant $b > 0$ with the following property. For every $\epsilon > 0$ there exists a weakly mixing pmp action $G \curvearrowright (Y, \eta)$ and a nonnegative function $f \in L^\infty(Y, \eta)$ such that $\|f\|_1 \le \epsilon$ and $\eta(\{y \in Y : (\mathsf{M}f)(y) \ge 1\}) \ge b$.*

Proof of Theorem 1.2 from Lemma 2.1. By Lemma 2.1 for each $k \in \mathbb{N}$ there exist a weakly mixing pmp action $G \curvearrowright (Y_k, \eta_k)$ and a nonnegative function $f'_k \in L^\infty(Y_k, \eta_k)$ such that $\|f'_k\|_1 \le \left(\frac{1}{2^k}\right)^2$ and if $E_k = \{y \in Y_k : (\mathsf{M}f'_k)(y) \ge 1\}$ then $\eta_k(E_k) \ge b$.

Let $f_k = 2^k f'_k$. So $\|f_k\|_1 \le \frac{1}{2^k}$ and $E_k = \{y \in Y_k : (\mathsf{M}f_k)(y) \ge 2^k\}$. Let (X, μ) be the product measure space $(X, \mu) := \prod_{k=1}^\infty (Y_k, \eta_k)$. Because each action $G \curvearrowright (Y_k, \eta_k)$ is weakly mixing, the diagonal action $G \curvearrowright (X, \mu)$ is ergodic. Let $p_k : X \to Y_k$ be the projection onto the k^{th} coordinate and define $\widehat{f_k} = f_k \circ p_k \in L^\infty(X,\mu)$. Let $\widehat{f} = \sum_{k=1}^\infty \widehat{f_k}$. Then $\|\widehat{f_k}\|_1 = \|f_k\|_1 \leq \frac{1}{2^k}$ so that $\|\widehat{f}\|_1 \leq \sum_{n=1}^\infty \frac{1}{2^k} = 1$.

Let $\widehat{E}_k = p_k^{-1}(E_k) \subseteq X$ and, for a point $x \in X$, let $N(x) = \{k \in \mathbb{N} : x \in \widehat{E}_k\}$. Since the events $(\widehat{E}_k)_{k=1}^\infty$ are independent and $\sum_{k=1}^\infty \mu(\widehat{E}_k) = \sum_{k=1}^\infty \eta_k(E_k) = \infty$, the converse Borel-Cantelli Lemma implies that $N(x)$ is infinite for almost every $x \in X$.

Since each $\widehat{f_k}$ is non-negative,

$$(\mathsf{M}\widehat{f})(x) \geq \sup_{k\geq 1}(\mathsf{M}\widehat{f_k})(x).$$

Therefore $(\mathsf{M}\widehat{f})(x) \geq 2^k$ for every k such that $x \in \widehat{E}_k$. Since almost every x is contained in infinitely many $\widehat{E}_k$, it follows that $(\mathsf{M}\widehat{f})(x) = \infty$ for a.e. x.

□

3 Geometric preliminaries

This section reviews some standard facts needed for the next section which reduces Lemma 2.1 to a geometric problem. It will be convenient to identify the hyperbolic plane with the upper-half plane

$$\mathbb{H}^2 := \{x + iy \in \mathbb{C} : \; y > 0\}$$

equipped with the Riemannian metric $ds^2 = \frac{dx^2+dy^2}{y^2}$. The group $\mathrm{SL}_2(\mathbb{R})$ acts on $\mathbb{H}^2$ by fractional linear transformations:

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

The kernel of this action is the subgroup $\{\pm I\} \leq \mathrm{SL}_2(\mathbb{R})$. Therefore, the quotient $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm I\}$ acts on $\mathbb{H}^2$ as above. By abuse of notation, we will write elements of $\mathrm{PSL}_2(\mathbb{R})$ as matrices with the implicit understanding that the matrices are taken modulo $\{\pm I\}$.

The action $\mathrm{PSL}_2(\mathbb{R}) \curvearrowright \mathbb{H}^2$ is transitive and the stabilizer of $i \in \mathbb{H}^2$ is the subgroup of rotations

$$K = \left\{ \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} : \; \theta \in \mathbb{R} \right\}.$$

Therefore $\mathbb{H}^2$ can be identified with the quotient space $\mathrm{PSL}_2(\mathbb{R})/K$ via the map $g \cdot i \mapsto gK$.

The action $\mathrm{PSL}_2(\mathbb{R})\curvearrowright\mathbb{H}^2$ preserves the Riemannian metric. By taking derivatives, there is an induced action of $\mathrm{PSL}_2(\mathbb{R})$ on the unit tangent bundle, denoted by $T^1(\mathbb{H}^2)$. This action is simply-transitive. Therefore $\mathrm{PSL}_2(\mathbb{R})$ is the group of all orientation-preserving isometries of $\mathbb{H}^2$.

By choosing a unit vector v_0 in the tangent space of $i \in \mathbb{H}^2$, we may identify $\mathrm{PSL}_2(\mathbb{R})$ with $T^1(\mathbb{H}^2)$ via the map $g \mapsto gv_0$. Thus we have a commutative diagram:

$$\begin{tikzcd}
\mathrm{PSL}_2(\mathbb{R}) \arrow[r, leftrightarrow] \arrow[d] & T^1(\mathbb{H}^2) \arrow[d] \\
\mathrm{PSL}_2(\mathbb{R})/K \arrow[r, leftrightarrow] & \mathbb{H}^2
\end{tikzcd}$$

Moreover $\mathrm{PSL}_2(\mathbb{R})$ acts by left translations on all four spaces and these actions commute with the maps.

Suppose $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ is a discrete torsion-free subgroup. Then the quotient $\Gamma\backslash\mathbb{H}^2 \cong \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})/K$ is a hyperbolic surface. More generally, for the purposes of this paper, a **hyperbolic surface** is any Riemannian manifold isometric to a subset S of a quotient $\Gamma\backslash\mathbb{H}^2$ for some discrete torsion-free subgroup $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ such that S is equal to the closure of its interior.

By quotienting out the left-action of Γ on the four spaces above, we arrive at the following commutative diagram:

$$\begin{tikzcd}
\Gamma\backslash\mathrm{PSL}_2(\mathbb{R}) \arrow[r, leftrightarrow] \arrow[d] & \Gamma\backslash T^1(\mathbb{H}^2) \arrow[d] \\
\Gamma\backslash\mathrm{PSL}_2(\mathbb{R})/K \arrow[r, leftrightarrow] & \Gamma\backslash\mathbb{H}^2
\end{tikzcd}$$

The derivative of the covering map $\mathbb{H}^2 \to \Gamma\backslash\mathbb{H}^2$ is Γ-invariant. Therefore the unit tangent bundle of the surface $\Gamma\backslash\mathbb{H}^2$ is canonically isomorphic with the quotient space $\Gamma\backslash T^1(\mathbb{H}^2)$. Thus we have obtained an identification of $\Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ with $T^1(\Gamma\backslash\mathbb{H}^2)$.

4 Reduction to geometry

This section reduces the ergodic theory problem of Lemma 2.1 to a geometric problem. Towards that goal, suppose that $S = \Gamma\backslash\mathbb{H}^2$ is a hyperbolic surface where $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ is a discrete torsion-free subgroup. Let $\pi : \mathbb{H}^2 \to S$ denote the quotient map. For $f \in L^\infty(S)$ let

