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

Lewis Bowen* and Peter Burton

University of Texas at Austin

January 4, 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

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 $\{0, \dots, n\}$ 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 [**?**]. This theorem also holds for L^1-functions. Nevo and co-authors established the first pointwise ergodic theorems for free groups [**?**, **?**] and simple Lie groups [**?**, **?**, **?**, **?**] with respect to ball and sphere averages. See also [**?**, **?**] 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 [**?**]. This note shows 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 $G := \mathrm{PSL}_2(\mathbb{R})$. Moreover, G acts on the unit-tangent bundle $T^1(\mathbb{H}^2)$ simply transitively. Fix a base-point $p_0 \in \mathbb{H}^2$. Let $F_r \subset G$ be the set of all g such that $d(g, gp_0) \le r$.

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 [**?**]:

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

The heart of the proof is geometric. 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 $\text{area}(V)/\text{area}(S)$ bounded from below such that for all $x \in V$, there is some radius r so that 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 bounds as (S, f) so that $\|\widehat{f}\|_1 \leq \|f\|_1(1 - \|f\|_1/2)$. 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 $\|f\|_1(1 - \|f\|_1/2)$, finishing the argument.

2 Quantitative counterexample

This section reduces Theorem 1.1 to the next lemma (which is similar to [**?**, 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 $\|f\|_1 \leq \epsilon$ and $\eta(\{y \in Y : (\mathsf{M}f)(y) \geq 1\}) \geq b$.*

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 $\|f_k\|_1 \leq \frac{1}{2^k}$ and if $E_k = \{y \in Y_k : (\mathsf{M}f_k)(y) \geq 1\}$ then $\eta_k(E_k) \geq b$. 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, sub-additivity of the maximal operator implies

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

Therefore $(\mathsf{M}\widehat{f})(x) \geq N(x)$. By the previous paragraph, this means that $(\mathsf{M}\widehat{f})(x) = \infty$ for a.e. x.

□

3 Reduction to geometry

Throughout this paper, a **hyperbolic surface** is a complete Riemann surface (possibly with non-empty boundary) with constant curvature -1. This section reduces the ergodic theory problem of Lemma 2.1 to a geometric problem. Towards that goal, suppose that S is a connected hyperbolic surface such that there exists a locally isometric covering map

$\pi : X \to S$ where $X \subset \mathbb{H}^2$ is a simply-connected subspace. For $f \in L^\infty(S)$ let $\widetilde{f}$ be its lift to $\mathbb{H}^2$ defined by

$$\widetilde{f}(x) = \begin{cases} f(\pi(x)) & x \in X \\ 0 & \text{otherwise} \end{cases}$$

Define the **geometric average** $\beta_r(f) \in L^\infty(S)$ by

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

where $\widetilde{x} \in X$ is any lift of x (so $\pi(\widetilde{x}) = x$). This does not depend on the choice of lift because π is invariant under the deck-transformation group.

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 3.1. *There exists a constant $b > 0$ such that for every $\epsilon > 0$ there exists a complete 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 3.1. The constant b is the same in both Lemmas 2.1 and 3.1. Let $\epsilon > 0$ be given and let S and f be as in Lemma 3.1. Let $\mathrm{T}^1 S$ be the unit tangent bundle of S and let η_S be the probability measure on $\mathrm{T}^1 S$ given by integrating normalized Lebesgue measure on the unit circle over ν_S. The canonical action of $\mathrm{PSL}_2(\mathbb{R})$ on $\mathrm{T}^1\mathbb{H}^2$ descends to an action onto $\mathrm{T}^1 S$. This action preserves η_S. We take $(Y, \eta) = (\mathrm{T}^1 S, \eta_S)$.

If we write $q : \mathrm{T}^1 S \to S$ for the natural projection then $f \circ q$ is an element of $L^\infty(\mathrm{T}^1 S, \eta_S)$ 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. $\square$

4 Geometric background

This section reviews some standard facts and introduces some not-so-standard notation around the geometry of hyperbolic surfaces needed for the proof of Lemma 3.1. 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 $ds^2 = dz^2/y^2$. The **canonical horoball** is the subset

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

A **cusp** is a surface isometric to a quotient of the form $C := H_0/\{z \mapsto z + x_0\}$ for some $x_0 > 0$.

A **pair of pants** with $k \in \{0, 1, 2, 3\}$ cusps is an oriented complete hyperbolic surface P satisfying:

1. P is homeomorphic to the 2-sphere minus k points and $3 - k$ pairwise disjoint open disks,

2. the $3 - k$ boundary components of P are geodesic curves.

The following facts are classical [**?**]:

1. $\text{area}(P) = 2\pi$,

2. P is determined up to orientation-preserving isometry by the number of cusps k and the lengths of its boundary components,

3. there exist k pairwise-disjoint cusps on P.

