

Adelic dynamics and arithmetic quantum unique ergodicity

Elon Lindenstrauss

1. Introduction

Let M be a complete Riemannian manifold with finite volume which we initially assume to be compact. Then since M is compact, $L^2(M)$ is spanned by the eigenfunctions of the Laplacian Δ on M .

Many interesting questions can be asked about these eigenfunctions and their properties, and of these we focus on one, quantum ergodicity, which to the best of my knowledge was first considered by A.I. Šnirel'man, and was substantially sharpened in the work of Z. Rudnick and P. Sarnak which deals with the equidistribution properties of these eigenfunctions.

Specifically, let ϕ_n be a complete orthonormal sequence of eigenfunctions of Δ ordered by eigenvalue. These can be interpreted for example as the steady states for Schroedinger's equation

$$i \frac{\partial \psi}{\partial t} = \Delta \psi$$

describing the quantum mechanical motion of a free (spinless) particle on M . According to Bohr's interpretation of quantum mechanics $\tilde{\mu}_n(A) := \int_A |\phi_n(x)|^2 d\text{vol}(x)$ is the probability of finding a particle in the state ϕ_n inside the set A at any given time.

A.I. Šnirel'man, Y. Colin de Verdière and S. Zelditch [26, 8, 30] have shown that whenever the geodesic flow on M is ergodic, for example if M has negative curvature, there is a subsequence n_k of density one on which $\tilde{\mu}_{n_k}$ converge in the weak* topology to the uniform measure $\frac{1}{\text{vol}(M)} d\text{vol}_M$ ¹. This phenomenon is called quantum ergodicity.

The author gratefully acknowledges the support of the Clay Mathematics Institute in the form of a Clay research fellowship and of NSF grant DMS-0434403.

¹I.e., for every continuous (compactly supported in more general situations) function f one has that $\int f d\mu_{n_k} \rightarrow \frac{1}{\text{vol}(M)} \int f d\text{vol}_M$.

This leaves unanswered the question of what happens with the remaining eigenfunctions. In the analogue case of planar billiards E.J. Heller has numerically observed eigenfunctions which seem to be enhanced near-closed geodesics, a phenomenon he called scarring. In its strongest form this phenomenon will be manifested if there is some (necessarily zero density) subsequence n_k so that

$$\lim_{k \rightarrow \infty} \tilde{\mu}_{n_k}(C) > 0$$

for some closed geodesic C in M .

It is worth explaining why closed geodesics are singled out in this context. Let S^*M denote the unit cotangent bundle of M with $\pi : S^*M \rightarrow M$ the natural projections. In addition to proving quantum ergodicity, in [26, 8, 30] Šnirel'man, Colin de Verdière and Zelditch defined the micro-local lift of these measures $\tilde{\mu}_k$. In one of its variants, the microlocal lift of $\tilde{\mu}_k$ (a measure on M) is a measure μ_k on S^*M , explicitly defined in terms of ϕ_k , so that its projection to M is close to μ_k , and is close to being invariant under the geodesic flow (in a quantifiable way depending on the eigenvalue λ_k). In particular any weak* limit $\tilde{\mu}$ of a subsequence $\tilde{\mu}_k$ is of the form $\pi_*\mu$ with μ invariant under the geodesic flow on S^*M . The uniform measure on a periodic trajectory of the geodesic flow is the most degenerate invariant measure possible, and its projection to M is the uniform measure on a periodic geodesic. We shall call any measure μ on S^*M that actually arises as the weak star limit of these micro-local lifts a **quantum limit**.

In their paper [21], Rudnick and Sarnak have made the following conjecture:

CONJECTURE 1 (Quantum unique ergodicity). *If M is a compact manifold of negative curvature, the normalized volume measure on S^*M is the only quantum limit.*

It follows in particular that scarring in its strong form stated above is impossible on negatively curved manifolds. We stress that in addition to normalized volume and the periodic geodesics there is a whole slew of invariant measures of varying entropy, support, and mixing properties. For some M , namely arithmetic quotients of \mathbb{H} , and for an appropriate choice of complete orthonormal sequence of eigenfunctions, Rudnick and Sarnak also proved in [21] that scarring in this strong form is indeed impossible (actually, they have shown that a slightly stronger form of scarring is impossible, but a modification of the same techniques [11] establishes that scarring as defined above does not occur). In its general form, the strongest evidence towards this conjecture seems to be

extensive numerics by A. Barnett [2], and a result of N. Anantharaman [1] which gives under some assumption on the L^p -norms of the eigenfunctions that any quantum limit has positive entropy.²

This special case considered by Rudnick and Sarnak has attracted much attention, if only because there are many more tools at our disposal in dealing with these nice arithmetic eigenfunctions. Since this is precisely the case we will consider we give some more details.

A hyperbolic surface of constant negative curvature can be written as $M = \Gamma \backslash \mathbb{H}$ with $\Gamma < \mathrm{PSL}(2, \mathbb{R})$. For some Γ , this manifold has extra symmetry: for all but finitely many $n \in \mathbb{N}$ there is a **correspondence** T_n which maps a point $x \in M$ to a set $T_n(x)$ of finitely many points in M , and each branch of the map $x \mapsto T_n(x)$ is locally an isometry. Using these correspondences one can define operators, the **Hecke operators**, on various function spaces on M by setting

$$[T_n(f)](x) = n^{-1/2} \sum_{y \in T_n(x)} f(y).$$

These operators (possibly after throwing away finitely many of them) turn out to be self adjoint, commute with each other and with the Laplacian, and so (at least in the compact case) have a complete orthonormal sequence of joint eigenfunctions of Δ and all T_n . We give a construction of these operators from the Adelic viewpoint in Section 5.

Such joint eigenfunctions are called Hecke Maass forms and are of great intrinsic interest: in particular, their study is a key tool in modern analytic number theory. We shall call any quantum limit on an arithmetic surface coming from Hecke Maass forms an **arithmetic quantum limit**.

In this note we will deal exclusively with the case Γ a congruence lattice in $\mathrm{SL}(2, \mathbb{Z})$ which alas fails to be cocompact. However, everything we will do is also valid for Γ arising from quaternionic division algebras over \mathbb{Q} that are unramified over \mathbb{R} which are cocompact (the details of this case are fully written in [4, 12]).

If $\Gamma \backslash \mathbb{H}$ is not compact then in general the quantum unique ergodicity problem needs to be reformulated, since in general one cannot expect to have many L^2 -eigenfunctions of the Laplacian. This is less of an issue in the arithmetic quotients we consider here since in this case Selberg [22] has shown that there is an abundance of eigenfunctions (in fact, as many as we expect to have in a compact surface of the same volume); furthermore the continuous spectrum given by Eisenstein series is very well understood, and the appropriate version of quantum unique

²(Added in proof) N. Anantharam has recently eliminated the need for any assumptions regarding the L^p -norms, proving for any compact hyperbolic surface that any quantum limit has positive entropy. This recent and remarkable result cannot be used as a substitute for Theorem 1.3 below as it does not give sufficiently strong information regarding the entropy of each ergodic component.

ergodicity has been established for these Eisenstein series using analytic methods by Luo, Sarnak and Jakobson [14, 10]. For a simple proof for the abundance of L^2 -eigenfunctions of Δ on arithmetic $\Gamma \backslash \mathbb{H}$ see [13].

Assuming the Riemann hypothesis for suitable automorphic L-functions, T. Watson [27] has shown that the only arithmetic quantum limit, both in the compact and noncompact case, is the normalized volume measure. In fact, to obtain this conclusion one does not need the full force of the Riemann hypothesis but only subconvexity estimates on the value of these L-functions at $1/2$, which are known for some families of L-functions but not for the ones appearing in Watson's work. Assuming the full force of the Riemann hypothesis gives a rate of convergence of the $\tilde{\mu}_k$ to the uniform measure that is known to be best possible [15].

Using a result about flows on homogeneous spaces, the author has been able to show unconditionally the following regarding arithmetic quantum limits:

THEOREM 1.1 ([12]). *Let M be $\Gamma \backslash \mathbb{H}$ for Γ a congruence lattice as above. Then any arithmetic quantum limit for M is $\frac{c}{\text{vol} M} \text{vol}_{S^* M}$ for some $c \in [0, 1]$; in the compact case, $c = 1$.*

This (almost) proves the arithmetic case of the quantum unique ergodicity conjecture. The only missing piece is showing $c = 1$ also in the non compact case. Note that this theorem does not exclude the possibility that $c = 0$, i.e., that all the mass escapes to the cusp(s).

It is an amusing coincidence that in the subconvexity approach, establishing that arithmetic quantum limits are probability measures (i.e., that no mass escapes to the cusp(s)) involves proving subconvexity for simpler (lower order) L -functions than what is needed to show that what remains in the limit is equidistributed. Needless to say, from the ergodic theoretic point of view controlling this hypothetical escape of mass seems to be the harder part, or at least the part which is at present unknown. . .

The proof of this theorem consists of two separate parts. One of these parts is a partial classification of invariant measures satisfying an auxiliary, recurrence condition:

Let L be an S -algebraic group, K a compact subgroup of L , $G = \text{SL}(2, \mathbb{R}) \times L$ and Γ a discrete subgroup of G (for example, Γ can be a lattice of G), and consider the quotient $X = \Gamma \backslash G/K$.

The diagonal subgroup

$$A = \left\{ \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} : t \in \mathbb{R} \right\} \subset \text{SL}(2, \mathbb{R})$$

acts on X by right translation. In this paper we wish to study probability measures μ on X invariant under this action.

Without further restrictions, one does not expect any meaningful classification of such measures. For example, one may take $L =$

$\mathrm{SL}(2, \mathbb{Q}_p)$, $K = \mathrm{SL}(2, \mathbb{Z}_p)$ and Γ the diagonal embedding of $\mathrm{SL}(2, \mathbb{Z}[\frac{1}{p}])$ in G . As is well-known,

$$(1.1) \quad \Gamma \backslash G / K \cong \mathrm{SL}(2, \mathbb{Z}) \backslash \mathrm{SL}(2, \mathbb{R}).$$

Any A -invariant measure μ on $\Gamma \backslash G / K$ is identified with an A -invariant measure $\tilde{\mu}$ on $\mathrm{SL}(2, \mathbb{Z}) \backslash \mathrm{SL}(2, \mathbb{R})$. The A -action on $\mathrm{SL}(2, \mathbb{Z}) \backslash \mathrm{SL}(2, \mathbb{R})$ is very well understood, and in particular such measures $\tilde{\mu}$ are in finite-to-one correspondence with shift invariant measures on a specific shift of finite type [23] — and there are plenty of these.

In order to get a useful classification of invariant measures, one needs to impose an additional assumption relating μ with the foliation of X by L -orbits. The condition we consider is that of recurrence: that is that for every $B \subset X$ with $\mu(B) > 0$, for almost every $x \in X$ with $x \in B$ there are elements $x' \in xL$ arbitrarily far along this orbit with $x' \in B$; for a formal definition see Definition 2.2.

THEOREM 1.2. *Let $G = \mathrm{SL}(2, \mathbb{R}) \times L$, where L is an S -algebraic group, $H < G$ the $\mathrm{SL}(2, \mathbb{R})$ factor of G and K a compact subgroup of L . Take Γ to be a discrete subgroup of G (not necessarily a lattice) such that $\Gamma \cap L$ is finite. Suppose μ is a probability measure on $X = \Gamma \backslash G / K$, invariant under multiplication from the right by elements of the diagonal group $\begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix}$. Assume that*

- (1) *All ergodic components of μ with respect to the A -action have positive entropy.*
- (2) *μ is L/K -recurrent.*

Then μ is a linear combination of algebraic measures invariant under H .