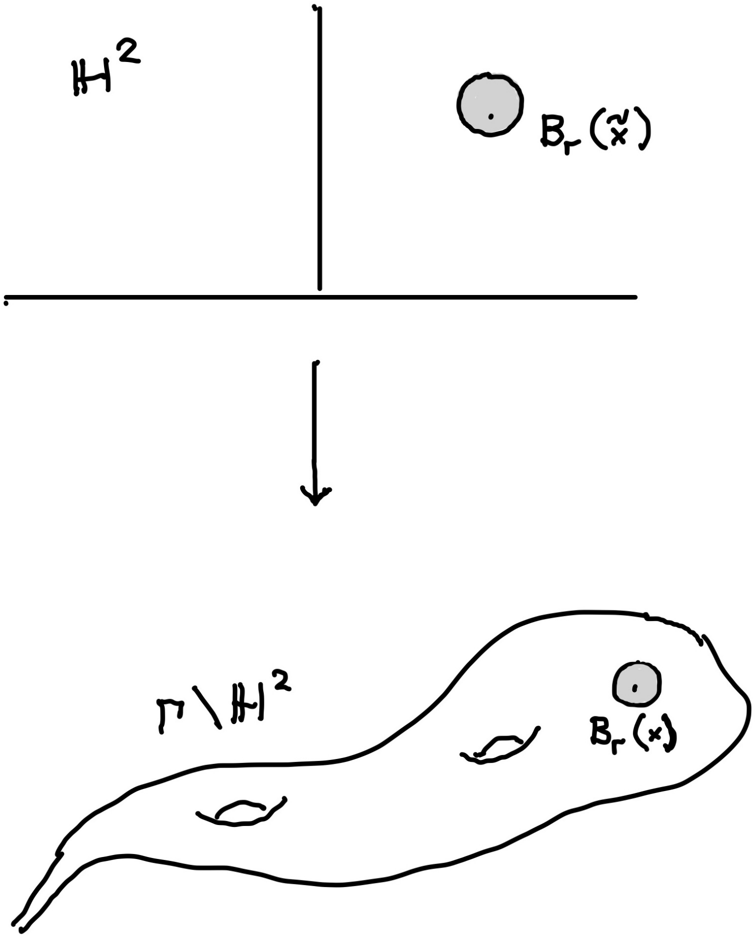

Figure 1: Geodesic balls in the hyperbolic plane and in a finite area surface

$\widetilde{f} = f \circ \pi$ be its lift to $\mathbb{H}^2$. Define the **geometric average** $\beta_r(f) \in L^\infty(S)$ by

$$(\beta_r f)(x) := \text{area}(B_r(\widetilde{x}))^{-1} \int_{B_r(\widetilde{x})} \widetilde{f}(y)\, \mathrm{d}y$$

where $\widetilde{x} \in X$ is any lift of x (so $\pi(\widetilde{x}) = x$) and $B_r(\widetilde{x})$ denotes the ball of radius r centered at $\widetilde{x}$. This does not depend on the choice of lift because π is invariant under the deck-transformation group Γ.

In the special case in which S has finite area, let ν_S denote the hyperbolic area form on S normalized so that $\nu_S(S) = 1$. Also let $\|f\|_1$ denote the $L^1(S, \nu_S)$ norm.

Lemma 4.1. *There exists a constant $b > 0$ such that for every $\epsilon > 0$ there exists a complete connected finite-area hyperbolic surface S with empty boundary and a function $f \in L^\infty(S, \nu_S)$ satisfying*

1. $f \geq 0$,
2. $\|f\|_1 \leq \epsilon$,
3. $\nu_S(\{x \in S : \sup_{r \geq 1}(\beta_r f)(x) \geq 1\}) \geq b$.

Proof of Lemma 2.1 from Lemma 4.1. The constant b is the same in both Lemmas 2.1 and 4.1. Let $\epsilon > 0$ be given and let S and f be as in Lemma 4.1. Then $S = \Gamma\backslash\mathbb{H}^2 = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})/K$ where $\Gamma \leq \mathrm{PSL}_2(\mathbb{R})$ is a torsion-free lattice. Let η_S be the probability measure on $\Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ given by integrating normalized Lebesgue measure on the unit circle K over ν_S. The right action $\mathrm{PSL}_2(\mathbb{R})$ on $\Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ preserves η_S. We take $(Y, \eta) = (\Gamma\backslash\mathrm{PSL}_2(\mathbb{R}), \eta_S)$. This action is ergodic because there is only orbit. It is weakly mixing because every ergodic action of $\mathrm{PSL}_2(\mathbb{R})$ is weakly mixing by the Howe-Moore Theorem [BM00].

If we write $q : \Gamma\backslash\mathrm{PSL}_2(\mathbb{R}) \to S = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})/K$ for the natural projection then $f \circ q$ is an element of $L^\infty(\Gamma\backslash\mathrm{PSL}_2(\mathbb{R}), \eta_S)$ and $\|f \circ q\|_1 = \|f\|_1$. Let $x \in S$ and let $\xi \in q^{-1}(x)$. Then

$$(\mathsf{A}_r(f \circ q))(\xi) = (\beta_r f)(x).$$

So the action $G \curvearrowright (Y, \eta)$ and function $f \circ q$ satisfy the conclusions of Lemma 2.1. □

5 Pants and cusps

This section introduces notation to describe pants and cusps that will be useful in the main construction.

A **right-angled hexagon** is a hexagon H in the hyperbolic plane such that all of its edges are geodesic segments and its interior angles are right angles. It will be convenient to label the sides of a hexagon by $f_0, e_{01}, f_1, e_{12}, f_2, e_{20}$ so that e_{ij} is adjacent to both f_i and f_j. See figure 5.

By [Bus92, Theorem 2.4.2], for every triple $(l_0, l_1, l_2) \in (0, \infty)^3$ there is a right-angled hexagon $H = H(l_0, l_1, l_2)$ such that the length of f_i is l_i for $i \in \{0, 1, 2\}$. Moreover, the

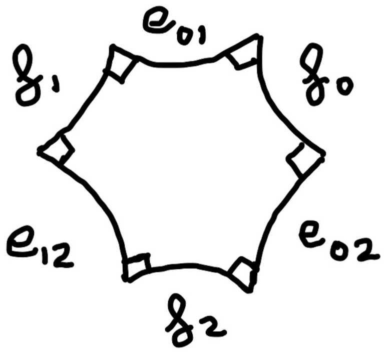

Figure 2: A right-angled hexagon

lengths of the other edges (e_{ij}) are determined by the lengths of f_0, f_1, f_2 so that H is uniquely determined up to isometry. For example, by [Bus92, Theorem 2.4.1],

$$\cosh(l_0) \quad = \quad \sinh(l_1)\sinh(l_2)\cosh(\text{length}(e_{12})) - \cosh(l_1)\cosh(l_2). \tag{1}$$

By taking limits, we can allow (l_0, l_1, l_2) to be in $[0, \infty]^3$ [Bus92, §4.4]. For example, if $(l_0, l_1, l_2) = (0, 0, 0)$ then H is an ideal triangle with its 'vertices' on the boundary at infinity. We will still refer to H as a right-angled hexagon even if some of its sides have zero or infinite length.

A **pair of pants** is a hyperbolic surface that is homeomorphic to a sphere minus three disjoint open disks such that each boundary component is a closed geodesic. For example, suppose for $k \in \{1, 2\}$, H^k is a right-angled hexagons with edges e^k_{ij}, f^k_i for $i, j \in \{0, 1, 2\}$. In addition suppose that the length of e^1_{ij} equals the length of e^2_{ij} for all i, j so that the hexagons are isometric. Let P be the surface obtained by glueing e^1_{ij} to e^2_{ij} isometrically for $i, j \in \{0, 1, 2\}$. This is a pair of pants (for details see [Bus92, §3.1] where it is called a Y-piece). The lengths of the boundary components are twice the lengths of the sides f^k_i. Conversely, if P is any pair of pants with boundary components $\partial_i P$ for $i \in \{0, 1, 2\}$ then for every pair $\{i, j\} \in \{0, 1, 2\}$ there exists a unique shortest geodesic segment γ_{ij} from $\partial_i P$ to $\partial_j P$. By cutting along these geodesic segments, we obtain two isometric right-angled hexagons (the **canonical right-angled hexagons of** P). Thus for every triple of numbers

$(l_0, l_1, l_2) \in (0, \infty)$ there exists a pair of pants P with boundary lengths equal to l_0, l_1, l_2 and P is unique up to isometry. See [Bus92, Theorem 3.1.7] for a formal proof of this statement.

A **pair of pants with k-cusps** (for $k \in \{0, 1, 2, 3\}$) is a hyperbolic surface that is homeomorphic to a sphere minus k points and $3 - k$ disjoint open disks such that each boundary component is a closed geodesic. They can be constructed exactly as in the previous paragraph by allowing the lengths of the edges f_i^k to take values in $[0, \infty)$. See [Bus92, Lemma 4.4.1] for a formal proof.

The **canonical horoball** is the subset

$$H_0 := \{x + iy \in \mathbb{C} : \ y \geq 1\} \subset \mathbb{H}^2.$$

For any $x_0 \in \mathbb{R}$, the map $z \mapsto z + x_0$ is an orientation-preserving isometry of the hyperbolic plane and therefore is represented as an element of $\mathrm{PSL}_2(\mathbb{R})$. A **cusp** is a surface isometric to a quotient of the form $C := H_0 / \{z \mapsto z + x_0\}$ for some $x_0 > 0$. For example, if P is a pair of pants with k cusps as defined above, then there really are k disjoint cusps on P [Bus92, Proposition 4.4.4].

By Gauss-Bonet, the area of a right-angled hexagon is π. So the area of a pair of pants is 2π [Bea95, p.153].