5 Deformations of surfaces

The proof of Lemma 3.1 constructs surfaces and L^1-functions inductively by cutting, pasting and deforming. This 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 one cusp and two boundary components, denoted by $\partial^1 P, \partial^2 P$,
- the complement $S \setminus \mathrm{int}(P)$ is disconnected.

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

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

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.

5.1 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 $x \in S \setminus \mathrm{int}(P)$, let $x_\alpha = i_\alpha(x) \in S_\alpha$ and for $f \in L^1(S \setminus \mathrm{int}(P))$, define $f_\alpha \in L^1(S_\alpha)$ by

$$f_\alpha(x_\alpha) = \begin{cases} f(x) & x \in S \setminus \mathrm{int}(P) \\ 0 & \text{otherwise} \end{cases}$$

Proposition 5.1. *Let (S, P) be a panted surface and $f \in L^1(S \setminus \mathrm{int}(P))$. For any $r > 0$, the map*

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

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

Proof. For convenience, set $S_0 = S, P_0 = P, p^i_0 = p^i$, etc.

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 v^1_α and v^2_α point towards each other. Since γ_α is the shortest geodesic from $\partial^1 P_\alpha$ to $\partial^2 P_\alpha$, v^i_α meets $\partial^i P_\alpha$ at a right angle.

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 $\widetilde{g}_\alpha$ be the component of $\pi_\alpha^{-1}(g_\alpha)$ that contains the basepoint of w^1. Let w^2_α be the unit vector based at the other end point of $\widetilde{g}_\alpha$ so that w^2_α and w^1 point towards each other. Then the derivative of π_α maps w^2_α to v^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.

Define the deck-transformation groups

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

Then Λ_α is generated by Λ^1_α and Λ^2_α (by Van Kampen's Theorem).

We claim that for each i, Λ^i_α varies continuously with α. In fact, the choice of π_α implies that Λ^1_α does not depend on α. Let g_α be the unique orientation-preserving isometry of the hyperbolic plane that maps w^2_0 to w^2_α. Then $\Lambda^2_\alpha = g_\alpha \Lambda^2_0 g_\alpha^{-1}$. Because g_α varies continuously in $\alpha \in [0,\infty)$, Λ_α varies continuously in α. It follows that if $\widetilde{f}_\alpha : \mathbb{H}^2 \to \mathbb{C}$ is defined by $\widetilde{f}_\alpha(x) := f_\alpha(\pi_\alpha(x))$ for $x \in X_\alpha$ and $\widetilde{f}_\alpha = 0$ otherwise, then $\widetilde{f}_\alpha$ varies continuously in the sense that for any $r > 0$ the map

$$(x,\alpha) \in \mathbb{H}^2 \times [0,\infty) \mapsto \mathrm{area}(B_r)^{-1} \int_{B_r(x)} \widetilde{f}_\alpha(y)\,\mathrm{d}y$$

is continuous where the integral is with respect to the area measure on $\mathbb{H}^2$. However the right-hand side equals $\beta_r f_\alpha(\pi_\alpha(x))$ by definition. So this implies the proposition.

$\square$

6 Averaging around cusps

The main result of this section is a comparison between the averages of the form $\beta_r(f)(p)$ and $\beta_r(f1_C)$ where C is a cusp of the surface. This is used in the proof of Lemma 3.1 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 $H_0 = \{x + iy : \ y \geq 1\}$ is the canonical horoball and $x_0 > 0$ is the length of the boundary of C (which is a horocycle). For $t > 0$, let

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

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

Proposition 6.1. *Let S be a hyperbolic surface with pairwise disjoint cusps $C_1, \dots, C_k \subset S$. 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 $t, r \geq 0$,*

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

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 $p \in \mathbb{H}^2 \setminus H_0$,

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

Before estimating the above, here are some general facts about area 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 depends on the choice of B and H only through R and T) and for any fixed $t_0 > 0$, $g(T+t_0,T)$ is monotone increasing in T. To see this, we may assume $H = H_0$. Set B_T equal to the ball of hyperbolic radius $T + t_0$ and hyperbolic center $e^{-T}i$. Then $g(T+t_0,T) = \text{area}(H_0 \cap B_T)$. Also B_T coincides with the Euclidean disk centered on the imaginary axis that contains $e^{t_0}i$ and e^{-2T-t_0} 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

$$\text{area}(B(r+t,p)\cap\{x+iy:\ y \geq e^t\}) = g(r+t, d(p,H_0)+t) \geq g(r, d(p,H_0)) = \text{area}(B(r,p)\cap H_0).$$

So it suffices to show

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

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

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

□

7 The inductive step

To prove Lemma 3.1, 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 $(S, P, \{C_i\}_{i=1}^k, U, f)$ is **good** if

1. (S,P) is a panted surface,
2. S is a complete hyperbolic surface with finite area and no boundary,
3. $C_1,\dots,C_k \subset S$ 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. $\|f\|_1 \le 2$.