The proof of this theorem relies on ideas developed by Ratner to deal with unipotent flows [19, 18, 20], in conjunction with some elements from work of M. Einsiedler and A. Katok [9] and a maximal inequality joint with D. Rudolph [12, Appendix].³

If no entropy assumption is made there is no reason for μ to be H -invariant but it is still reasonable to conjecture that μ must be a linear combination of algebraic measures. At the moment showing this seems quite hard.

It is fairly straightforward to show that arithmetic quantum limits are $G(\mathbb{Q}_p) / G(\mathbb{Z}_p)$ -recurrent; indeed implicitly it has been done already in [11]. More complicated is studying the entropy, but in a joint work with J. Bourgain we show

THEOREM 1.3 ([4]). *Every ergodic component of an arithmetic quantum limit has entropy $> \frac{2}{9}$ (in this normalization, the entropy of the uniform Lebesgue measure is 2).*

³In the proof one also uses a very simple case of the S -arithmetic version of Ratner's theorem ([17], [16]), which can also be shown directly.

We remark that the first attempt to use Hecke operators to give explicit bounds on the modulus of continuity of quantum limits (which is what entropy estimates eventually boil down to) is due to S. Wolpert [28].

The purpose of this paper is to place these two results in the appropriate more general context. In particular we show how exactly the same techniques give the following seemingly unrelated result, announced in [12]:

THEOREM 1.4. *Let \mathbb{A} denote the ring of Adeles over \mathbb{Q} . Let $A(\mathbb{A})$ denote the diagonal subgroup of $\mathrm{PGL}(2, \mathbb{A})$, and let μ be an $A(\mathbb{A})$ -invariant probability measure on $X = \mathrm{PGL}(2, \mathbb{Q}) \backslash \mathrm{PGL}(2, \mathbb{A})$. Then μ is the $\mathrm{PGL}(2, \mathbb{A})$ -invariant measure on X .*

Using the same general strategy, L. Silberman and A. Venkatesh have been able to prove a version of arithmetic quantum unique ergodicity for other $\Gamma \backslash G/K$, specifically for locally symmetric spaces arising from division algebras of prime degree. While the strategy remains the same, several new ideas are needed for this extension, in particular a new micro-local lift for higher rank groups [24].

2. Measures on G -spaces

Let G be a second countable locally compact group. It is customary to call a space X on which G acts a G -space. The type of space and the regularity of the action is usually implicit. In this paper we will consider some explicit X with lots of structure, and try to analyze how a probability measure μ interacts with the action of G . The abstract framework in which we can perform much of the analysis is that of a continuous action: we assume that X is a second countable locally compact space and that the map $(g, x) \mapsto g.x$ is continuous,⁴ and accordingly unless we specifically state otherwise, all G -space we shall consider are of this type. Under suitable assumptions these general results about decomposition of measures etc. are likely extendable to Borel actions on a standard Borel space⁵. To avoid pathologies, we will assume that for μ a.e. x the map $g \mapsto g.x$ is injective. Nothing further is assumed on the measure μ ; in particular the action of G need not preserve μ nor even its measure class.

For our purposes the action would be much less important than the foliation of X into orbits. One can study more general spaces, which we

⁴We will always denote our group actions as a left actions $x \mapsto g.x$. If $x \in \Gamma \backslash G$ and $g \in G$ we set $g.x = xg^{-1}$.

⁵We recall that a space X equipped with a sigma algebra of measurable sets \mathcal{B} is a standard Borel space if there is a bijective map $\iota : X \rightarrow [0, 1]$ which sends \mathcal{B} to the sigma algebra of Borel measurable subsets of the unit interval. For example, if X is any locally compact metric space, then X equipped with the sigma algebra of Borel sets is a standard Borel space.

have called (G, T) -spaces in [12, Section 2] with T a homogeneous space of G (i.e., a space in which G acts transitively with compact stabilizer). Roughly a (G, T) -space is a space foliated by leaves modeled on T . For simplicity, in this paper we consider only group actions and not these more general objects.

EXAMPLE 2.1. Suppose X_0 is any locally compact metric space, and G a locally compact group. Then $X = X_0 \times G$ is a G space with $g.(x_0, g') = (x_0, gg')$.

For any G -space X and $x \in X$, we shall let $t_x : G \rightarrow X$ be the map $g \mapsto g.x$.

DEFINITION 2.2. We say that a Radon measure μ on a G -space X is **recurrent** if for every measurable $B \subset X$ with $\mu(B) > 0$, for almost every $x \in B$ it holds that $t_x^{-1}(B)$ is unbounded, i.e., has a non-compact closure.

The opposite of recurrence is transience:

DEFINITION 2.3. We say that a Radon measure μ on a G -space X is **transient** if there is a sequence $B_n \subset X$ with $\mu(X \setminus \bigcup_n B_n) = 0$ so that for every $x \in B_n$, $t_x^{-1}(B_n)$ is bounded.

PROPOSITION 2.4. *Let μ be a radon measure on a G -space X .*

- (1) *If μ is recurrent (or transient), and if $\mu(B) > 0$ then the same is true for $\mu|_B$.*
- (2) *If μ is recurrent (or transient), then the same holds for any measure in its measure class.*
- (3) *μ can be uniquely written as $\mu = \mu_1 + \mu_2$ with μ_1 recurrent and μ_2 transient.*

PROOF. The first two statements of this proposition follow directly from the definitions; in particular (2) holds since the definitions only involve notions which are measure class invariant.

Suppose μ is not recurrent. Then there exist sets $B' \subset B$ with $\mu(B') \neq 0$ so that for every $x \in B'$ the set $t_x^{-1}(B)$ is bounded. In particular, $t_x^{-1}(B')$ is bounded.

Let \mathcal{C}_0 be the collection of such B' , and let $B'_0 \in \mathcal{C}$ be such that

$$\mu(B'_0) > \frac{1}{2} \sup_{B' \in \mathcal{C}} \mu(B').$$

Define inductively for $n > 0$ the collection \mathcal{C}_n of all such B' disjoint from $\bigcup_{k < n} B'_k$ and choose a $B'_n \in \mathcal{C}_n$ so that

$$\mu(B'_n) > \frac{1}{2} \sup_{B' \in \mathcal{C}_n} \mu(B').$$

Let $B_\infty = \bigcup_n B'_n$. Then by definition $\mu_2 = \mu|_{B_\infty}$ is transient, and it is a simple exercise to show that $\mu_1 = \mu|_{X \setminus B_\infty}$ is recurrent.

Regarding uniqueness: by definition the measures μ_1 and μ_2 we have constructed above are mutually singular. Suppose μ can also be represented as $\mu = \nu_1 + \nu_2$ with ν_1 recurrent and ν_2 transient.

Then if $\nu_2(B) > 0$ there is a $B' \subset B$ with positive ν_2 measure satisfying that $t_x^{-1}(B')$ is bounded for every $x \in B'$. Indeed, since ν_2 is transient there is a B_n with $t_x^{-1}(B_n)$ bounded for every $x \in B_n$ for which $\nu_2(B \cap B_n) > 0$; take $B' = B \cap B_n$.

Conversely, if $t_x^{-1}(B)$ is bounded for every $x \in B'$ then since ν_1 is recurrent we have that $\nu_1(B) = 0$. If $\mu(B) > 0$ this implies that $\nu_2(B) > 0$. We conclude that $\nu_2(B) > 0$ iff there is a $B' \subset B$ with positive μ measure satisfying that $t_x^{-1}(B')$ is bounded for every $x \in B'$. The same holds of course also for μ_2 , so μ_2 and ν_2 are in the same measure class. Similarly one can show that μ_1 and ν_1 are in the same measure class, and since μ_1 and μ_2 are mutually singular this means that $\mu_1 = \nu_1$ and $\mu_2 = \nu_2$. \square

Another proof of Proposition 2.4 can be deduced from Theorem 4.2 below.

3. Conditional measure on G -orbits

A basic construction which is very helpful in analyzing measures on G -spaces is the construction of conditional measures on G -orbits, which we shall also refer to as G -leaves.

Throughout this section, we take μ to be a Radon measure on X and assume t_x is injective for μ -almost every $x \in X$.

Let $\mathcal{M}_\infty(G)$ denote the space of all Radon (in particular, locally finite) measures on G , equipped with the smallest topology so that the map $\nu \mapsto \int f d\nu$ is continuous for every continuous compactly supported $f \in C_c(G)$. Note that since G is a locally compact separable metric space, $\mathcal{M}_\infty(G)$ is separable and metrizable (though in general not locally compact).

The purpose of this section is to show how the measure μ on X induces a locally finite measure on almost every G -orbit which is well defined up to a normalizing constant. More formally, we define a measurable map $x \mapsto \mu_{x,G}$ from X to $\mathcal{M}_\infty(G)$ with the properties described below in Theorem 3.3; we emphasize that in general $\mu_{x,G}$ will not be finite measures.

We take d_G to be some right G invariant metric on G (i.e., $d_G(xg, yg) = d_G(x, y)$), and use $B_r^G(g)$ (respectively $\bar{B}_r^G(g)$) to denote the open (closed) ball of radius r around $g \in G$; it would also be convenient to use the notations $B_r^G = B_r^G(e)$ and for $x \in X$, $B_r^G(x) = t_x(B_r^G)$.

We recall that a sigma ring is a collection of sets \mathcal{A} which is closed under countable unions and under set differences (i.e., if $A, B \in \mathcal{A}$ then so is $A \setminus B$). Unless specified otherwise, all sigma rings we consider will

be countably generated sigma rings of Borel sets, and in particular have a maximal element.

DEFINITION 3.1. Let $\mathcal{A} \subset \mathcal{B}$ be a countably generated sigma ring, and let $\mathcal{C} \subset \mathcal{A}$ be a countable ring of sets which generates \mathcal{A} . The **atom** $[x]_{\mathcal{A}}$ of a point $x \in X$ in \mathcal{A} is defined as

$$[x]_{\mathcal{A}} = \bigcap_{C \in \mathcal{C}: x \in C} C = \bigcap_{A \in \mathcal{A}: x \in A} A.$$

Let $\mathcal{A} \subset \mathcal{B}$ be a countably generated sigma ring, μ a Radon measure, and assume that the μ -measure of the maximal element of \mathcal{A} is finite. Then we can consider the decomposition of μ with respect to the sigma ring \mathcal{A} , i.e., a set of probability measures $\{\mu_x^{\mathcal{A}} : x \in X\}$ on X with the following properties.

- (1) For all $x, x' \in X$ with $[x]_{\mathcal{A}} = [x']_{\mathcal{A}}$,
- (3.1) $\mu_x^{\mathcal{A}} = \mu_{x'}^{\mathcal{A}}$ and $\mu_x^{\mathcal{A}}([x]_{\mathcal{A}}) = 1$,
- (2) For every $B \in \mathcal{B}$, the map $x \mapsto \mu_x^{\mathcal{A}}(B)$ is \mathcal{A} -measurable.
- (3) For every $A \in \mathcal{A}$ and $B \in \mathcal{B}$,

$$(3.2) \quad \mu(A \cap B) = \int_A \mu_x^{\mathcal{A}}(B) d\mu(x).$$

We consider first the rather trivial G space $X = X_0 \times G$ as in Example 2.1. Let $\mathcal{B}_0 \subset \mathcal{B}$ be the inverse image of the sigma algebra of Borel subsets of X_0 with respect to the projection $X \rightarrow X_0$.