6 Deformations of surfaces

The proof of Lemma 4.1 constructs surfaces and L^1-functions inductively by cutting, pasting and deforming. The main result of this section is that the averages $\beta_r f$ vary continuously under deforming the boundary of surfaces equipped with additional structure. To make this precise, we need the following ad hoc definition.

A **panted surface** is a pair (S, P) such that S is a connected oriented hyperbolic surface and $P \subset S$ is a closed subsurface satisfying:

- P is a pair of pants with ≤ 1 cusp,
- the complement $S \setminus P$ has two connected components,
- two of the boundary components of P are contained in the interior of S. These are

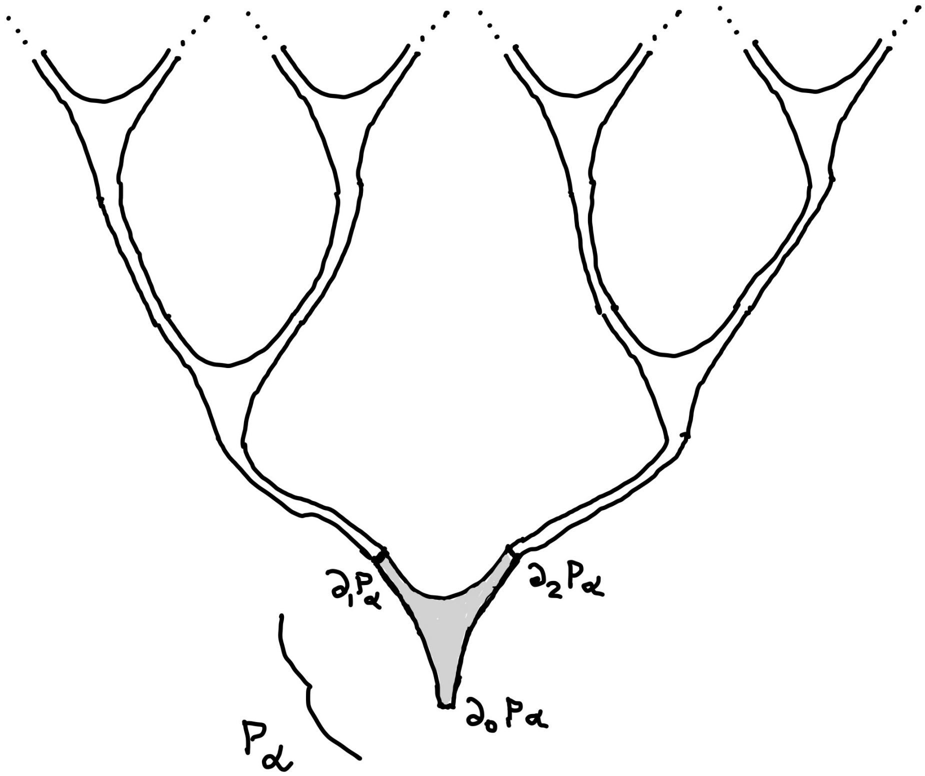

Figure 3: The surface S_α

denoted by $\partial^1 P, \partial^2 P$. If there is a third boundary component then it is denoted by $\partial^0 P$.

For $\alpha > 0$, the **α-deformation** of (S, P) is a panted surface (S_α, P_α) defined as follows. Let P_α be the (compact) oriented hyperbolic pair of pants with geodesic boundary $\partial P_\alpha = \cup_{i=0}^2 \partial^i P_\alpha$ satisfying

$$\begin{aligned} \text{length}(\partial^0 P_\alpha) &= \alpha \\ \text{length}(\partial^1 P_\alpha) &= \text{length}(\partial^1 P) \\ \text{length}(\partial^2 P_\alpha) &= \text{length}(\partial^2 P). \end{aligned}$$

This uniquely determines P_α up to orientation-preserving isometry.

Define a local isometry $\psi : \partial^1 P_\alpha \cup \partial^2 P_\alpha \to \partial^1 P \cup \partial^2 P$ as follows. There exists a unique shortest geodesic γ in P from $\partial^1 P$ to $\partial^2 P$. Let p^i be the point of intersection of γ with $\partial^i P$.

Similarly, let γ_α be the unique shortest geodesic in P_α from $\partial^1 P_\alpha$ to $\partial^2 P_\alpha$. Let p^i_α be the point of intersection of γ_α with $\partial^i P_\alpha$. Finally, let ψ be the map defined by

- for $i = 1, 2$, the restriction of ψ to $\partial^i P_\alpha$ is an isometry onto $\partial^i P$,
- $\psi(p^i_\alpha) = p^i$,
- ψ preserves orientation, where the orientation on ∂P is induced from the given orientation on P and the orientation on ∂P_α is induced from the given orientation on P_α.

This uniquely specifies ψ.

Finally, let $S_\alpha = (S \setminus \mathrm{int}(P)) \cup P_\alpha / \{x \sim \psi(x)\}$ be the surface obtained from (S minus the interior of P) and P_α by gluing together along ψ.

6.1 Continuity

This subsection studies how the averages $\beta_r f$ vary with α when f is a function on S_α. To make this precise, let $i_\alpha : S \setminus \mathrm{int}(P) \to S_\alpha$ be the inclusion map. For $f \in L^1(S \setminus \mathrm{int}(P))$, define $f_\alpha \in L^1(S_\alpha)$ by

$$f_\alpha(x) = \begin{cases} f(i_\alpha^{-1}(x)) & x \in S_\alpha \setminus \mathrm{int}(P_\alpha) \\ 0 & \text{otherwise} \end{cases}$$

Proposition 6.1. *Let (S, P) be a panted surface and $f \in L^\infty(S \setminus \mathrm{int}(P))$. For any $r > 0$, the map*

$$(x, \alpha) \mapsto \beta_r f_\alpha(i_\alpha(x))$$

is continuous as a map from $(S \setminus P) \times [0, \infty)$ to $\mathbb{C}$.

To begin, we introduce notation for describing the universal covers of the surfaces S_α and their deck-transformation groups. For $i = 1, 2$, let v^i_α be the unit tangent vector based at p^i_α, tangent to γ_α and oriented so that geodesic flow moves v^i_α immediately into γ_α.

Fix a unit tangent vector w^1 in the tangent bundle of $\mathbb{H}^2$. Because S_α is connected, there exists a unique orientation-preserving universal covering map $\pi_\alpha : X_\alpha \to S_\alpha$ such that

- $X_\alpha \subset \mathbb{H}^2$ is a closed simply-connected subset containing the base point of w^1,

- the derivative of π_α maps w^1 to v^1_α.

Let $\widetilde{\gamma}_\alpha$ be the component of $\pi_\alpha^{-1}(\gamma_\alpha)$ that contains the basepoint of w^1. Let w^2_α be the unit vector based at the other end point of $\widetilde{\gamma}_\alpha$ so that geodesic flow moves w^2_α immediately into $\widetilde{\gamma}_\alpha$. Then the derivative of π_α maps w^2_α to v^2_α. Let g_α be the unique orientation-preserving isometry of the hyperbolic plane that maps w^2_0 to w^2_α.

Let S^1_α, S^2_α be the two connected components of $S_\alpha \setminus \mathrm{int}(P_\alpha)$, indexed so that $\partial^i P_\alpha \subset S^i_\alpha$ for $i = 1, 2$. To make the notation uniform, set $w^1_\alpha = w^1$. Then let $X^i_\alpha \subset X_\alpha$ be the connected component of $\pi_\alpha^{-1}(S^i_\alpha)$ that contains the base point of w^i_α. So the restriction of π_α to X^i_α is the universal cover of S^i_α. Note that $X^1_\alpha = X^1$ and $X^2_\alpha = \gamma_\alpha X^2$ for all α.

Define the deck-transformation groups

$$\begin{aligned}\Lambda^i_\alpha &= \{g \in \mathrm{Isom}^+(\mathbb{H}^2) : \ \pi_\alpha \circ g = \pi_\alpha \text{ and } gX^i_\alpha = X^i_\alpha\}\\ \Lambda_\alpha &= \{g \in \mathrm{Isom}^+(\mathbb{H}^2) : \ \pi_\alpha \circ g = \pi_\alpha\}.\end{aligned}$$

By Van Kampen's Theorem, Λ_α is generated by Λ^1_α and Λ^2_α. Indeed, it is the free product of these subgroups. So there is a unique isomorphism $\phi_\alpha : \Lambda_0 \to \Lambda_\alpha$ defined by

$$\phi_\alpha(g) = \begin{cases} g & \text{if } g \in \Lambda^1_0 \\ g_\alpha g g_\alpha^{-1} & \text{if } g \in \Lambda^2_0 \end{cases}$$

To simplify notation, we will drop the subscripts when they equal zero. For example, $S = S_0, \Lambda = \Lambda_0$, and so on.