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

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

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

The next result forms the inductive step in the proof of Lemma 3.1.

Proposition 7.1. *Let $(S, P, \{C_i\}_{i=1}^k, U, f)$ be a good tuple, $\rho > 0$ and $\epsilon > 0$. Let*

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

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

1. $\operatorname{area}(\widehat{S}) = 2\operatorname{area}(S) + 2\pi$,
2. *if*
$$\widehat{V} = \{p \in \widehat{S} \setminus (\widehat{P} \cup \widehat{U}) : \ \mathsf{M}_\rho \widehat{f}(p) \ge 1\}$$
then $\operatorname{area}(\widehat{V}) \ge 2\operatorname{area}(V) - 3\epsilon$,
3. $\|\widehat{f}\|_1 \le \dfrac{\|f\|_1(1 - \|f\|_1/2)}{1 - 4\epsilon - 2e^{-\rho}}$.

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

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

for all $p \in G$.

By Proposition 5.1, there exists $\alpha > 0$ such that if S_α and f_α are defined as in §5.1 then

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

for all $p \in G$. Here we are identifying G 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 $V^{(i)}, U^{(i)}, G^{(i)} \subset S^{(i)}$ 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

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

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 the Theorem 1.1, there exists $t > 0$ and $W' \subset W^{(2)}$ such that area$(W') \geq$ 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

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

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

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

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

Set

$$\widehat{f} := \frac{\bar{f}}{1 - 4\epsilon - 2e^{-\rho}}.$$

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

The next step is to verify the maximal function estimates. If $p \in W^{(1)}$, then the definition of W implies

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

Therefore

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

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 6.1,

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

Therefore,

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

Therefore, $\mathsf{M}_\rho \widehat{f}(p) \geq 1$. Together with inequality (1) 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).

The equality $\left\|1_{U^{(2)}[t]} f^{(2)}\right\|_1 = e^{-t}\|f\|_1$ follows from linearity and the fact that area$(C[t]) = e^{-t}$area(C) for any cusp C. So

$$\text{area}(\widehat{S}) \int \bar{f}\ d\nu_{\widehat{S}} \quad = \quad \text{area}(S) \left(\int f\ d\nu_S \right) (2 - \|f\|_1).$$

Therefore,

$$\|\bar{f}\|_1 = \frac{\text{area}(S)(2 - \|f\|_1)}{2\,\text{area}(S) + 2\pi} \|f\|_1 \leq \|f\|_1 (1 - \|f\|_1/2)$$

which implies conclusion (3).

$\square$

8 The end of the proof

The next lemma establishes the base case of the induction in the proof of Lemma 3.1.

Lemma 8.1. *For every $\rho \geq 0$, there exists a good tuple $(S, P, \{C_i\}_{i=1}^4, U, f)$ such that*

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

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.

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

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

Let $C_1^{(i)}, C_2^{(i)} \subset Y_i$ be disjoint cusps such that for any $p \in V_i$ and $q \in C_1^{(i)} \cup C_2^{(i)}$, $d(p, q) \geq \rho$. 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

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

By the pointwise ergodic theorem 1.1, for a.e. $p \in S$, $\mathsf{M}f(p) \geq 2$. 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

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

this finishes the proof.

$\square$

Lemma 8.2. *Let $t_1, t_2, \ldots$ be a sequence of real numbers $t_i \in [0, 2)$ such that $t_{i+1} \leq t_i(1-t_i/2)$ for all i. Then $\lim_{i\to\infty} t_i = 0$.*

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

Proof of Lemma 3.1. For $b, \rho > 0$, let $\Sigma(b, \rho)$ be the set of all numbers $\epsilon > 0$ such that there exists a good tuple $(S, P, \{C_i\}_{i=1}^k, U, f)$ satisfying

1. $f \geq 0$,
2. $\|f\|_1 \leq \epsilon$,
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 $0 \in \overline{\Sigma(b, 1)}$ for some $b > 0$.

Lemma 8.1 proves that $1 \in \Sigma(1/2, \rho)$ for all ρ. Proposition 7.1 proves: if $\delta \in \Sigma(b, \rho)$ for all $\rho \geq 1$ then $\delta(1 - \delta/2) \in \overline{\Sigma(b - \epsilon, \rho)}$ for all $\epsilon > 0$ and $\rho \geq 1$. By iterating and using Lemma 8.2, this implies $0 \in \overline{\Sigma(1/2 - \epsilon, \rho)}$ for all $\epsilon > 0$ and $\rho \geq 1$ which finishes the lemma.

$\square$

9 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 for this action?