Then \mathcal{B}_0 is a countably generated sigma algebra, and its atoms are precisely the fibers $\{x_0\} \times G$, i.e., the G -leaves for this G -structure. We can decompose μ with respect to the sigma algebra \mathcal{B}_0 as above, obtaining a system of conditional measures (in this case, probability measures) $\mu_x^{\mathcal{B}_0}$ on the G -leaves, which we can then push forward to G using the map t_x^{-1} .

Such a countably generated sigma algebra whose atoms are precisely the G -leaves exists only rarely, and furthermore in this case the dynamics is rather dull (for example, in this case any probability measure is G -transient). In more interesting cases, one needs to work with countably generated sigma rings whose atoms are pieces of G -leaves, which we will call G -plaques:

DEFINITION 3.2. A set $A \subset X$ is an open G -plaque if for any $x \in A$:

- (i) $A \subset t_x(G)$ and
- (ii) $t_x^{-1}A$ is open in G .

By considering a sequence of such sigma rings whose atoms are bigger and bigger open G -plaques one can construct the full system of conditional measures, which we will denote by $\mu_{x,G}$. In general, particularly when $\mu_{x,G}$ is infinite, there is no natural normalization. However,

the identity $e \in G$ can be shown to be almost surely in the support of $\mu_{x,G}$, and so we can use the normalizer $\mu_{x,G}$ e.g., by requiring that

$$\mu_{x,G}(B_1^G(e)) = 1.$$

The following theorem which we quote from [12] uniquely determines $\mu_{x,G}$:

THEOREM 3.3. *Let X be a G -space, and μ a probability measure on X with t_x injective a.s. Then there is a Borel measurable map $\mu_{x,G} : X \rightarrow \mathcal{M}_\infty(G)$ which is uniquely determined (up to μ -measure 0) by the following two conditions:*

- (1) *For μ almost every x , we have that $\mu_{x,G}(B_1^G) = 1$.*
- (2) *For any countably generated sigma ring $\mathcal{A} \subset \mathcal{B}$ with maximal element E , if for every $x \in E$ the atom $[x]_{\mathcal{A}}$ is an open G -plaque, then for μ -almost every $x \in E$,*

$$t_x^{-1} * \mu_x^{\mathcal{A}} \propto \mu_{x,G}|_{t_x^{-1}[x]_{\mathcal{A}}}.$$

In addition, $\mu_{x,G}$ satisfies the following:

- (3) *There is a set $X_0 \subset X$ of full μ -measure so that for every $x, y \in X_0$ with $x \stackrel{T}{\sim} y$, and for any $g \in G$ satisfying $t_y \circ g = t_x$ we have that*

$$g * \mu_{x,G} \propto \mu_{y,G}.$$

For example, it follows that for $X = X_0 \times G$ as in Example 2.1⁶,

$$\mu_{x,G} = \frac{(t_x^{-1})_*(\mu_x^{\mathcal{B}_0})}{\mu_x^{\mathcal{B}_0}(B_1^G(x))}.$$

4. Recurrence and conditional measures

It turns out that recurrence and conditional measures are intimately connected. What we have called recurrence is more commonly called conservativity of measures; we will give the following definition for conservative measures which we will immediately show is equivalent to recurrence; however, quantifying Definition 2.2 and Definition 4.1 gives different notions.

DEFINITION 4.1. A probability measure μ on a G -space X is **G -conservative** if for μ -a.e. x one has that $\mu_{x,G}$ is an infinite measure.

THEOREM 4.2. *Let μ be a probability measure on a G -space X . Then*

- (1) *μ is G -recurrent if and only if $\mu_{x,G}$ is an infinite measure a.s., i.e., if μ is G -conservative;*
- (2) *μ is G -transient if and only if $\mu_{x,G}$ is finite a.s.*

⁶More generally, for any G -space for which there is a countably generated sigma algebra \mathcal{B}_0 whose atoms are precisely the G -leaves.

PROOF. Let $B = \{x : \mu_{x,G}(G) < \infty\}$. Then for any $\epsilon > 0$ there is a $r > 0$ and a Borel set $B_r \subset B$ so that

- (1) $\mu(B_r) > \mu(B) - \epsilon$
- (2) for every $x \in B_r$, one has that $\mu_{x,G}(B_r^G) > 0.9\mu_{x,G}(G)$.

Then for every $x \in B_r$, it holds that $t_x^{-1}(B_r) \subset B_{2r}^G$: indeed, if $t \in G \setminus B_{2r}^G(x^G)$ with $x, t_x(t) \in B_r$ then

$$\mu_{x,G}(G) \geq \mu_{x,G}(B_r^G \cup B_r^G(t)) = \mu_{x,G}(B_r^G) + \mu_{x,G}(B_r^G(t)) \geq 1.8\mu_{x,G}(G)$$

is a contradiction since $\mu_{x,G}(G) > 0$ almost everywhere.

We conclude that if $\mu_{x,G}$ is finite a.s. this sequence of sets B_r show that μ is transient; furthermore it follows that unless $\mu_{x,G}$ is infinite a.s., μ cannot be recurrent.

For the reverse direction, let $B' = \{x : \mu_{x,G}(x) = \infty\}$, and suppose $C \subset B'$ with $t_x^{-1}(C)$ bounded for every $x \in C$. We need to show $\mu(C) = 0$, which will imply that if μ is transient B' is a null set as well as that if B' is co-null then μ is recurrent.

By definition of C and since $C \subset B'$, there are for any $\epsilon > 0$ real numbers $R > r > 0$ and a compact $C' \subset C$ so that:

- (1) $\mu(C') > \frac{1}{2}\mu(C)$
- (2) $t_x^{-1}(C) \subset B_r^G$ for all $x \in C'$
- (3) $\mu_{x,G}(B_r^G) < \epsilon\mu_{x,G}(B_R^G(x^G))$ for any $x \in C'$.

Consider now the set $C_R = \bar{B}_R^G(C') = \bigcup_{x \in C'} \bar{B}_R^G(x)$. This is a compact set, as the image of the compact set $C' \times \bar{B}_R^G$ under the continuous map $(x, s) \mapsto x.s$. For every $x \in C_R$ we have that $t_x^{-1}(C_R)$ is bounded: indeed $t_x^{-1}(C_R) \subset \bar{B}_{R+r}^G$.

The map $\phi : x \mapsto \bar{B}_{R+r}^G(x) \cap C_R$ is a continuous map from C_R to 2^{C_R} , the space of closed subsets of C_R , equipped with the Hausdorff metric. The inverse image under ϕ of every $p = \phi(x) \in 2^{C_R}$ is precisely the slice $G.x \cap C_R$.

Therefore the pullback under ϕ of the sigma algebra of Borel sets in 2^{C_R} is a countably generated sigma algebra \mathcal{A} of (Borel) subsets of C_R whose atoms are precisely these slices.

By Theorem 3.3, $\mu_x^{\mathcal{A}} \propto (t_x)_*(\mu_{x,G}|_{t_x^{-1}(C_R)})$ and so by 1-3 above, for every $x \in C_R$ for which $\mu_x^{\mathcal{A}}(C') > 0$, for $\mu_x^{\mathcal{A}}$ a.e. $x' \in C' \cap [x]_{\mathcal{A}}$

$$\begin{aligned} \mu_x^{\mathcal{A}}(C') &= \mu_{x'}^{\mathcal{A}}(C') = \frac{\mu_{x',G}(t_{x'}^{-1}(C'))}{\mu_{x',G}(t_{x'}^{-1}(C_R))} \\ &\leq \frac{\mu_{x',G}(B_r^G)}{\mu_{x',G}(B_R^G)} < \epsilon. \end{aligned}$$

Since the above inequality also clearly holds if $\mu_x^{\mathcal{A}}(C') = 0$, we conclude that

$$\begin{aligned}\mu(C') &= \int_{C_R} \mu_x^{\mathcal{A}}(C') d\mu(x) \\ &< \epsilon \int_{C_R} \mu_x^{\mathcal{A}}(C_R) d\mu(x) \\ &= \epsilon \mu(C_R) \leq \epsilon \mu(X) = \epsilon.\end{aligned}$$

Thus $\mu(C) < 2\epsilon$ for arbitrary $\epsilon > 0$, establishing the theorem. \square

REMARK. Theorem 4.2 can be used to give an alternative proof of Proposition 2.4.(3), by setting for arbitrary μ its transient and recurrent parts to be $\mu_1 = \mu|_B$ and $\mu_2 = \mu|_{B'}$ respectively, B, B' as in the proof of Theorem 4.2.

5. The space $\mathrm{PGL}(2, \mathbb{Q}) \backslash \mathrm{PGL}(2, \mathbb{A})$

In this section we recall some well-known facts relating the Adelic quotient $\mathrm{PGL}(2, \mathbb{Q}) \backslash \mathrm{PGL}(2, \mathbb{A})$ and $\mathrm{PGL}(2, \mathbb{Z}) \backslash \mathrm{PGL}(2, \mathbb{R})$, and explain how the Hecke correspondence on the latter can be explained from the Adelic viewpoint.

Set

$$\begin{aligned}G_R &= \mathrm{PGL}(2, R) \quad \text{for every ring } R \\ A &= \begin{pmatrix} e^t & 0 \\ 0 & 1 \end{pmatrix} \in G_{\mathbb{R}} \\ G_f &= \prod'_{p \neq \infty} G_{\mathbb{Q}_p} \\ K_f &= \prod_{p \neq \infty} G_{\mathbb{Z}_p}.\end{aligned}$$

By slight abuse of notation we have used 2×2 -matrices, and not equivalence classes modulo the center of $\mathrm{GL}(2)$ to denote elements in $\mathrm{PGL}(2)$. It would also be convenient to use K_p as a shorthand for $G_{\mathbb{Z}_p}$. As is customary, the rational points $G_{\mathbb{Q}}$ can be considered as a subgroup of $G_{\mathbb{R}}$ and of $G_{\mathbb{Q}_p}$ for every finite prime p ; we can also consider these rational points as a subgroup of any product of the above groups via the diagonal embedding.

We shall say that a 2×2 matrix M with integer entries is primitive if $\det M \neq 0$ and M is not a nontrivial integer multiple of another matrix with integer entries. Denote the set primitive integer matrices by Λ . For each $\gamma \in G_{\mathbb{Q}}$ there is a unique $\tilde{\gamma} \in \Lambda$ proportional to it; and when convenient we shall implicitly identify between γ and $\tilde{\gamma}$. For every $n \in \mathbb{N}$ we let

$$\Lambda_n = \{\gamma \in \Lambda : \det \tilde{\gamma} = n\}.$$

We will also use the notation $d(\gamma) = |\det \tilde{\gamma}|$.

We recall that there is a natural isomorphism

$$(5.1) \quad G_{\mathbb{Z}} \backslash G_{\mathbb{R}} \cong G_{\mathbb{Q}} \backslash G_{\mathbb{A}} / K_f.$$

Let $\pi_{\mathbb{A} \rightarrow \mathbb{R}}$ denote the projection $G_{\mathbb{A}} = G_{\mathbb{R}} \times G_f \rightarrow G_{\mathbb{R}}$. The isomorphism (5.1) is explicitly given by the right K_f -invariant map $\pi : G_{\mathbb{Q}} \backslash G_{\mathbb{A}} \mapsto G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$

$$\pi : G_{\mathbb{Q}} g \mapsto \pi_{\mathbb{A} \rightarrow \mathbb{R}}(G_{\mathbb{Q}} g \cap [G_{\mathbb{R}} \times K_f])$$

since $G_{\mathbb{Q}} \cap K_f = G_{\mathbb{Z}}$ the image of this map is left $G_{\mathbb{Z}}$ invariant — indeed a single left $G_{\mathbb{Z}}$ coset.