Lemma 6.2. *For every $\widetilde{x} \in X^1$, radius $r > 0$, $\alpha_{\max} \geq 0$ and $i \in \{1, 2\}$ there exists a finite subset $F \subset \Lambda$ such that the ball $B_r(\widetilde{x})$ has trivial intersection with $\phi_\alpha(g)X^i_\alpha$ for all $g \in \Lambda$ with $g \notin F\Lambda^i_\alpha$. In symbols,*

$$\bigcup_{i=1}^{2} \bigcup_{0 \leq \alpha \leq \alpha_{\max}} \bigcup_{g \in \Lambda \setminus F\Lambda^i} B_r(\widetilde{x}) \cap \phi_\alpha(g)X^i_\alpha = \emptyset.$$

Proof. Let $i_0 \in \{1, 2\}$, $0 \leq \alpha \leq \alpha_{\max}$ and let $\lambda : [0, r'] \to \mathbb{H}^2$ be a unit-speed geodesic from $\widetilde{x}$ to a point in $B_r(\widetilde{x}) \cap \phi_\alpha(h)X^{i_0}_\alpha$ for some $h \in \Lambda$ (and $r' \leq r$). It suffices to show there is a finite set $F \subset \Lambda$ such that $h \in F\Lambda^{i_0}$ and F does not depend on α (although it may depend on $\alpha_{\max}$ and r).

If the image $\pi_\alpha(\lambda) \subset S_\alpha$ is contained in S^1_α then $i_0 = 1$ and $h \in \Lambda^1$. So in this case, we may let $F = \{1_\Lambda\}$ and we are done.

So we assume $\pi_\alpha(\lambda)$ is not contained in S^1_α. This implies $\pi_\alpha(\lambda)$ is transverse to $\partial^1 P_\alpha \cup \partial^2 P_\alpha$. So there is a maximal discrete set $0 \le t_0 < t_1 < \cdots < t_n \le r'$ of times satisfying $\pi_\alpha(\lambda(t_i)) \in \partial^1 P_\alpha \cup \partial^2 P_\alpha$. Suppose $\pi_\alpha(\lambda(t_i)) \in \partial^j P_\alpha$ for some $j \in \{1, 2\}$. Then there exist one or two elements $g \in \Lambda$ such that

$$d_{\mathbb{H}^2}(\lambda(t_i), \phi_\alpha(g)\widetilde{p}^j_\alpha) \le \text{length}(\partial^j P)/2 \tag{2}$$

where $\widetilde{p}^j_\alpha$ is the basepoint of w^j_α. Choose an element $g_i \in \Lambda$ satisfying this inequality. Note $g_n\Lambda^{i_0} = h\Lambda^{i_0}$. So it suffices to prove: for each i with $1 \le i < n$:

1. there exists a finite set $F \subset \Lambda$ (depending only on r and $\alpha_{\max}$) such that $g_i^{-1}g_{i+1} \in F$;

2. there is a $\delta_0 > 0$ (depending only on r and $\alpha_{\max}$) such that $t_{i+1} - t_i \ge \delta_0$.

Indeed, these claims imply $g_n \in F^n$ and $n \le r/\delta_0$.

To begin, we translate the problem to a neighborhood of $\{\widetilde{p}^1_\alpha, \widetilde{p}^2_\alpha\}$ as follows. To ease notation, let $\epsilon_i \in \{1, 2\}$ be such that $\pi_\alpha(\lambda(t_i)) \in \partial^{\epsilon_i} P_\alpha$ and let

$$\ell = \max(\text{length}(\partial^1 P), \text{length}(\partial^2 P)).$$

By the triangle inequality,

$$\begin{aligned}
& d_{\mathbb{H}^2}(\widetilde{p}^{\epsilon_i}_\alpha, \phi_\alpha(g_i^{-1}g_{i+1})\widetilde{p}^{\epsilon_{i+1}}_\alpha) && (3)\\
\le\ & d_{\mathbb{H}^2}(\widetilde{p}^{\epsilon_i}_\alpha, \phi_\alpha(g_i^{-1})\lambda(t_i)) + d_{\mathbb{H}^2}(\phi_\alpha(g_i^{-1})\lambda(t_i), \phi_\alpha(g_i^{-1})\lambda(t_{i+1})) && (4)\\
& + d_{\mathbb{H}^2}(\phi_\alpha(g_i^{-1})\lambda(t_{i+1}), \phi_\alpha(g_i^{-1}g_{i+1})\widetilde{p}^{\epsilon_{i+1}}_\alpha) && (5)\\
\le\ & \ell + r && (6)
\end{aligned}$$

where the last inequality comes from two applications of (2) and the fact that $d_{\mathbb{H}^2}(\lambda(t_{i+1}), \lambda(t_i)) \le r$.

Case 1. Suppose the geodesic segment $\pi_\alpha(\lambda[t_i, t_{i+1}])$ is contained in S^j_α for some $j \in \{1, 2\}$.

In this case, there is a positive lower bound on the length $t_{i+1} - t_i$ because the surface S^j_α does not depend on α (up to isometry) and $t_{i+1} - t_i$ is at least as large as the shortest curve in S^j from $\partial^j P$ to itself that is not homotopic into the boundary.

If $\pi_\alpha(\lambda[t_i,t_{i+1}])$ is contained in $S^1_\alpha = S^1$ then (3) reduces to

$$d_{\mathbb{H}^2}(\widetilde{p}^1, g_i^{-1}g_{i+1}\widetilde{p}^1) \leq \ell + r.$$

This is because $g_i^{-1}g_{i+1} \in \Lambda^1$, ϕ_α is the identity on Λ^1 and $\widetilde{p}^1_\alpha = \widetilde{p}^1$. Since Λ^1 is discrete, there are only finitely many elements of Λ^1 that move $\widetilde{p}^1$ by distance at most $\ell + r$.

If $\pi_\alpha(\lambda[t_i,t_{i+1}])$ is contained in S^2_α then (3) reduces to

$$d_{\mathbb{H}^2}(\widetilde{p}^2, g_i^{-1}g_{i+1}\widetilde{p}^2) \leq \ell + r.$$

This is because $g_i^{-1}g_{i+1} \in \Lambda^2$, $\phi_\alpha(g_i^{-1}g_{i+1}) = g_\alpha g_i^{-1}g_{i+1}g_\alpha^{-1}$ and $\widetilde{p}^2_\alpha = g_\alpha\widetilde{p}^2$ (and the hyperbolic metric is left-invariant so we can cancel the g_α's). Since Λ^2 is discrete, there are only finitely many elements of Λ^2 that move $\widetilde{p}^2$ by distance at most $\ell + r$. This finishes Case 1.

Case 2. Suppose the geodesic segment $\pi_\alpha(\lambda[t_i,t_{i+1}])$ is contained in P_α.

Suppose $\pi_\alpha(\lambda[t_i,t_{i+1}]) = \gamma_\alpha$. Then $g_i = g_{i+1}$, so we can choose F to consist of the identity element. By equation (1) applied to either of the canonical right-angled hexagons inside P_α,

$$\begin{aligned} \cosh(\alpha) &= \sinh(\text{length}(\partial^1 P)/2)\sinh(\text{length}(\partial^2 P)/2)\cosh(\text{length}(\gamma_\alpha)) && (7)\\ &\quad - \cosh(\text{length}(\partial^1 P)/2)\cosh(\text{length}(\partial^2 P)/2). && (8)\end{aligned}$$

Since $\cosh(\alpha) \geq 1$,

$$\cosh(\text{length}(\gamma_\alpha)) \geq \frac{1+\cosh(\text{length}(\partial^1 P)/2)\cosh(\text{length}(\partial^2 P)}{\sinh(\text{length}(\partial^1 P)/2)\sinh(\text{length}(\partial^2 P)/2)} > 1.$$

So the length of γ_α admits a positive lower bound that does not depend on α. Since $\pi_\alpha(\lambda[t_i,t_{i+1}]) = \gamma_\alpha$ this implies a positive lower bound on $t_{i+1} - t_i$ that does not depend on α.

So assume $\pi_\alpha(\lambda[t_i,t_{i+1}]) \neq \gamma_\alpha$. Let e_{jk} be the shortest geodesic segment from $\partial^j P_\alpha$ to $\partial^k P_\alpha$ (for $j,k \in \{0,1,2\}$). This is well-defined even when $\alpha = 0$ by the requirement that e_{0j} meets $\partial^j P_\alpha$ in a right-angle for $j \in \{1,2\}$. Note $e_{12} = \gamma_\alpha$.

Since $\pi_\alpha(\lambda[t_i,t_{i+1}]) \neq \gamma_\alpha$, $\pi_\alpha(\lambda[t_i,t_{i+1}])$ is transverse to $\cup_{j,k} e_{jk}$. So there exists a maximal set of times $t_i < s_1 < s_2 < \ldots < s_m < t_{i+1}$ and elements $\eta_j \in \{01, 02, 12\}$ such that $\pi_\alpha(\lambda(s_j)) \in e_{\eta_j}$ for all j. Moreover, $g_i^{-1}g_{i+1}$ is determined by the sequence $\eta_1, \ldots, \eta_m$ of sides and $\epsilon_i, \epsilon_{i+1}$. So it suffices to show there are only finitely many such sequences possible. To do this, it suffices to show there is a lower bound on $s_{j+1} - s_j$ that depends only on $\alpha_{\max}$ and r (for all $1 \leq j < m$). This also implies the required lower bound on $t_{i+1} - t_i$.

Suppose $12 \in \{\eta_j, \eta_{j+1}\}$. In this case, $\pi_\alpha(\lambda[s_j, s_{j+1}])$ is a geodesic from a point in $e_{12} = \gamma_\alpha$ to a segment of the form e_{0k} for some $k \in \{1, 2\}$. But the shortest geodesic from γ_α to e_{0k} is along $\partial^k P_\alpha$ and has length equal to half the length of $\partial^k P_\alpha$. Since this length does not depend on α, it provides a positive lower bound on $s_{j+1} - s_j$ independent of α.

We may now assume $\{\eta_j, \eta_{j+1}\} = \{01, 02\}$. Let u_k be the point of intersection of $\partial^k P_\alpha$ with e_{0k} (for $k \in \{1, 2\}$). Note that $\pi_\alpha(\lambda(s_j))$ and $\pi_\alpha(\lambda(s_{j+1}))$ each have distance at most r from $\{u_1, u_2\}$.

Suppose the claim is false. By considering the canonical right-angled hexagons associated with P_α, we see that for every $\epsilon > 0$ there exist a right-angled hexagon H_ϵ bounded by sides f_k, e_{kl} $(k, l \in \{0, 1, 2\})$ and points $u'_k \in e_{0k}$ satisfying

1. $\text{length}(f_k) = \text{length}(\partial^k P)/2$ for $k \in \{1, 2\}$,
2. $\text{length}(f_0) \in [0, \alpha_{\max}]$,
3. if u_k is the vertex at the intersection of f_k and e_{0k} then $d_{\mathbb{H}^2}(u_k, u'_k) \leq r$,
4. $d_{\mathbb{H}^2}(u'_1, u'_2) \leq \epsilon$.

Here, the points u'_1, u'_2 correspond with $\pi_\alpha(\lambda(s_j))$ and $\pi_\alpha(\lambda(s_{j+1}))$. See figure 6.1.

By (1), the length of e_{12} is bounded from above and below by positive constants depending only on $\alpha_{\max}$. Thus the sides f_1, e_{12}, f_2 and points u'_1, u'_2 are all contained in a ball B whose radius is bounded in terms of $\alpha_{\max}, r$ and the constants $\text{length}(\partial^k P)$ $(k \in \{1, 2\})$. Let us consider u_1 to be fixed in the hyperbolic plane (independent of ϵ) and consider taking a subsequential limit of these hexagons as $\epsilon \searrow 0$ in the Fell topology. The limit polygon is such that its sides e_{01} and e_{02} intersect in $\mathbb{H}^2$. So it is a compact convex pentagon. However, it is not possible to obtain a compact pentagon as a limit of right-angled hexagons (even allowing that some of the sides of the right-angled hexagons have zero length). Indeed, if it was possible then it would be possible to do it with right-angled hexagons of bounded diameter such that at least one of the side-lengths tends to zero in the limit. But the formula (1) shows that for every $D > 0$ there is $\delta > 0$ such that if a right-angled hexagon H has a

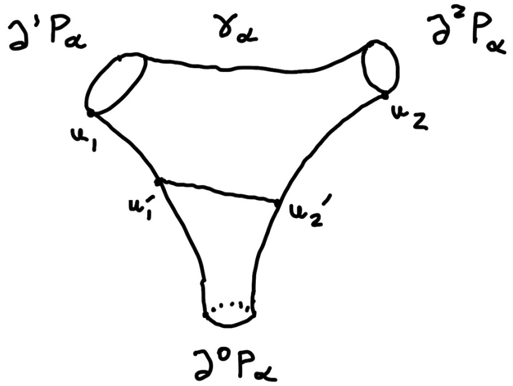

Figure 4: The pair of pants P_α.

side, say e_{12}, with length $< \delta$ then the diameter of H is $> D$. This contradiction shows that there is a positive lower bound on $s_{j+1} - s_j$ depending only on r and $\alpha_{\max}$ as required. This finishes the last case and therefore, finishes the proof.

□

Proof of Proposition 6.1. The quantity $\beta_r f_\alpha(i_\alpha(x))$ is uniformly continuous in x. Indeed, suppose for some $j \in \{1, 2\}$, $x, y \in S^j$. Let $\pi_\alpha : X_\alpha \to S_\alpha$ be the universal covering map and let $\widetilde{x}, \widetilde{y} \in X_\alpha \subset \mathbb{H}^2$ be lifts of $i_\alpha(x), i_\alpha(y)$ such that $d_{S_\alpha}(i_\alpha(x), i_\alpha(y)) = d_{\mathbb{H}^2}(\widetilde{x}, \widetilde{y})$. Then

$$\begin{aligned} |\beta_r(f_\alpha)(i_\alpha(x)) - \beta_r(f_\alpha)(i_\alpha(y))| &= \frac{1}{\text{area}(B_r(\widetilde{x}))} \left| \int_{B_r(\widetilde{x})} \widetilde{f}(z)\ dz - \int_{B_r(\widetilde{y})} \widetilde{f}(z)\ dz \right| \\ &\leq \|f\|_\infty \frac{\text{area}(B_r(\widetilde{x}) \vartriangle B_r(\widetilde{y}))}{\text{area}(B_r(\widetilde{x}))} \end{aligned}$$

where $\vartriangle$ denotes symmetric difference. Because the map i_α restricted to S^j is an isometry the distance $d_{S_\alpha}(i_\alpha(x), i_\alpha(y)) = d_S(x, y)$. Since the bound above tends to zero uniformly in the distance $d_{S_\alpha}(i_\alpha(x), i_\alpha(y))$, this proves the claim. Therefore, it suffices to prove that for any fixed $x \in S \setminus P$, the map $\alpha \mapsto \beta_r f_\alpha(i_\alpha(x))$ is continuous.

Recall

$$\beta_r f_\alpha(i_\alpha(x)) = \text{area}(B_r(\widetilde{x}))^{-1} \int_{B_r(\widetilde{x})} \widetilde{f}_\alpha(y)\, \mathrm{d}y$$

where $\widetilde{x}$ is a preimage of $i_\alpha(x)$. By symmetry, we may assume that $x \in S^1_\alpha$. Since $X^1_\alpha = X^1$ for all α, we can choose $\widetilde{x} \in X^1$ so that it does not depend on α.