The group G_f acts on $G_{\mathbb{Q}} \backslash G_{\mathbb{A}}$ in the obvious way ($g_f \cdot G_{\mathbb{Q}} g = G_{\mathbb{Q}} g g_f^{-1}$), and traces of this action persist in $G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$ in the form of the Hecke correspondence.

It follows from the definitions that

$$(5.2) \quad G_f = \bigcup_{n \in \mathbb{N}} K_f \begin{pmatrix} n & 0 \\ 0 & 1 \end{pmatrix} K_f \quad (\text{disjoint union}).$$

Given an element $g_f \in K_f \begin{pmatrix} n & 0 \\ 0 & 1 \end{pmatrix} K_f$, if one takes a point $G_{\mathbb{Z}} g_{\infty} \in G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$, and looks at $\pi(\pi^{-1}(G_{\mathbb{Z}} g_{\infty}) g_f)$ one obtains a finite subset of $G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$ that depends only on n and the point $G_{\mathbb{Z}} g_{\infty}$, which we shall denote by $T_n(G_{\mathbb{Z}} g_{\infty})$. The map $G_{\mathbb{Z}} g_{\infty} \mapsto T_n(G_{\mathbb{Z}} g_{\infty})$ is called the n -Hecke correspondence.

It will be convenient for us to have a more explicit description of the Hecke correspondences. For every n , the set Λ_n decomposes into finitely many left $G_{\mathbb{Z}}$ cosets. Take T_n to be a subset of Λ_n containing exactly one element of every left $G_{\mathbb{Z}}$ coset (we intentionally use the same notation T_n we have used for the Hecke correspondence). The double coset $K_f \begin{pmatrix} n & 0 \\ 0 & 1 \end{pmatrix} K_f$ can be decomposed into right K_f cosets as follows:

$$K_f \begin{pmatrix} n & 0 \\ 0 & 1 \end{pmatrix} K_f = \bigcup_{\gamma^{-1} \in T_n} \gamma K_f.$$

Then for $g_f \in K_f \begin{pmatrix} n & 0 \\ 0 & 1 \end{pmatrix} K_f$ as above and $g_{\infty} \in G_{\mathbb{R}}$,

$$\begin{aligned} \pi(\pi^{-1}[G_{\mathbb{Z}} g_{\infty}] g_f) &= \pi(G_{\mathbb{Q}} g_{\infty} \times [K_f g_f K_f]) \\ &= \pi \left(G_{\mathbb{Q}} g_{\infty} \times \left[K_f \begin{pmatrix} n & 0 \\ 0 & 1 \end{pmatrix} K_f \right] \right) \\ &= \bigcup_{\gamma^{-1} \in T_n} \pi(G_{\mathbb{Q}} g_{\infty} \times [\gamma^{-1} K_f]) \\ &= \bigcup_{\gamma^{-1} \in T_n} \pi(G_{\mathbb{Q}} [\gamma g_{\infty}] \times K_f) \\ &= G_{\mathbb{Z}} T_n g_{\infty} \end{aligned}$$

which is indeed a finite set of points as claimed, and we have, for $x = G_{\mathbb{Z}}g_{\infty} \in G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$, that $T_n(x) = G_{\mathbb{Z}}T_ng_{\infty} = \Lambda_n g_{\infty}$ is precisely the Hecke correspondence we have defined earlier. The Hecke correspondences give rise to the Hecke operators which are defined on functions over $G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$ by taking

$$[T_n f](x) = n^{-1/2} \sum_{y \in T_n(x)} f(y).$$

If $n = p^k$ with p a prime, we can choose

$$T_{p^k} = \left\{ \begin{pmatrix} p^k & 0 \\ b & 1 \end{pmatrix} : 0 \leq b < p^k \right\} \cup \left\{ \begin{pmatrix} 1 & 0 \\ 0 & p^k \end{pmatrix} \right\} \cup \bigcup_{j=1}^{k-1} \left\{ \begin{pmatrix} p^{k-j} & 0 \\ b & p^j \end{pmatrix} : 0 \leq b < p^{k-j}, (b, p) = 1 \right\},$$

and in general if $n = p_1^{k_1} \dots p_{\ell}^{k_{\ell}}$,

$$T_n = T_{p_1^{k_1}} \dots T_{p_{\ell}^{k_{\ell}}}.$$

We also remark that if n, m are relatively prime, $G_{\mathbb{Z}}T_nT_m = G_{\mathbb{Z}}T_mT_n = \Lambda_{nm}$. It will be convenient to set $T_1 = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}$ and $T_{*|n} = \bigcup_{m|n} T_m$.

6. Entropy of uniformly Adelic-conservative measures

Let X be a G -space, and μ a Radon measure on X . By definition, μ is G -conservative if and only if $\mu_{x,G}(G) = \infty$. One can quantify this as follows: suppose F_n is a sequence of subsets of G and $\rho_n \rightarrow \infty$. Then μ is (F_n, ρ_n) -uniformly conservative if for almost every x ,

$$\mu_{x,G}(F_n) \geq \rho_n.$$

We now fix our attention on the space $X = \mathrm{PGL}(2, \mathbb{Q}) \backslash \mathrm{PGL}(2, \mathbb{A})$ considered in the previous section.

We shall say that a probability measure μ on X is **ℓ -uniformly adelicly conservative** if there is some $c > 0$ so that for every prime p and every $x \in X$

$$(6.1) \quad \mu_{x,G_f} \left(\bigcup_{i=1}^{\ell} K_f \begin{pmatrix} p^i & 0 \\ 0 & 1 \end{pmatrix} K_f \right) \geq c \mu_{x,G_f}(K_f).$$

By (5.2) we see that this is indeed a quantification of G_f -conservativity in the sense considered above.

We can now state the following theorem, which is the main theorem we will prove in this note. For $\ell = 2$ this theorem is implicit in [4, 5], and the extension to general ℓ using the same techniques does not pose an additional difficulty. However, it seems useful to state and prove this result explicitly.

THEOREM 6.1. *Let μ be a probability measure on X invariant under A which is ℓ -uniformly adelicly conservative. Then the entropy of every ergodic component of μ with respect to A is positive.*

The same techniques applies also to the following case: let H be a quaternion division algebra defined over \mathbb{Q} which splits over \mathbb{R} . Let G be the group of invertible elements of H divided by the center. Then $G(\mathbb{Z})$ is a cocompact lattice in $G(\mathbb{R}) \cong \mathrm{PGL}(\mathbb{R})$. Indeed, this case is somewhat simpler since the space X is compact (again, the case $\ell = 2$ is implicitly proved in [4]).

Except for the number theoretic Theorem 7.6 this discussion generalizes also for quaternion division algebras over any number field (while we do not explore this direction here, in [11, Sect. 5] some of the techniques which eventually led to the work [4] are developed in precisely this context). It seems it should be possible also to find a suitable substitute to Theorem 7.6.

We will discuss two applications to Theorem 6.1. One is arithmetic quantum unique ergodicity, and the other the analysis of probability measures on $G_{\mathbb{Q}} \backslash G_{\mathbb{A}}$ invariant under the full Adelic diagonal group.

There is a trade-off regarding Theorem 6.1: one can simplify the proof if one makes a stronger assumption regarding the measure μ and its behavior with respect to the G_f -action. This stronger assumption is valid for the quantum unique ergodicity application, but not for the case of measures on $G_{\mathbb{Q}} \backslash G_{\mathbb{A}}$ invariant by the full Adelic torus.

This approach was independently found by Silberman and Venkatesh [25] and by Bourgain and myself (unpublished), and Silberman and Venkatesh have taken it much further and were able to establish analogs for much more general groups, for example for general \mathbb{R} -split algebraic groups.

Both arithmetic quantum limits and measures invariant under the full Adelic torus satisfy this uniformly adelicly conservativity criterion. For arithmetic quantum limits this is (implicitly) shown with $\ell = 2$ in [4].

We note the following:

PROPOSITION 6.2. *If μ is a measure on $G_{\mathbb{Q}} \backslash G_{\mathbb{A}}$ invariant under the full Adelic \mathbb{Q} -split torus then μ is 1-uniformly adelicly conservative.*

PROOF. It can be shown, similarly to [12, Prop. 4.3], that if X is a G -space, $H < G$, and ν is an H -invariant measure on X then for ν -a.e. x the conditional measure $\nu_{x,G}$ is left⁷ H -invariant. Since μ is $A(\mathbb{A})$ -invariant, and in particular $A(\mathbb{A}_f) < G_f$ invariant, we see that

⁷In [12], all actions were right actions; here they are all left actions, hence the sides are switched here vs. [12, Prop. 4.3].

μ_{x,G_f} is a.e. invariant under $\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$ for every p , hence

$$\mu_{x,G_f} \left(K_f \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} K_f \right) \geq \mu_{x,G_f} \left(\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix} K_f \right) = \mu_{x,G_f}(K_f)$$

and μ is 1-uniformly adelicly conservative with the implicit constant $c = 1$. \square

The proposition above, Theorem 6.1 and Theorem 1.2 imply Theorem 1.4.

7. Entropy and volume estimates for ℓ -uniformly adelicly conservative measures

In this section we prove an estimate on the measure of small tubes around geodesics for ℓ -uniformly conservative measures which will imply Theorem 6.1. First we set up some notations.

First we define the tubes: Identifying the Lie algebra of $G_{\mathbb{R}}$ with $\mathfrak{sl}(2, \mathbb{R})$ in the obvious way we define

$$\tilde{B}_0(\epsilon, \tau) = \left\{ \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & -x_{11} \end{pmatrix} : |x_{12}|, |x_{21}| < \epsilon, |x_{11}| < \tau \right\} \subset \mathfrak{sl}(2, \mathbb{R})$$

$$\tilde{B}(\epsilon, \tau) = \exp \tilde{B}_0(\epsilon, \tau) \subset G_{\mathbb{R}}$$

$$\tilde{B}(x, \epsilon, \tau) = x \tilde{B}(\epsilon, \tau) \subset G_{\mathbb{Z}} \backslash G_{\mathbb{R}} \quad \text{for } x \in G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$$

$$B(x, \epsilon, \tau) = x[\tilde{B}(\epsilon, \tau) \times K_f] \subset G_{\mathbb{Q}} \backslash G_{\mathbb{A}} \quad \text{for } x \in G_{\mathbb{Q}} \backslash G_{\mathbb{A}}.$$

We also recall the definition of the maximal function $M(f, x)$ for a Borel measurable function f on X :

$$M(f, x) = \sup_{T>0} \int_0^T f(xa(-t)) dt.$$

We can now state the main result of this section. In the next section we will show how this theorem indeed implies Theorem 6.1.

THEOREM 7.1. *Let μ be a ℓ -uniformly adelicly conservative probability measure on X , and let $\Omega \subset X$ be a compact subset. Then there are constants $\alpha, \beta > 0$, depending only on ℓ and $t > 0$, so that for every $x \in \Omega$ and $\epsilon > 0$ at least one of the following two possibilities holds:*

- (1) $\mu(B(x, \epsilon, t)) \ll_{t, \Omega} \epsilon^{\alpha}$;
- (2) $M(1_{\Omega^c}, x) > \beta$.