Note that the preimage of $S^1_\alpha \cup S^2_\alpha$ in X_α is the disjoint union of the translates of X^1_α and X^2_α. In symbols,

$$\bigcup_{i=1}^{2} \bigcup_{g\Lambda^i \in \Lambda/\Lambda^i} \phi_\alpha(g) X^i_\alpha.$$

So

$$\int_{B_r(\widetilde{x})} \widetilde{f}_\alpha(y)\,\mathrm{d}y = \sum_{i=1}^{2} \sum_{g\Lambda^i \in \Lambda/\Lambda^i} \int_{B_r(\widetilde{x}) \cap \phi_\alpha(g) X^i_\alpha} \widetilde{f}_\alpha(y)\,\mathrm{d}y.$$

By Lemma 6.2 there are finite sets $F^1, F^2 \subset \Lambda$ (depending only on an upper bound for α and r) such that

$$\int_{B_r(\widetilde{x})} \widetilde{f}_\alpha(y)\,\mathrm{d}y = \sum_{i=1}^{2} \sum_{g \in F^i} \int_{B_r(\widetilde{x}) \cap \phi_\alpha(g) X^i_\alpha} \widetilde{f}_\alpha(y)\,\mathrm{d}y. \tag{9}$$

The integrals can be rewritten as follows:

$$\int_{B_r(\widetilde{x}) \cap \phi_\alpha(g) X^i_\alpha} \widetilde{f}_\alpha(y)\,\mathrm{d}y = \int_{\phi_\alpha(g^{-1}) B_r(\widetilde{x}) \cap X^i_\alpha} \widetilde{f}_\alpha(\phi_\alpha(g) y)\,\mathrm{d}y = \int_{\phi_\alpha(g^{-1}) B_r(\widetilde{x}) \cap X^i_\alpha} \widetilde{f}_\alpha(y)\,\mathrm{d}y \tag{10}$$

where the first equality follows from the change of variables $y \mapsto \phi_\alpha(g) y$ and the second from the Λ_α-invariance of $\widetilde{f}_\alpha$. If $i = 1$ then $\widetilde{f}_\alpha(y) = \widetilde{f}(y)$ for all $y \in X^1_\alpha = X^1$. So

$$\int_{B_r(\widetilde{x}) \cap \phi_\alpha(g) X^1_\alpha} \widetilde{f}_\alpha(y)\,\mathrm{d}y = \int_{\phi_\alpha(g^{-1}) B_r(\widetilde{x}) \cap X^1} \widetilde{f}(y)\,\mathrm{d}y.$$

If $i = 2$ then $\widetilde{f}_\alpha(g_\alpha y) = \widetilde{f}(y)$ for $y \in X^2$ (and $g_\alpha X^2 = X^2_\alpha$). By a change of variables

$$\int_{\phi_\alpha(g^{-1}) B_r(\widetilde{x}) \cap X^2_\alpha} \widetilde{f}_\alpha(y)\,\mathrm{d}y \quad = \quad \int_{\phi_\alpha(g^{-1}) B_r(\widetilde{x}) \cap X^2_\alpha} \widetilde{f}(g_\alpha^{-1} y) = \int_{g_\alpha^{-1} \phi_\alpha(g^{-1}) B_r(\widetilde{x}) \cap X^2} \widetilde{f}(y)\,\mathrm{d}y.$$

Combined with (9) and (10) this implies

$$\mathrm{area}(B_r(\widetilde{x}))\beta_r f_\alpha(i_\alpha(x)) \quad = \quad \sum_{g \in F^1} \int_{\phi_\alpha(g^{-1}) B_r(\widetilde{x}) \cap X^1} \widetilde{f}(y)\,\mathrm{d}y, + \sum_{g \in F^2} \int_{g_\alpha^{-1} \phi_\alpha(g^{-1}) B_r(\widetilde{x}) \cap X^2} \widetilde{f}(y)\,\mathrm{d}y.$$

Observe that each of the integrals above is continuous in α because $\alpha \mapsto g_\alpha$ and $\alpha \mapsto \phi_\alpha(g)$ are continuous (for fixed g). So we have expressed $\beta_r f_\alpha(i_\alpha(x))$ as a finite sum of functions that are continuous in α. Thus $\beta_r \widetilde{f}_\alpha(i_\alpha(x))$ is continuous in α.

□

7 Averaging around cusps

The main result of this section is a comparison between the averages of the form $\beta_r(f)$ and $\beta_r(f1_C)$ where C is a cusp of the surface. This is used in the proof of Lemma 4.1 to control the maximal function under these kinds of deformations of functions. To be precise, we need the following definitions.

Let $C = H_0/\{z \mapsto z + x_0\}$ be a cusp where $H_0 = \{x + iy \in \mathbb{H}^2 : \ y \geq 1\}$ is the canonical horoball and $x_0 > 0$ is the length of the boundary of C (which is a horocycle). For $t > 0$, let

$$C[t] = \{x + iy \in \mathbb{H}^2 : \ y \geq e^t\}/\{z \mapsto z + x_0\} \subset C.$$

This is the unique cusp contained in C such that the distance between the boundaries ∂C and $\partial C[t]$ is t.

Proposition 7.1. *Let S be a hyperbolic surface with pairwise disjoint cusps $C_1, \ldots, C_k \subset S$. Let $U = \cup_{i=1}^k C_i$ be the union of the cusps and $U[t] = \cup_{i=1}^k C_i[t]$ the union of the shortened cusps for $t \geq 0$. Let $f \in L^\infty(S)$ be a non-negative function such that (1) f is constant on C_i for all i and (2) $f(p) = 0$ for all $p \in S \setminus U$. Then for all $p \in S \setminus U$ and $t, r \geq 0$,*

$$\beta_{r+t}(f1_{U[t]})(p) \geq e^{-t}(1 - 2e^{-r})\beta_r(f)(p).$$

Proof. Because β_r is linear, it suffices to consider the special case in which $f(p) = 1$ for all $p \in U$. By passing to the universal cover, it suffices to prove: for any $p \in \mathbb{H}^2 \setminus H_0$,

$$\frac{\text{area}(B(r+t,p) \cap \{x + iy : \ y \geq e^t\})}{\text{area}(B(r+t,p))} \geq e^{-t}(1 - 2e^{-r})\frac{\text{area}(B(r,p) \cap H_0)}{\text{area}(B(r,p))}.$$

Before estimating the above, here are some general facts about areas of intersections of balls and horoballs.

For $R > T > 0$, let $g(R,T)$ be the area of the intersection of a ball B and a horoball H such that the radius of B is R and the distance between the center of B and the boundary of H is T. Then $g(R,T)$ is well-defined (in that it depends on the choice of B and H only through R and T) and for any fixed t_0, $g(T + t_0, T)$ is monotone increasing in T. To see this, we may assume $H = H_0$ and $t_0 > 0$ (since if $t_0 \leq 0$ then $g(T + t_0, T) = 0$). Set B_T equal to the ball of hyperbolic radius $T + t_0$ and hyperbolic center $e^{-T}i$ in the upper half-plane model $\mathbb{H}^2$. Recall that the hyperbolic distance between two points on the imaginary axis is the absolute difference between their logarithms (so $d_{\mathbb{H}^2}(e^a i, e^b i) = |a-b|$). So $g(T+t_0, T) = \text{area}(H_0 \cap B_T)$. Also B_T coincides with the Euclidean disk centered on the imaginary axis that contains $e^{t_0}i$ and $e^{-2T-t_0}i$ in its boundary. In particular, $B_T \subset B_{T'}$ for any $T \leq T'$. So $g(T+t_0,T) \leq g(T'+t_0,T')$.

It follows that

$$\text{area}\left(B(r+t,p) \cap \{x+iy:\ y \geq e^t\}\right) = g(r+t, d_{\mathbb{H}^2}(p,H_0)+t) \geq g(r, d_{\mathbb{H}^2}(p,H_0)) = \text{area}(B(r,p) \cap H_0).$$

So it suffices to show

$$\frac{\text{area}(B(r,p))}{\text{area}(B(r+t,p))} \geq e^{-t}(1-2e^{-r}).$$

Since $\text{area}(B(r,p)) = 2\pi(\cosh(r)-1)$,

$$\begin{aligned}\frac{\text{area}(B(r,p))}{\text{area}(B(r+t,p))} &= \frac{\cosh(r)-1}{\cosh(r+t)-1} = \frac{e^r - 2 + e^{-r}}{e^{t+r} - 2 + e^{-t-r}} \\ &\geq \frac{e^r - 2}{e^{t+r}} = e^{-t}(1-2e^{-r}).\end{aligned}$$

□

8 The inductive step

To prove Lemma 4.1, we will construct surfaces S with functions $f \in L^1(S)$ by induction. To be precise, we need the next two definitions.

Definition 1. A tuple $\left(S, P, \{C_i\}_{i=1}^k, U, f\right)$ is **good** if

1. (S,P) is a panted surface,

2. S is a complete hyperbolic surface with finite area and no boundary,

3. $C_1, \ldots, C_k \subset S$ are pairwise disjoint cusps,

4. P is disjoint from $U = \cup_i C_i$,

5. $f \in L^1(S)$ is non-negative,

6. f is constant on each cusp C_i,

7. $f(p) = 0$ for all $p \in S \setminus U$,

8. $\|f\|_1 \leq 2$.

Definition 2. For $\rho \geq 0$ and $f \in L^1(S)$, let

$$\mathsf{M}_\rho f(p) = \sup_{\rho \leq r} \beta_r(|f|)(p)$$

be the **ρ-truncated maximal function** of f.

The next result forms the inductive step in the proof of Lemma 4.1.