Note that since our tubes $B(x, \epsilon, t)$ are right K_f -invariant, and without loss of generality Ω can be taken to be K_f invariant, Theorem 7.1 is precisely equivalent to the analogous statement regarding the $\tilde{\mu} = \pi_*(\mu)$ measure of tubes $\tilde{B}(x, \epsilon, t)$ on $\pi(X) = G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$. This later point of view has the advantage that the space $G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$ is easier to visualize than the

larger space X — indeed, as is well-known, $G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$ is the unit cotangent bundle of the hyperbolic surface $\mathrm{SL}(2, \mathbb{Z}) \backslash \mathbb{H}$. In addition to the projection $\pi : G_{\mathbb{Q}} \backslash G_{\mathbb{A}} \rightarrow G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$, we will make use of the projection $\tilde{\pi} : G_{\mathbb{R}} \rightarrow G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$.

To enable us to work on $G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$ we need to understand how μ being ℓ -uniformly adelicly conservative is reflected by the properties of $\tilde{\mu}$:

PROPOSITION 7.2. *If μ is a ℓ -uniformly adelicly conservative probability measure on X then for the same $c > 0$ as in (6.1) it holds that for every Borel $B \subset G_{\mathbb{R}}$ and every prime p*

$$\sum_{\gamma \in T_{*|p^{\ell}}} \tilde{\mu}(G_{\mathbb{Z}}\gamma B) \geq (1+c)\tilde{\mu}(G_{\mathbb{Z}}B)$$

where $\tilde{\mu} = \pi_*\mu$.

PROOF. By dividing B into sufficiently many small pieces, without loss of generality we can assume that the collection of sets $\{G_{\mathbb{Z}}\gamma B\}_{\gamma \in T_{*|p^{\ell}}}$ is disjoint.

Define \mathcal{A} to be the following sigma ring of subsets of X

$$\mathcal{A} = \left\{ \pi^{-1} \left(T_{*|p^{\ell}}(A) \right) : A \text{ a Borel subset of } B \right\}.$$

Since the Borel subsets of B are countably generated, so is \mathcal{A} . For $x \in \pi^{-1}(B)$ the atom $[x]_{\mathcal{A}}$ satisfies

$$[x]_{\mathcal{A}} = x \left(\bigcup_{i=0}^{\ell} K_f \begin{pmatrix} p^i & 0 \\ 0 & 1 \end{pmatrix} K_f \right).$$

We can now apply Theorem 3.3.(2) and the ℓ -uniformly adelicly conservative criterion (6.1) to deduce that for every $x \in \pi^{-1}(T_{*|p^{\ell}}B)$,

$$\mu_x^{\mathcal{A}}(\pi^{-1}(B)) = \frac{\mu_{y, G_f}(K_f)}{\mu_{y, G_f} \left(\bigcup_{i=0}^{\ell} K_f \begin{pmatrix} p^i & 0 \\ 0 & 1 \end{pmatrix} K_f \right)} \leq \frac{1}{1+c}$$

where y is any point in $\pi^{-1}(B \cap T_{*|p^{\ell}}(\pi(x)))$. Integrating over $\pi^{-1}(T_{*|p^{\ell}}B)$, we get

$$\begin{aligned} \tilde{\mu}(T_{*|p^{\ell}}(B)) &\geq (1+c) \int_{\pi^{-1}(T_{*|p^{\ell}}B)} \mu_x^{\mathcal{A}}(\pi^{-1}(B)) d\mu(x) \\ &= (1+c)\mu(\pi^{-1}(B)) = (1+c)\tilde{\mu}(B). \end{aligned}$$

□

We can now outline our strategy, following closely [4, 5], to proving Theorem 7.1.

We want to bound the $\tilde{\mu}$ -measure of a tube $\tilde{B}(x, \epsilon, \tau)$ for a given $x = \Gamma g \in \Omega$. Consider the collection of tubes

$$\mathcal{V} = \left\{ \tilde{B}(y, \epsilon, \tau) : y \in T_{*|p^{\ell}}(x), p \text{ prime} \leq N \right\}$$

for suitably chosen N which would be a power of ϵ^{-1} .

If all of these tubes were disjoint, we would have that

$$\begin{aligned}
 (7.1) \quad 1 &\geq \sum_{V \in \mathcal{V}} \tilde{\mu}(V) \\
 &= \tilde{\mu}(\tilde{B}(x, \epsilon, \tau)) + \sum_{p \text{ prime} \leq N} \sum_{y \in T_{*|p^\ell}(x) \setminus \{x\}} \tilde{\mu}(\tilde{B}(y, \epsilon, \tau)) \\
 &\geq c |\{\text{primes} \leq N\}| \tilde{\mu}(\tilde{B}(x, \epsilon, \tau))
 \end{aligned}$$

and $\tilde{\mu}(\tilde{B}(x, \epsilon, \tau)) \ll N^{-1} \log N$. More generally, for any set of primes P

$$\begin{aligned}
 (7.2) \quad &\text{all sets in } \mathcal{V}_P := \left\{ \tilde{B}(y, \epsilon, \tau) : y \in T_{*|p^\ell}(x), p \in P \right\} \text{ are disjoint} \\
 &\Downarrow \\
 &\tilde{\mu}(\tilde{B}(x, \epsilon, \tau)) \leq c^{-1} |P|^{-1}.
 \end{aligned}$$

The key fact implicitly behind the proof is that if we take N large but not too large then either:

- (1) All these tubes are indeed disjoint.
- (2) There is a g_0 very close to g with $\Gamma g_0 A$ is compact so that for $\gamma \in \Lambda(p^j), \gamma' \in \Lambda(q^k)$ with $p, q \leq N$ and $j, k \leq \ell$

$$\begin{aligned}
 &\tilde{B}(\gamma g, \epsilon, \tau) \cap \tilde{B}(\gamma' g, \epsilon, \tau) \neq \emptyset \\
 &\Downarrow
 \end{aligned}$$

$$\gamma g_0 = \gamma' g_0 a(t) \text{ for some } |t| \leq 4\tau.$$

- (3) There is a g_0 very close to g so that the trajectory of Γg_0 under A is divergent (i.e., $x_0 a(t) \rightarrow \infty$ as $|t| \rightarrow \infty$) so that the same statement as in (2) holds.

If (1) holds there is nothing left to do but rejoice.

If (2) holds things are slightly more delicate. To any periodic A -orbit one can attach a real quadratic number field: since $G_{\mathbb{Z}} g_0 A$ is compact $G_{\mathbb{Z}} \cap g_0 A g_0^{-1} \neq \emptyset$. The eigenvalues of elements of $G_{\mathbb{Z}} \cap g_0 A g_0^{-1} \neq \emptyset$ are integers (in fact units) in a single real quadratic number field $\mathbb{Q}(\sqrt{D})$, and one can bound D in terms of N .

It turns out that in this case

$$\left\{ B(y, \epsilon, \tau) : y \in T_{*|p^\ell}(x), p \text{ prime} \leq N \text{ inert in } \mathbb{Q}(\sqrt{D}) \right\}$$

is disjoint, and one needs to estimate how many primes $\leq N$ are indeed inert in $\mathbb{Q}(\sqrt{D})$ (equivalently, $\left(\frac{D}{p}\right) = -1$).

If (3) holds, i.e., x is close to a divergent trajectory, the trajectory of x will also spend much time outside Ω — more precisely the maximal function $M(1_{\Omega^c}, x) > \beta$ for a fixed β which does not depend on Ω .

We now turn to actually implementing the above strategy. Recall from §5 that for any $\alpha \in G_{\mathbb{Q}}$, there corresponds a unique primitive

integer matrix $\tilde{\alpha}$, and we shall often implicitly identify between α and $\tilde{\alpha}$.

The key observation for carrying out this strategy is contained in the following simple lemma:

LEMMA 7.3 ([4, Lemma 4.1]). *Let Ω be a compact subset of $G_{\mathbb{R}}$, and let $\alpha, \beta \in G_{\mathbb{Q}}$ and $g \in \Omega$ satisfy*

$$(7.3) \quad \beta g, \alpha g \in g\tilde{B}(\epsilon, \tau) \quad \text{for } \epsilon \ll_{\Omega, t} [d(\alpha)d(\beta)]^{-1/2}.$$

Then α, β commute.

PROOF. Throughout this proof C, C_1, \dots will denote constants depending only on τ and Ω . Define $t_{\alpha}, t_{\beta} \in (-\tau, \tau)$ by

$$\begin{aligned} \alpha g &\in ga(t_{\alpha})\tilde{B}(1, 0) \\ \beta g &\in ga(t_{\beta})\tilde{B}(1, 0). \end{aligned}$$

Then in fact $\alpha g \in ga(t_{\alpha})\tilde{B}(C\epsilon, 0)$, and similarly for β .

$$\begin{aligned} g^{-1}\alpha\beta g &\in a(t_{\alpha} + t_{\beta})\tilde{B}(C_1\epsilon, C_1\epsilon^2) \\ g^{-1}\beta\alpha g &\in a(t_{\alpha} + t_{\beta})\tilde{B}(C_1\epsilon, C_1\epsilon^2). \end{aligned}$$

In other words, since $g^{-1}\tilde{B}(C_1\epsilon, C_1\epsilon^2)g \subset \tilde{B}(C_2\epsilon, C_2\epsilon)$

$$\alpha\beta \in \beta\alpha\tilde{B}(C_3\epsilon, C_3\epsilon).$$

By (7.3), both α and β are in some fixed compact region in $G_{\mathbb{R}}$. It follows that both $\tilde{\alpha}\tilde{\beta}$ and $\tilde{\beta}\tilde{\alpha}$ are integer matrices with all entries $\leq C_4d(\alpha\beta)^{1/2}$ in absolute value. Since integers are at least one apart this implies that if $\epsilon \leq C_5d(\alpha\beta)^{-\frac{1}{2}}$ in fact $\alpha\beta = \beta\alpha$, i.e., α, β commute. \square

In $\mathrm{SL}(2, \mathbb{R})$ there are two types of commutative one dimensional subgroups: tori and unipotent subgroups. The following helps distinguish between the two:

LEMMA 7.4. *Let Ω be a compact subset of $G_{\mathbb{R}}$, $\alpha \in G_{\mathbb{Q}}$ and $g \in \Omega$ satisfy*

$$\alpha g \in g\tilde{B}(\epsilon, \tau) \quad \text{for } \epsilon \ll_{\Omega, \tau} d(\alpha)^{-\frac{1}{2}}.$$

Then α is \mathbb{R} -diagonalizable.

PROOF. A straightforward calculation shows that

$$\tilde{B}(\epsilon, \tau) \cap \{g \in \mathrm{PGL}(2, \mathbb{R}) : g \text{ is not } \mathbb{R}\text{-split}\} \subset \tilde{B}(\epsilon, C\epsilon).$$

Indeed, this follows easily from estimating from below the trace of an arbitrary element $h \in \tilde{B}(\epsilon, \tau) \setminus \tilde{B}(\epsilon, C\epsilon)$ and showing it is $\geq 2\det(h)^{1/2}$; hence h is \mathbb{R} -split.

Therefore if α is not \mathbb{R} -diagonalizable,

$$\alpha \in g\tilde{B}(\epsilon, C\epsilon)g^{-1} \subset \tilde{B}(C_1\epsilon, C_1\epsilon).$$

In other words, if $\tilde{\alpha} = (a_{ij})$ the integer entries of $\tilde{\alpha}$ satisfy the inequalities

$$|a_{11} - a_{22}|, |a_{12}|, |a_{21}| \leq C_2 \epsilon d(\alpha)^{1/2} < 1;$$

hence α is the identity (in particular, \mathbb{R} -diagonalizable) in contradiction to our assumption. \square

Slightly less simple, but of the same general flavor is the following which is only needed because $G_{\mathbb{Z}}$ is not cocompact; it is not relevant e.g., for lattices arising from quaternion division algebras over \mathbb{Q} .

LEMMA 7.5. *Let Ω be a compact subset of $G_{\mathbb{R}}$, $\eta \in (0, \frac{1}{2})$, $\alpha \in G_{\mathbb{Q}}$ and $g \in \Omega$ satisfy*

$$(7.4) \quad \alpha g \in g\tilde{B}(\epsilon, \tau) \quad \text{for } \epsilon \ll_{\Omega, \tau} d(\alpha)^{-1-\eta}$$

with $\tilde{\alpha} \neq I$ diagonalizable over \mathbb{Q} . Then

$$M(1_{\tilde{\pi}(\Omega)\mathbb{C}}, \tilde{\pi}(g)) \geq \eta.$$

PROOF. In this proof we let c_1, c_2, \dots denote constants depending only on τ and Ω and C, C_1, \dots denote constants which are either absolute or depend only on τ .

As before, (7.4) implies in particular that α is in some fixed compact subset of $G_{\mathbb{R}}$; hence all entries of $\tilde{\alpha}$ are integers $< c_1 d(\alpha)^{1/2}$. Let v_1, v_2 be the two eigenvalues of $\tilde{\alpha}$ with $|v_1| > |v_2|$ and let z_1, z_2 be the corresponding eigenvectors. By the assumption that $\tilde{\alpha}$ is diagonalizable over \mathbb{Q} , v_1, v_2 are integers and z_1, z_2 can be chosen to be integral and primitive (i.e., the coefficients of each eigenvector are relatively prime integers). Note that it follows from the bounds on the coefficients of $\tilde{\alpha}$ that $v_1, v_2 < c_1 d(\alpha)^{1/2}$ and

$$(7.5) \quad c_2 d(\alpha)^{-\frac{1}{2}} < |t_{\alpha}| < c_3$$

with c_2, c_3 appropriate functions of c_1 . Like any 2×2 degenerate integer matrix, $\tilde{\alpha} - v_2 I$ has the form $\begin{pmatrix} rp & rq \\ sp & sq \end{pmatrix}$ with r, p, s, q integers, and hence z_1, z_2 , which are also eigenfunctions of $\tilde{\alpha} - v_2 I$, are equal up to sign to $(s, -r), (p, q)$.

$$\|z_1\| \cdot \|z_2\| < 2c_1 d(\alpha)^{1/2}.$$

Since z_1 and z_2 are integer vectors, it follows that $\|z_1\|, \|z_2\| \leq 2c_1 d(\alpha)^{1/2}$.⁸

From (7.4), we see that $\alpha^g = g^{-1}\alpha g \in \tilde{B}(\epsilon, \tau)$. Since $\text{tr}(\tilde{\alpha}) = 2 \cosh(t_{\alpha}) d(\alpha)^{1/2}$

$$\begin{aligned} \alpha^g &\in \tilde{B}(\epsilon, \tau) \cap \left\{ h : \text{tr}(h) = 2 \cosh(t_{\alpha}) \det(h)^{1/2} \right\} \\ &\subset a(t_{\alpha}) \tilde{B}(C\epsilon, C\epsilon^2) \cup a(-t_{\alpha}) \tilde{B}(C\epsilon, C\epsilon^2). \end{aligned}$$

⁸By being slightly more careful and working with the smaller of the two eigenvectors (which has norm $\ll d(\alpha)^{1/4}$) one can get a slightly sharper lemma.

Without loss of generality, $\alpha^g \in a(-t_\alpha)\tilde{B}(C\epsilon, C\epsilon^2)$; otherwise replace α by α^{-1} which also satisfies (7.4) and which has the same value of t_α .

There is some constant C_1 so that as long as

$$(7.6) \quad \sum_{j=0}^n e^{2jt_\alpha} < (C_1\epsilon)^{-1}$$

one has

$$\tilde{B}(Ce^{2nt_\alpha}\epsilon, C\epsilon^2) \dots \tilde{B}(Ce^{2t_\alpha}\epsilon, C\epsilon^2)\tilde{B}(C\epsilon, C\epsilon^2) \subset \tilde{B}(1, 1).$$

Let n_α be the largest integer for which

$$(7.7) \quad e^{2n_\alpha t_\alpha} (C_2\epsilon)^{-1} t_\alpha.$$

Since $\sum_{j=0}^n e^{2jt_\alpha} < e^{2(n+1)t_\alpha}/t_\alpha$, the inequality (7.6) is satisfied for all $n \leq n_\alpha$. Also, since $\tilde{B}(\epsilon_1, \epsilon_2)a(t) \subset a(t)\tilde{B}(e^{2|t|}\epsilon_1, \epsilon_2)$ we see that for any $n \leq n_\alpha$

$$(7.8) \quad \begin{aligned} \alpha^n g &\in ga(-nt_\alpha)\tilde{B}(Ce^{2nt_\alpha}\epsilon, C\epsilon^2) \dots \tilde{B}(Ce^{2t_\alpha}\epsilon, C\epsilon^2)\tilde{B}(C\epsilon, C\epsilon^2) \\ &\subset ga(-nt_\alpha)g\tilde{B}(c_6, c_6). \end{aligned}$$

By Mahler's criterion for compactness in $G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$ applied to the compact set $\tilde{\pi}(\Omega)$, there is a $\delta > 0$ so that for every $g' \in G_{\mathbb{R}}$ with $\tilde{\pi}(g') \in \tilde{\pi}(\Omega)$ we have that

$$\|zg'\| > \delta |\det(g')|^{1/2} \quad \text{for every } z \in \mathbb{Z}^2.$$

But for $n \leq n_\alpha$, by (7.8),

$$\nu_1^n z_1 g = z_1 \tilde{\alpha}^n g \in d(\alpha)^{n/2} z_1 ga(-nt_\alpha)\tilde{B}(c_6, c_6),$$

and so for $nt_\alpha \leq t < (n+1)t_\alpha$ (with $n \leq n_\alpha$)

$$\begin{aligned} \|z_1 ga(-t)\| &= \nu_1^n d(\alpha)^{-n/2} \|z_1 gha(-t + nt_\alpha)\| \quad \text{for some } h \in \tilde{B}(c_6, c_6) \\ &\leq c_7 e^{-nt_\alpha} \|z_1\| |\det(g)| \\ &\leq c_8 e^{-nt_\alpha} d(\alpha)^{1/2} |\det(g)|^{1/2}. \end{aligned}$$

Using (7.6) and (7.5) we deduce that for all n in the interval $[(1-\eta)n_\alpha, n_\alpha]$ and $t \in [nt_\alpha, (n+1)t_\alpha]$ we have that

$$\begin{aligned} \|z_1 ga(-t)\| &\leq c_8 [e^{-2n_\alpha t_\alpha}]^{n/n_\alpha} d(\alpha)^{1/2} |\det(g)|^{1/2} \\ &\leq c_{10} \epsilon^{1-\eta} d(\alpha)^{-1+\eta/2} \end{aligned}$$

which is $\leq \delta$ as long as $\epsilon < c_{11} d(\alpha)^{-1-\eta}$ (for an appropriate c_{11}) for all $\eta \in (0, \frac{1}{2})$. It follows that $\tilde{\pi}(ga(-t)) \notin \tilde{\Omega}$ for all $t \in [(1-\eta)n_\alpha t_\alpha, n_\alpha t_\alpha]$ and

$$M(1_{\tilde{\Omega}^c}, \tilde{\pi}(g)) \geq \eta.$$

□

In order to show their sufficiently many primes $\leq N$ inert in $\mathbb{Q}(\sqrt{D})$ for case (2) we will use the following result [4, Theorem 5.1]:

THEOREM 7.6. *For any $\epsilon > 0$ there is a $\varrho > 0$ so that for every large enough integer D which is not a perfect square, and $N \geq D^{1/4+\epsilon}$, one has that the set P of primes $N^\varrho \leq p \leq N$ with $\left(\frac{D}{p}\right) = -1$ satisfies*

$$\sum_{p \in P} \frac{1}{p} > \frac{1}{2} - \epsilon.$$

In particular, the number of primes $\leq N$ with $\left(\frac{D}{p}\right) = -1$ is $\gg N^{\varrho'}$ with ϱ' some constant depending on ϵ , which is what we will use in this paper. We remark that in [4] we use Theorem 7.6 much more efficiently⁹ but there seems little point for doing this here in this mostly expository paper. We also remark that in order to work with $N \sim D^{1/4+\epsilon}$ in Theorem 7.6, one needs to use highly nontrivial estimates on character sums due to Burgess [7]; if one is willing to take e.g., $N \geq D^{1/2}$ (which is fine for our purposes) these estimates become trivial.

We now, finally, are in a position to prove the main theorem of this section:

PROOF OF THEOREM 7.1. Let $M = (c_1\epsilon)^{-1/(\ell+2)}$ and $N = M^\ell$ with c_1 some constant depending on Ω and τ to be determined later. Fix some compact $\Omega' \subset G_{\mathbb{R}}$ projecting to Ω and take $x = \tilde{\pi}(g)$ with $x \in \Omega, g \in \Omega'$.

Consider now the collection of tubes

$$\mathcal{V} = \left\{ \tilde{B}(y, \epsilon, \tau) : y \in T_{*|p^\ell}(x), p \text{ prime } \leq M \right\}.$$

If all are disjoint, then from (7.1) we get that $\tilde{\mu}(\tilde{B}(x, \epsilon, \tau)) \ll M^{-1} \log M$, which is $\ll \epsilon^\alpha$ for any $\alpha < (\ell(\ell+2))^{-1}$.

Therefore we may assume that there are two distinct $\gamma_1, \gamma_2 \in G_{\mathbb{Q}}$ with $d(\gamma_1), d(\gamma_2) < N^{\ell/2}$ so that

$$\gamma_1 g \tilde{B}(\epsilon, \tau) \cap \gamma_2 g \tilde{B}(\epsilon, \tau) \neq \emptyset.$$

There is some constant C depending only on τ so that $\tilde{B}(\epsilon, \tau)^2 \subset \tilde{B}(C\epsilon, C\tau)$, and so the above equation implies that

$$(7.9) \quad \gamma g \in g \tilde{B}(C\epsilon, C\tau)$$

for $\gamma := \gamma_2^{-1}\gamma_1$; note that $d(\gamma) \mid d(\gamma_1)d(\gamma_2) \leq M^{2\ell}$.

By Lemma 7.4, γ is diagonalizable over \mathbb{R} . There are now two cases, depending on whether γ is diagonalizable over \mathbb{Q} or not.

If γ is diagonalizable over \mathbb{Q} then

$$\epsilon = c_1^{-1} M^{-\ell(\ell+2)} \leq c_1^{-1} d(\gamma)^{-3/2},$$

⁹In particular, in the way Theorem 7.6 is applied in [4] the value of α is immaterial.

and so as long as c_1 was chosen sufficiently large depending only on Ω and τ , by Lemma 7.5 we have that $M(1_{\Omega^c}, x) \geq \frac{1}{2}$. We note that this case arises when x is very near a divergent geodesic in a sense which can be made precise but is not very important for our purposes.

There remains the case of γ diagonalizable over \mathbb{R} but not over \mathbb{Q} .