Proposition 8.1. *Let $\left(S, P, \{C_i\}_{i=1}^k, U, f\right)$ be a good tuple and let ρ, ϵ be parameters such that $10 \leq \rho$ and $0 < \epsilon < 1/10$. Let*

$$V = \{p \in S \setminus (P \cup U) : \ \mathsf{M}_\rho f(p) \geq 1\}.$$

Then there exists a good tuple $\left(\widehat{S}, \widehat{P}, \{\widehat{C}_j\}_{j=1}^{2k}, \widehat{U}, \widehat{f}\right)$ satisfying

1. $\operatorname{area}(\widehat{S}) = 2\operatorname{area}(S) + 2\pi$*,*

2. if

$$\widehat{V} = \left\{p \in \widehat{S} \setminus (\widehat{P} \cup \widehat{U}) : \ \mathsf{M}_\rho \widehat{f}(p) \geq 1\right\}$$

then $\operatorname{area}(\widehat{V}) \geq 2\operatorname{area}(V) - 3\epsilon$*,*

3. $\|\widehat{f}\|_1 \leq \dfrac{\|f\|_1(1 - \|f\|_1/6)}{1 - 4\epsilon - 4e^{-\rho}}$.

Proof. By definition of V, there exist $R > 0$ and a compact subset $W \subset V$ such that $\operatorname{area}(W) \geq \operatorname{area}(V) - \epsilon$ and

$$\sup_{\rho \leq r \leq R} \beta_r(f)(p) \geq 1 - \epsilon$$

for all $p \in W$.

By Proposition 6.1, there exists $\alpha > 0$ such that if S_α and f_α are defined as in §6.1 then

$$\sup_{\rho \leq r \leq R} \beta_r(f_\alpha)(p) \geq 1 - 2\epsilon$$

for all $p \in W$. Here we are identifying W with a subset of S_α. This makes sense because $S \setminus P$ is naturally isometric to $S_\alpha \setminus P_\alpha$ and $W \subset V \subset S \setminus P$.

Let $S^{(1)}, S^{(2)}$ be two isometric copies of S_α. For $i = 1, 2$ and $1 \leq j \leq k$, let $C_j^{(i)} \subset S^{(i)}$ be the copy of the cusp C_j in $S^{(i)}$ and let $f^{(i)} \in L^1(S^{(i)})$ be a copy of f_α. Define $V^{(i)}, U^{(i)}, W^{(i)} \subset S^{(i)}$ similarly.

The surface S_α has a single boundary component which is of length α. Let Y_α be the pair of pants with one cusp and two geodesic boundary components $\partial^1 Y_\alpha$ and $\partial^2 Y_\alpha$, both of length α. For $i = 1, 2$, let $\psi^{(i)} : \partial^i Y_\alpha \to \partial S^{(1)}$ be an isometry and let $\psi : \partial Y_\alpha \to \partial(S^{(1)} \sqcup S^{(2)})$ be the union of these two maps. Finally, let

$$\widehat{S} = \left(S^{(1)} \sqcup S^{(2)} \sqcup Y_\alpha\right) / \{x \sim \psi(x)\}$$

be the result of gluing Y_α to $S^{(1)} \sqcup S^{(2)}$ via ψ. Let $\widehat{P}$ be the copy of Y_α in $\widehat{S}$. Conclusion (1) is immediate.

Extend $f^{(i)}$ to all of $\widehat{S}$ by setting $f^{(i)}(p) = 0$ for all $p \in \widehat{S} \setminus S^{(i)}$. By Nevo's Pointwise Ergodic Theorem (Theorem 1.1) applied to $f^{(1)}$, there exists $t > 0$ and $W' \subset W^{(2)}$ such that $\text{area}(W') \geq \text{area}(W^{(2)}) - \epsilon$ and for all $p \in W'$ and $r \geq t$,

$$\beta_r\left(f^{(1)}\right)(p) \geq -\epsilon + \int f^{(1)}\, d\nu_{\widehat{S}}.$$

Define cusps

$$\widehat{C}_j := C_j^{(1)}, \quad \widehat{C}_{k+j} := C_j^{(2)}[t]$$

for $1 \leq j \leq k$.

Define $\bar{f} \in L^1(\widehat{S})$ by

$$\bar{f} = f^{(1)} + \left[1 - \int f^{(1)}\, d\nu_{\widehat{S}}\right] e^t 1_{U^{(2)}[t]} f^{(2)}$$

where $U^{(2)}[t] = \cup_{j=1}^k C_j^{(2)}[t]$ is as defined in §7.

Because $\|f\|_1 \leq 2$ (by definition of a good tuple), it follows that

$$1 - \int f^{(1)}\, d\nu_{\widehat{S}} = 1 - \frac{\text{area}(S)}{\text{area}(\widehat{S})} \int f\, d\nu_S > 0.$$

So both summands defining $\bar{f}$ are non-negative. In particular, $\bar{f} \geq 0$.

Set

$$\widehat{f} := \frac{\bar{f}}{1 - 4\epsilon - 4e^{-\rho}}.$$

It is immediate that $\left(\widehat{S}, \widehat{P}, \{\widehat{C}_j\}_{j=1}^{2k}, \widehat{U}, \widehat{f}\right)$ is a good tuple.

The next step is to verify the maximal function estimates. We claim that if $p \in W^{(1)} \cup W'$ then $\mathsf{M}_\rho \widehat{f}(p) \geq 1$. So suppose $p \in W^{(1)}$. Then the definition of W implies

$$\mathsf{M}_\rho \bar{f}(p) \geq \mathsf{M}_\rho f^{(1)}(p) \geq 1 - 2\epsilon.$$

Therefore

$$\mathsf{M}_\rho \widehat{f}(p) \geq \frac{1-2\epsilon}{1-4\epsilon-4e^{-\rho}} \geq 1. \tag{11}$$

If $p \in W' \subset W^{(2)}$, then there exists $r \geq \rho$ such that

$$\beta_r\left(f^{(2)}\right)(p) \geq 1 - \epsilon.$$

By Proposition 7.1,

$$\beta_{r+t}\left(1_{U^{(2)}[t]} f^{(2)}\right)(p) \geq e^{-t}(1-2e^{-r})\beta_r\left(f^{(2)}\right)(p) \geq e^{-t}(1-2e^{-r})(1-\epsilon).$$

Therefore,

$$\begin{aligned}
\mathsf{M}_\rho \bar{f}(p) &\geq \beta_{r+t}(\bar{f})(p) \geq \beta_{r+t}\left(f^{(1)}\right)(p) + \left[1 - \int f^{(1)}\, d\nu_{\widehat{S}}\right] e^t \beta_{r+t}\left(1_{U^{(2)}[t]} f^{(2)}\right)(p) \\
&\geq -\epsilon + \int f^{(1)}\, d\nu_{\widehat{S}} + \left[1 - \int f^{(1)}\, d\nu_{\widehat{S}}\right](1-2e^{-r})(1-\epsilon) \\
&= -\epsilon + (1-2e^{-r})(1-\epsilon) + \left(\int f^{(1)}\, d\nu_{\widehat{S}}\right)\left[1-(1-2e^{-r})(1-\epsilon)\right] \\
&\geq 1 - 3\epsilon - 4e^{-r} \geq 1 - 4\epsilon - 4e^{-\rho}
\end{aligned}$$

where the lower bound on $\beta_{r+t}\left(f^{(1)}\right)(p)$ follows from the definition of W'. Therefore, $\mathsf{M}_\rho \widehat{f}(p) \geq 1$. Together with inequality (11) this implies $\mathsf{M}_\rho \widehat{f}(p) \geq 1$ for all $p \in W^{(1)} \cup W'$. So $\widehat{V} \supset W^{(1)} \cup W'$ which implies

$$\operatorname{area}(\widehat{V}) \geq 2\operatorname{area}(V) - 3\epsilon.$$

This verifies conclusion (2).