Since $d(\gamma) \leq N^\ell$ and $\epsilon \leq c_1^{-1} N^{-\ell} M^{-\ell} = c_1^{-1} M^{-\ell(\ell+1)}$, it follows from Lemma 7.3 that for an appropriate choice of c_1 (depending on Ω and τ), for any other two $\beta_1, \beta_2 \in G_{\mathbb{Q}}$ with $d(\beta_1), d(\beta_2) < N^\ell$ if

$$\beta_1 g \tilde{B}(\epsilon, \tau) \cap \beta_2 g \tilde{B}(\epsilon, \tau) \neq \emptyset$$

(hence $\beta_2^{-1} \beta_1 g \in g \tilde{B}(C\epsilon, C\tau)$) then $\beta_2^{-1} \beta_1$ and γ commute. The eigenvalues of $\tilde{\gamma}$ as well as of any integer matrix commuting with $\tilde{\gamma}$ are in the field $\mathbb{Q}(\sqrt{D})$ for

$$D = (\text{tr } \tilde{\gamma})^2 - 4 \det(\tilde{\gamma}) \ll_{\Omega, \tau} d(\tilde{\gamma}) < M^{2\ell} = N^2.$$

Furthermore, if $\tilde{\beta}$ is a primitive integer matrix commuting with $\tilde{\gamma}$ and if ν is an eigenvalue of $\tilde{\beta}$ then $\nu \notin \mathbb{Q}$ and $\det(\tilde{\beta}) = N(\nu)$ with $N(\nu)$ denoting the norm of ν in the field $\mathbb{Q}(\sqrt{D})$. This implies that $\det(\tilde{\beta})$ is divisible by some prime p which is split in $\mathbb{Q}(\sqrt{D})$, i.e., $(\frac{p}{D}) = 1$. We conclude that the collection of tubes

$$\mathcal{V}_P := \left\{ \tilde{B}(y, \epsilon, \tau) : y \in T_{*|p^\ell}(x), p \in P \right\}$$

for $P = \{p \leq N \text{ prime} : (\frac{p}{D}) = -1\}$ is disjoint, and hence by (7.2)

$$\tilde{\mu}(\tilde{B}(x, \epsilon, \tau)) \leq c^{-1} |P|^{-1}$$

for c the constant in the definition of ℓ -uniformly adelicly conservative applied to μ . By Theorem 7.6 we get $|P| \geq N^{\varrho'}$ for some fixed ϱ' and so

$$\tilde{\mu}(\tilde{B}(x, \epsilon, \tau)) \ll \epsilon^{\varrho' / (\ell+2)}.$$

□

8. Proof that Theorem 7.1 implies Theorem 6.1

In this section we explain how the bound on the volume of small tubes in Theorem 7.1 translates to an estimate on the entropy.

Though this is not essential for our purposes we would like to point out that the entropy of μ with respect to the flow $a(t)$ (where as we have already done earlier we implicitly identify $a(t) \in G_{\mathbb{R}}$ with the element $(a(t), e, e, \dots) \in G_{\mathbb{A}}$), is equal to the entropy of $\tilde{\mu} = \pi_*(\mu)$ on $\pi(X) = G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$ with respect to the flow $a(t)$ on $\pi(X)$. Indeed, π commutes with the respective actions, the fibers are a compact group (namely K_f), and the flow acts on these fibers isometrically. Therefore $\pi : X \rightarrow G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$ is an isometric extension of flows and such isometric extensions preserve entropy — indeed, preserve the set of values of the entropy of the ergodic components.

Thus though it makes no difference to the argument it is enough to relate entropy and volume of small tubes for the action of $a(t)$ on $\pi(X)$.

We recall some basic definitions and facts regarding measure preserving systems and their entropy (which is also known as the Kolmogorov-Sinai entropy): We first recall the ergodic decomposition theorem. Any probability measure ν on a (Lebesgue) measure space Y invariant under a measurable map T (or a flow $a(t)$) can be written as the average of ergodic T invariant probability measures. This decomposition can be explicitly given in terms of the conditional measures $\nu_x^\mathcal{E}$ where \mathcal{E} is the sigma algebra of invariant Borel sets in Y ¹⁰: by definition of conditional measures $\nu = \int \nu_x^\mathcal{E} d\nu(x)$, the measures $\nu_x^\mathcal{E}$ are easily seen to be T -invariant, and using the Birkhoff pointwise ergodic theorem one shows that they are in fact ergodic. For any finite partition \mathcal{P} we will let $h_T(\nu, \mathcal{P})$ denote the T, ν - entropy of \mathcal{P}

$$h_T(\nu, \mathcal{P}) = \lim_{n \rightarrow \infty} H_\nu \left(\bigvee_{i=0}^n T^{-i} \mathcal{P} \right) \quad H_\nu(\mathcal{Q}) = - \sum_{\text{atoms of } \mathcal{Q}} \nu(A) \log \nu(A)$$

and $h_T(\nu) = \sup_{\mathcal{P}} h_T(\nu, \mathcal{P})$. We recall that for any T invariant measure ν , its entropy is the average of the entropies of its ergodic components:

$$h_T(\nu) = \int h_T(\nu_x^\mathcal{E}) d\nu(x).$$

For many purposes, the entropies $h_T(\nu_x^\mathcal{E})$ of the ergodic components of ν are more relevant than $h_T(\nu)$. For example, the Shannon McMillan Breiman theorem states that ν -a.s.

$$(8.1) \quad h_T(\nu_x^\mathcal{E}, \mathcal{P}) = - \lim_{n \rightarrow \infty} \log \frac{1}{n} \nu \left([x]_{\bigvee_{i=0}^{n-1} T^{-i} \mathcal{P}} \right).$$

The relation between entropy and volumes of tubes is well-known and fairly standard. If we were dealing with a compact quotient of $\mathrm{PGL}(2, \mathbb{R})$ then this relation would have been completely standard: e.g., it follows easily from the Brin-Katok theorem [6] (see also [29] for a survey of several important related results) that

$$h_{a(t)}(\tilde{\mu}_x^\mathcal{E}) = \lim_{\epsilon \rightarrow 0} \frac{\log \tilde{\mu}(\tilde{T}(x, \epsilon))}{\log \epsilon} \quad \tilde{\mu}\text{-a.e.}$$

There seems to be no convenient reference for this fact in the non-compact case, particularly when $\tilde{\mu}$ is not ergodic, and therefore we will give a direct proof of the implication Theorem 7.1 \implies Theorem 6.1.

We will need the following slightly stronger form of the Shannon-McMillan-Breiman theorem (8.1):

¹⁰We gloss over one slightly delicate technical point: this sigma algebra is almost never countably generated; however since Y is a nice measure space there is a countably generated sigma algebra \mathcal{E}' whose sets agree with \mathcal{E} up to measure zero.

PROPOSITION 8.1. *Let ν be a T -invariant probability measure on a space Y . For any set B of positive ν measure,*

$$h_T(\nu_x^\mathcal{E}, \mathcal{P}) = - \lim_{n \rightarrow \infty} \frac{1}{n} \log \nu \left([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}} \cap B \right) \\ \text{for } \nu\text{-almost every } x \in B.$$

PROOF. Clearly $\nu([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}} \cap B) \leq \nu([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}})$ so by (8.1)

$$h_T(\nu_x^\mathcal{E}, \mathcal{P}) \leq - \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \nu([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}} \cap B).$$

Let $\nu_B = \frac{1}{\nu(B)}\nu|_B$. Consider the function

$$\psi_n(x) = \frac{\nu([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}})}{\nu_B([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}})}.$$

Then $\int_Y \psi_n(x) d\nu_B(x) = 1$ and hence for any $\epsilon > 0$

$$\nu_B \{ \psi_n > e^{\epsilon n} \} < e^{-\epsilon n}.$$

By Borel-Cantelli, for ν_B -almost every x , we have that $\psi_n < e^{\epsilon n}$ for all n large enough, and so for ν a.e. $x \in B$

$$\begin{aligned} \nu([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}} \cap B) &= \nu(B) \nu_B([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}}) \\ &= \nu(B) \nu([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}}) \psi_n(x)^{-1} \\ &\geq \nu(B) \nu([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}}) e^{-\epsilon n} \quad \text{for } n \text{ large,} \end{aligned}$$

and

$$\begin{aligned} - \lim_{n \rightarrow \infty} \frac{1}{n} \log \nu([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}} \cap B) &\leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \nu([x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}}) - n\epsilon \\ &= h_T(\nu_x^\mathcal{E}, \mathcal{P}) - n\epsilon. \end{aligned}$$

□

PROOF OF THEOREM 7.1 \implies THEOREM 6.1.

The proof will consist of three stages. Let $\epsilon > 0$ be arbitrary. For convenience we set $T : \pi(X) \rightarrow \pi(X)$ to be the map $x \mapsto xa(-1)$.

Step 1: Construction of a partition \mathcal{P} .

Let Ω be a compact subset of $\pi(X)$ with $\tilde{\mu}(\Omega a^\mathbb{G}) < \epsilon$. Let $B_0 \subset G_\mathbb{R}$ be an open neighborhood of the identity sufficiently small so that for every $x = G_\mathbb{Z}g \in \Omega$ the collection of sets $\{\gamma g \in B_0\}_{\gamma \in G_\mathbb{Z}}$ is disjoint.

Find $B \subset B_0$ a symmetric neighborhood of the identity (i.e., $B^{-1} = B$) such that

$$(B \cup a(1)Ba(-1) \cup a(-1)Ba(1))^3 \subset B_0.$$

We now take \mathcal{P} to be a finite measurable partition of $\pi(X)$ so that for every $x \in \Omega$, $[x]_\mathcal{P} \subset x \in B$.

Step 2: Understanding the atoms $[x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}}$.

Set for $x \in \pi(X)$ and $n \in \mathbb{N}$

$$S_a^b(\Omega, x) = \sum_{i=a}^{b-1} 1_{\Omega}(T^i x).$$

We shall call any set $V \subset G_{\mathbb{R}}$ satisfying

$$V \subset \bigcap_{i=0}^{n-1} g_i a(i) B a(-i) \quad \text{for some choice of } g_i \in G_{\mathbb{R}}$$

an n, B -ball. A set $V \subset G_{\mathbb{Z}} \backslash G_{\mathbb{R}}$ will be said to be an n, B -ball if it is the projection of a n, B -ball in $G_{\mathbb{R}}$. Take C to be such that B can be covered by C left translates of $a(1)Ba(-1)$. We claim that for every $x \in \Omega$, $[x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}}$ is a union of $C^{S_0^n(\Omega, x, n)}$ n, B -balls.

The proof of this claim is by induction: the case $n = 1$ is obvious. Suppose $[x]_{\bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}}$ is a union of at most $C^{S_0^n(\Omega, x)}$ n, B -balls, say V_1, \dots, V_r . Then if $T^n x \in \Omega^{\mathbb{G}}$, by definition of C we can cover each V_i by C $n+1, B$ -balls, and we are done, and if $T^n x \in \Omega$ we have that for every i the sets $V_i \cap [x]_{T^{-n}\mathcal{P}}$ are $n+1, B$ balls and together these $\leq C^{S_0^n(\Omega, x, n)} = C^{S_0^n(\Omega, x, n+1)}$ balls cover $[x]_{\bigvee_{i=0}^n T^{-i}\mathcal{P}}$.

Finally we remark that any n, B -ball in $\pi(X)$ is contained in a set of the form

$$\{yu_-(s_-)a(t)u_+(s_+) : |s_-| < c, |t| < c \text{ and } |s_+| < ce^{-2n}\}.$$

Similarly, one shows that for every $x \in \Omega$ the atom $[x]_{\bigvee_{i=-n}^{-1} T^{-i}\mathcal{P}}$ can be covered by at most $C^{S_{-n}^0(\Omega, x)}$ sets of the form

$$\{yu_-(s_-)a(t)u_+(s_+) : |s_-| < ce^{-2n}, |t| < c \text{ and } |s_+| < c\},$$

and combining these two facts we get that for every $x \in \Omega$ the atom $[x]_{\bigvee_{i=-n}^{-1} T^{-i}\mathcal{P}}$ can be covered by at most $C^{S_{-n}^n(\Omega, x)}$ tubes $\tilde{V}_c(y, ce^{-2n})$.