Next, we verify conclusion (3). Recall that our normalization conventions imply $\operatorname{area}(\widehat{S})\|f^{(1)}\|_1 = \operatorname{area}(S)\|f\|_1$ (for example). Because $\operatorname{area}(C[t]) = e^{-t}\operatorname{area}(C)$ for any cusp C,

$$\operatorname{area}(\widehat{S})\left\|1_{U^{(2)}[t]} f^{(2)}\right\|_1 = \operatorname{area}(S) e^{-t}\|f\|_1.$$

So

$$\begin{aligned} \mathrm{area}(\widehat{S})\|\bar{f}\|_1 &= \mathrm{area}(\widehat{S})\|f^{(1)}\|_1 + \mathrm{area}(\widehat{S})\left[1 - \int f^{(1)}\, d\nu_{\widehat{S}}\right] e^t \left\|1_{U^{(2)}[t]} f^{(2)}\right\|_1 \\ &= \mathrm{area}(S)\|f\|_1 + \mathrm{area}(S)\left[1 - \int f^{(1)}\, d\nu_{\widehat{S}}\right] \|f\|_1 \\ &= \mathrm{area}(S)\|f\|_1 \left(2 - \frac{\mathrm{area}(S)}{\mathrm{area}(\widehat{S})}\|f\|_1\right) \le \mathrm{area}(S)\|f\|_1\,(2 - \|f\|_1/3) \end{aligned}$$

where the last inequality comes from the fact that $\mathrm{area}(\widehat{S}) = 2\mathrm{area}(S) + 2\pi$ and since $\widehat{S}$ contains a pair of pants, $\mathrm{area}(\widehat{S}) \ge 2\pi$. Therefore, $\frac{\mathrm{area}(S)}{\mathrm{area}(\widehat{S})} \ge 1/3$.

Divide both sides by $\mathrm{area}(\widehat{S})$ and use the estimate $\mathrm{area}(S)/\mathrm{area}(\widehat{S}) \le 1/2$ to obtain

$$\|\bar{f}\|_1 \le \|f\|_1(1 - \|f\|_1/6)$$

which implies conclusion (3).

□

9 The end of the proof

The next lemma establishes the base case of the induction in the proof of Lemma 4.1.

Lemma 9.1. *For every $\rho \ge 0$, there exists a good tuple $(S, P, \{C_i\}_{i=1}^4, U, f)$ such that*

$$\nu_S\left(\{p \in S \setminus (P \cup U) :\ \mathsf{M}_\rho f(p) \ge 1\}\right) \ge 1/2.$$

Proof. Let $\alpha > 0$ and let Y_1 be a pair of pants with two cusps and one geodesic boundary component of length $\alpha > 0$. Let Y_2 be an isometric copy of Y_1. Let P be a pair of pants with one cusp and two geodesic boundary components each of length α. Let $\psi : \partial P \to \partial Y_1 \sqcup \partial Y_2$ be an isometry and let

$$S = [Y_1 \sqcup Y_2 \sqcup P]/\{x \sim \psi(x)\}$$

be the surface obtained by gluing Y_1, Y_2 and P together by way of ψ. Then (S, P) is a panted surface with area 6π.

For $i = 1, 2$, let $V_i \subset Y_i$ be a compact subsurface with

$$\mathrm{area}(V_i) \ge 3\,\mathrm{area}(Y_i)/4 = 3\pi/2.$$

Let $C_1^{(i)}, C_2^{(i)} \subset Y_i$ be disjoint cusps such that for any $p \in V_i$ and $q \in C_1^{(i)} \cup C_2^{(i)}$, $d_S(p,q) \geq \rho$. Let $f \in L^1(S)$ be any non-negative function such that $(S, P, \{C_i\}_{i=1}^4, U, f)$ is a good tuple and $\|f\|_1 = 1$. For example, one could define f by

$$f(p) = \begin{cases} \dfrac{\operatorname{area}(S)}{4 \operatorname{area}\left(C_j^{(i)}\right)} & p \in C_j^{(i)} \\ 0 & \text{otherwise} \end{cases}$$

By Nevo's Pointwise Ergodic Theorem 1.1, for a.e. $p \in S$, $\mathsf{M}f(p) \geq 1$. Since $\beta_r f(p) = 0$ for all $r < \rho$ and $p \in V_1 \cup V_2$, it follows that $\mathsf{M}_\rho f(p) \geq 1$ for all $V_1 \cup V_2$. Since

$$\operatorname{area}(V_1 \cup V_2) \geq 3\pi = \operatorname{area}(S)/2$$

this finishes the proof.

□

Lemma 9.2. *Let $t_1, t_2, \ldots$ be a sequence of real numbers $t_i \in [0,2)$ such that $t_{i+1} \leq t_i(1 - t_i/6)$ for all i. Then $\lim_{i\to\infty} t_i = 0$.*

Proof. Since $1 - t_i/6 < 1$, the sequence is monotone decreasing. So the limit exists $L = \lim_{i\to\infty} t_i$ exists, $L \in [0,2)$ and $L = L(1 - L/6)$. This implies $L = 0$. □

Proof of Lemma 4.1. For $b, \rho > 0$, let $\Sigma(b,\rho)$ be the set of all numbers $\delta > 0$ such that there exists a good tuple $\left(S, P, \{C_i\}_{i=1}^k, U, f\right)$ satisfying

1. $f \geq 0$,
2. $\|f\|_1 \leq \delta$,
3. $\nu_S\left(\{p \in S \setminus (P \cup U) : \mathsf{M}_\rho f(p) \geq 1\}\right) \geq b$.

Also let $\overline{\Sigma(b,\rho)}$ denote the closure of $\Sigma(b,\rho)$ in $[0,\infty)$. It suffices to prove that $0 \in \overline{\Sigma(b,10)}$ for some $b > 0$.

Note that if $b' \leq b$ and $\rho' \geq \rho$ then $\Sigma(b,\rho) \subset \Sigma(b',\rho')$. Lemma 9.1 proves that $1 \in \Sigma(1/2,\rho)$ for all ρ. Proposition 8.1 proves: if $\delta \in \Sigma(b,\rho)$ for all $\rho \geq 10$ then $\delta(1-\delta/6) \in \overline{\Sigma(b-\epsilon,\rho)}$ for all $\epsilon > 0$ and $\rho \geq 10$. By iterating and using Lemma 9.2, this implies $0 \in \overline{\Sigma(1/2-\epsilon,\rho)}$ for all $\epsilon > 0$ and $\rho \geq 10$ which finishes the lemma.

□

10 Two open problems

The main counterexample does not have spectral gap. This is because we are forced to make the "necks" in the construction of the surface arbitrarily narrow. Similarly, Tao's construction does not have spectral gap. This raises a question: does Nevo's Pointwise Ergodic Theorem 1.1 hold in L^1 if $G \curvearrowright (X, \mu)$ has spectral gap? It also raises the converse question: if $G \curvearrowright (X, \mu)$ is ergodic but does not have spectral gap then does the Pointwise Ergodic Theorem necessarily fail in L^1 for this action?

References

[Bea95] Alan F. Beardon. *The geometry of discrete groups*, volume 91 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995. Corrected reprint of the 1983 original.

[BM00] M. Bachir Bekka and Matthias Mayer. *Ergodic theory and topological dynamics of group actions on homogeneous spaces*, volume 269 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 2000.

[Bus92] Peter Buser. *Geometry and spectra of compact Riemann surfaces*, volume 106 of *Progress in Mathematics*. Birkhäuser Boston, Inc., Boston, MA, 1992.

[GN10] Alexander Gorodnik and Amos Nevo. *The ergodic theory of lattice subgroups*, volume 172 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2010.

[Lin01] Elon Lindenstrauss. Pointwise theorems for amenable groups. *Invent. Math.*, 146(2):259–295, 2001.

[MNS00] G. A. Margulis, A. Nevo, and E. M. Stein. Analogs of Wiener's ergodic theorems for semisimple Lie groups. II. *Duke Math. J.*, 103(2):233–259, 2000.

[Nev94a] Amos Nevo. Harmonic analysis and pointwise ergodic theorems for noncommuting transformations. *J. Amer. Math. Soc.*, 7(4):875–902, 1994.

[Nev94b] Amos Nevo. Pointwise ergodic theorems for radial averages on simple Lie groups. I. *Duke Math. J.*, 76(1):113–140, 1994.

[Nev97] Amos Nevo. Pointwise ergodic theorems for radial averages on simple Lie groups. II. *Duke Math. J.*, 86(2):239–259, 1997.

[Nev06] Amos Nevo. Pointwise ergodic theorems for actions of groups. In *Handbook of dynamical systems. Vol. 1B*, pages 871–982. Elsevier B. V., Amsterdam, 2006.

[NS94] Amos Nevo and Elias M. Stein. A generalization of Birkhoff's pointwise ergodic theorem. *Acta Math.*, 173(1):135–154, 1994.

[NS97] Amos Nevo and Elias M. Stein. Analogs of Wiener's ergodic theorems for semisimple groups. I. *Ann. of Math. (2)*, 145(3):565–595, 1997.

[Orn69] Donald Ornstein. On the pointwise behavior of iterates of a self-adjoint operator. *J. Math. Mech.*, 18:473–477, 1968/1969.

[Tao15] Terence Tao. Failure of the L^1 pointwise and maximal ergodic theorems for the free group. *Forum Math. Sigma*, 3:e27, 19, 2015.