Step 3: Conclusion.

Let $D = \{x : M(1_{\Omega^{\mathbb{G}}}, x) > \beta\} \cup \Omega^{\mathbb{G}}$; by the maximal inequality

$$\tilde{\mu}(D) \leq (1 + \beta^{-1})\tilde{\mu}(\Omega^{\mathbb{G}}) \leq (1 + \beta^{-1})\epsilon.$$

From Theorem 7.1 it follows that unless $V_{0.1}(x, 0.1\delta)$ is contained in D , $\tilde{\mu}(V_{0.1}(x, 0.1\delta)) \ll_{\tilde{\mu}, \Omega} \delta^\alpha$; indeed if $y \in V_{0.1}(x, 0.1\delta) \cap D^{\mathbb{G}}$ then $V_{0.1}(x, 0.1\delta) \subset V(y, \delta)$ and by Theorem 7.1.(2) it holds that $\tilde{\mu}(V(y, \delta)) \ll_{\tilde{\mu}, \Omega} \delta^\alpha$. In particular, for every $x \in \pi(X)$,

$$(8.2) \quad \tilde{\mu}(V_{0.1}(x, \delta) \setminus D) \ll_{\tilde{\mu}, \Omega} \delta^\alpha.$$

Now for \mathcal{P} as in Step 1, the atom $[x]_{\bigvee_{i=-n}^n T^{-i}\mathcal{P}}$ can be covered by at most $C^{S_{-n}^n(\Omega, x)}$ tubes $\tilde{V}_c(y, ce^{-2n})$, and so by (8.2) for every x

$$\tilde{\mu}([x]_{\bigvee_{i=-n}^{-1} T^{-i}\mathcal{P}} \setminus D) \ll_{\tilde{\mu}, \Omega} C^{S_{-n}^n(\Omega, x)} e^{-2\alpha n}.$$

By the Birkhoff pointwise ergodic theorem $\frac{1}{n}S_{-n}^n(\Omega, x) \rightarrow 2\tilde{\mu}(\Omega^{\mathbb{G}}) < 2\epsilon$ a.s. so applying Proposition 8.1 for a.e. $x \in D^{\mathbb{G}}$

$$h_T(\tilde{\mu}_x^{\mathcal{E}}, \mathcal{P}) = - \lim_{n \rightarrow \infty} \frac{1}{n} \log \tilde{\mu}([x]_{\bigvee_{i=-n}^n T^{-i}\mathcal{P}} \setminus D) \geq 2\alpha - 2 \log(C)\epsilon.$$

Since the measure of D is arbitrarily small, we are done. \square

9. Open problems

The geodesic flow on the sphere S^2 is very far from being ergodic — it is, in fact, completely periodic — and indeed the quantum ergodicity theorem of Šnirel'man, Colin de Verdière and Zelditch (let alone quantum **unique** ergodicity) fails on this surface. The reason for this is simple: the spectrum of the Laplacian on the sphere is highly degenerate (see [31] for a related discussion).

However, any quaternion algebra over \mathbb{Q} which does not split over \mathbb{R} , for example the ordinary Hamilton quaternions, acts on the sphere in a natural way. This action again gives rise to a commuting ring of Hecke operators, and as in the case of the hyperbolic plane one can look at joint eigenfunctions ϕ_i of the Laplacian and all Hecke operators. Since this takes care of the problem of large multiplicity of the spectrum, it is plausible to conjecture in analogy to Conjecture 1.

PROBLEM 1 ([3, Conj. 1]). Let ϕ_i be Hecke eigenfunctions on the sphere as above. Then $\tilde{\mu}_i = |\phi_i|^2 d\text{vol}$ converges weak* to the uniform measure.

In [3] S. Böcherer, Sarnak, and R. Schulze-Pillot have shown that similarly to $\Gamma \backslash \mathbb{H}$ this conjecture follows from the Riemann hypothesis for suitable automorphic L-functions. However, since the dynamics of the geodesic flow on the sphere is so vastly different from that on $\Gamma \backslash \mathbb{H}$ there seems to be no analogue of Theorem 1.2 which would make sense in this context.

Another important open problem is controlling the hypothetical escape of mass which might occur for e.g., $\text{SL}(2, \mathbb{Z}) \backslash \mathbb{H}$. Recall that in this case it has been proven in [12] that any arithmetic quantum limit is of the form $\frac{c}{\text{vol}(M)} \text{vol}_{S^*M}$ where $c \in [0, 1]$. By Watson's work [27] it is clear c should be 1, but no unconditional proof of this fact is known.

PROBLEM 2. Shown that $c = 1$.

Here is a simpler variant of this problem, also open, but which may be more amenable to a dynamical approach:

PROBLEM 3. Let $\tilde{\mu}_i = |\phi_i|^2 d\text{vol}_M$ with ϕ_i Hecke Maass forms on $M = \text{SL}(2, \mathbb{Z}) \backslash \mathbb{H}$. Show that for every pair f, g of compactly supported nonnegative continuous functions on M with g not identically zero

$$\frac{\int f(x) d\tilde{\mu}_i(x)}{\int g(x) d\tilde{\mu}_i(x)} \rightarrow \frac{\int f(x) d\text{vol}_M(x)}{\int g(x) d\text{vol}_M(x)} \quad \text{as } i \rightarrow \infty.$$

Acknowledgments. I would like to thank the organizers of the Current Developments In Mathematics 2004 workshop for providing me with this opportunity to present my work. My work on quantum unique ergodicity was initiated by a suggestion of Peter Sarnak, and the numerous discussions I have had with him about these matters have strongly influenced my work. Parts of this paper are based on my joint work with Jean Bourgain, and I would like to thank him for this fruitful and enjoyable collaboration. During my work with Bourgain we have received helpful suggestions regarding the sieve method in general and Theorem 7.6 in particular from Enrico Bombieri, Henryk Iwaniec and Kannan Soundararajan. In my choice of open problems I have been influenced by discussions with Alex Gamburd, Peter Sarnak, Lior Silberman and Akshay Venkatesh. I thank Peter Sarnak and Lior Silberman for their comments on earlier versions of this manuscript.

References

- [1] N. Anantharaman, *The eigenfunctions of the laplacian do not concentrate on sets of small topological entropy*, preprint, 2004.
- [2] A. Barnett, *Asymptotic rate of quantum ergodicity in chaotic euclidean billiards*, preprint, 2004.
- [3] S. Böcherer, P. Sarnak, and R. Schulze-Pillot, *Arithmetic and equidistribution of measures on the sphere*, Comm. Math. Phys. **242**(1-2) (2003), 67–80.
- [4] J. Bourgain and E. Lindenstrauss, *Entropy of quantum limits*, Comm. Math. Phys. **233**(1) (2003), 153–171.
- [5] J. Bourgain and E. Lindenstrauss, *Corrections and additions to ‘Entropy of quantum limits’*, 2004.
- [6] M. Brin and A. Katok, *On local entropy*, in ‘Geometric dynamics’ (Rio de Janeiro, 1981), Lecture Notes in Math., **1007**, 30–38, Springer, Berlin, 1983.
- [7] D.A. Burgess, *On character sums and L-series*, II, Proc. London Math. Soc. (3) **13** (1963), 524–536.
- [8] Y.C. de Verdière, *Ergodicité et fonctions propres du laplacien*, Comm. Math. Phys. **102**(3) (1985), 497–502.
- [9] M. Einsiedler and A. Katok, *Invariant measures on G/Γ for split simple Lie groups G* , Comm. Pure Appl. Math. **56**(8) (2003), 1184–1221, Dedicated to the memory of Jürgen K. Moser.
- [10] D. Jakobson, *Equidistribution of cusp forms on $\mathrm{PSL}_2(\mathbb{Z}) \backslash \mathrm{PSL}_2(\mathbb{R})$* , Ann. Inst. Fourier (Grenoble) **47**(3) (1997), 967–984.
- [11] E. Lindenstrauss, *On quantum unique ergodicity for $\Gamma \backslash \mathbb{H} \times \mathbb{H}$* , Internat. Math. Res. Notices **17** (2001), 913–933.
- [12] E. Lindenstrauss, *Invariant measures and arithmetic quantum unique ergodicity*, to appear in Annals of Math., 2003.
- [13] E. Lindenstrauss and A. Venkatesh, *Existence and Weyl’s law for spherical cusp forms*, in preparation, 2004.
- [14] W.Z. Luo and P. Sarnak, *Quantum ergodicity of eigenfunctions on $\mathrm{PSL}_2(\mathbb{Z}) \backslash \mathbb{H}^2$* , Inst. Hautes Études Sci. Publ. Math. **81** (1995), 207–237.
- [15] W.Z. Luo and P. Sarnak, *Quantum variance for Hecke eigenforms*, preprint, 2004.
- [16] G.A. Margulis and G.M. Tomanov, *Invariant measures for actions of unipotent groups over local fields on homogeneous spaces*, Invent. Math. **116**(1-3) (1994), 347–392.

- [17] M. Ratner, *Raghunathan's conjectures for Cartesian products of real and p -adic Lie groups*, Duke Math. J. **77**(2) (1995), 275–382.
- [18] M. Ratner, *Factors of horocycle flows*, Ergodic Theory Dynam. Systems **2**(3-4) (1982), 465–489.
- [19] M. Ratner, *Rigidity of horocycle flows*, Ann. of Math. (2) **115**(3) (1982), 597–614.
- [20] M. Ratner, *Horocycle flows, joinings and rigidity of products*, Ann. of Math. (2) **118**(2) (1983), 277–313.
- [21] Z. Rudnick and P. Sarnak, *The behaviour of eigenstates of arithmetic hyperbolic manifolds*, Comm. Math. Phys. **161**(1) (1994), 195–213.
- [22] A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. (N.S.) **20** (1956), 47–87.
- [23] C. Series, *The modular surface and continued fractions*, J. London Math. Soc. (2) **31**(1) (1985), 69–80.
- [24] L. Silberman and A. Venkatesh, *On quantum unique ergodicity for locally symmetric spaces I: a micro local lift*, preprint, 2004.
- [25] L. Silberman and A. Venkatesh, *On quantum unique ergodicity for locally symmetric spaces*, II, in preparation, 2004.
- [26] A.I. Šnirel'man, *Ergodic properties of eigenfunctions*, Uspehi Mat. Nauk **29**(6(180)) (1974), 181–182.
- [27] T. Watson, *Rankin triple products and quantum chaos*, Ph.D. thesis, Princeton University, 2001.
- [28] S.A. Wolpert, *The modulus of continuity for $\Gamma_0(m)\backslash\mathbb{H}$ semi-classical limits*, Comm. Math. Phys. **216**(2) (2001), 313–323.
- [29] L.-S. Young, *Entropy in dynamical systems*, in ‘Entropy’, Princeton Ser. Appl. Math., 313–327, Princeton Univ. Press, Princeton, NJ, 2003.
- [30] S. Zelditch, *Uniform distribution of eigenfunctions on compact hyperbolic surfaces*, Duke Math. J. **55**(4) (1987), 919–941.
- [31] S. Zelditch, *Quantum ergodicity on the sphere*, Comm. Math. Phys. **146**(1) (1992), 61–71.

DEPARTMENT OF MATHEMATICS, FINE HALL, WASHINGTON RD., PRINCETON NJ 08544

