From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Mathematics

2015

Two dynamical perspectives on the randomness of the Mobius function

Peckner, Ryan Nathaniel

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01v979v546w

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Two dynamical perspectives on the randomness of the Möbius function

Ryan Peckner

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Mathematics
Adviser: Peter Sarnak

June 2015

© Copyright by Ryan Peckner, 2015.

All Rights Reserved

Abstract

Sarnak's approach to the Möbius randomness heuristic from the standpoint of dynamical systems is studied in two complementary settings. First, the Möbius function is realized in the context of symbolic dynamics, and we prove that its associated square-free factor has a unique measure of maximal entropy. We then show that the Möbius function is linearly disjoint from all zero-entropy translations on homogeneous spaces of connected Lie groups, generalizing work of Vinogradov, Green-Tao, and Bourgain-Sarnak-Ziegler.

Acknowledgements

It would be a gross understatement to say that this work would have been impossible without the constant guidance, support, and insights of my adviser, Peter Sarnak. His rigor, patience, and concern for quality have allowed the last five years to be intellectually challenging and fulfilling in ways far beyond what I'd imagined before coming to Princeton.

Yakov Sinai has had a formative influence on my understanding of dynamical systems, ergodic theory, and probability, and his kindness and willingness to share new ideas have been inspirational. Moreover, his work with Francesco Cellarosi on squarefree numbers paved the way for the first part of this thesis.

Discussions with Mariusz Lemańczyk and Manfred Einsiedler, as well as the input of Benjamin Weiss, provided critical insights for arguments involving entropy, skew-products, and joinings.

A substantial portion of this thesis was completed as I recuperated at home from a broken leg, and the constant care of my dad and stepmom gave me the strength to keep working even when I couldn't walk. Their infinite love and support is what has gotten me to where I am today.

Ben, Etha, Mim, and Martin - you guys are the best friends anyone could have.

To Dad, Sue, J, Bess and Andre.

Contents

Abstract iii

Acknowledgements iv

Introduction **1**

1 Uniqueness of the measure of maximal entropy for the squarefree flow **11**

1.1 Induced transformations 11

1.2 The Squarefree Flow 12

1.3 Failure of the Gibbs property 30

2 Möbius disjointness for homogeneous dynamics **34**

2.0.1 Necessary background 34

2.1 The case $G = SL_2(\mathbb{R})$ 37

2.1.1 $\Gamma\backslash G$ is compact 38

2.1.2 $\Gamma\backslash G$ is noncompact 41

2.2 The general semisimple case 42

2.3 The general case 53

2.3.1 The case of abelian R 67

2.3.2 The case of general R 74

Introduction

One of the most important objects in analytic number theory is the Möbius function, defined for positive integers n by

$$\mu(n) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n \text{ is not squarefree} \\ (-1)^r & \text{if } n = p_1 \cdots p_r \text{ is a product of } r \text{ distinct primes.} \end{cases}$$

It is widely believed that the Möbius function behaves as randomly as possible, due to the chaotic distribution of the prime numbers. The goal of this dissertation, broadly speaking, is to clarify the meaning of this heuristic by providing novel results toward a rigorous description of Möbius randomness.

Sarnak has developed an approach to the idea of Möbius randomness based on the study of certain dynamical systems [S]. Let $\Omega^{(3)} = \{-1, 0, 1\}^{\mathbb{N}}$ with the product topology, and let $\sigma : \Omega^{(3)} \to \Omega^{(3)}$ be the left shift defined by $(\sigma x)_n = x_{n+1}$. $(\Omega^{(3)}, \sigma)$ is a topological dynamical system, that is, a pair (X, S) where X is a compact metric space and $S : X \to X$ is continuous and surjective.

The Möbius function defines a point $\mu = (\mu(1), \mu(2), \mu(3), \dots) \in \Omega^{(3)}$, and we let $\mathcal{M}$ be the closure in $\Omega^{(3)}$ of the orbit $\{\sigma^k \mu : k \geq 0\}$. Then the 'Möbius flow' $\mathcal{M}$ is a closed, σ-invariant subset of $\Omega^{(3)}$, but due to the mysterious behavior of μ one cannot say much else about it. We approach the question of Möbius randomness via an indirect study of the Möbius flow, based on its relations to simpler systems.

In the first part of this thesis, we study the factor of the Möbius flow obtained by ignoring the sign changes in $\mu(n)$, that is, by changing every occurrence of -1 in μ to a $+1$. This yields a nicely structured system whose properties are governed by the distribution of squarefree numbers, and which thereby provides a basic level of randomness in μ. We examine the complexity of this system by proving that it carries a unique invariant probability measure with maximal possible entropy, and characterize its structure in terms of Bernoulli and Kronecker systems.

In the second part, we allow ourselves to consider the sign changes of $\mu(n)$ in the context of Sarnak's Möbius disjointness conjecture, proving that it holds for all zero-entropy translations on homogeneous spaces of real connected Lie groups. This may be viewed as a reflection of an expected topological disjointness of the Möbius flow from all sufficiently simple systems, conditioned on the presence of its squarefree factor. Together with the first part of our work, this gives some insight into the notion that the Möbius flow is "as random as possible": it should have the largest possible relative topological entropy as an extension of the squarefree flow.

We now lay a more detailed groundwork for both parts of this thesis.

The squarefree flow

Let $\Omega^{(2)} = \{0,1\}^{\mathbb{N}}$ and let $\phi : \Omega^{(3)} \to \Omega^{(2)}$ be the squaring map $(x_n) \mapsto (x_n^2)$. Observe that ϕ is continuous, surjective, and intertwines the shift maps on the two spaces (this is an example of a *factor map* between topological dynamical systems). Let $\mathcal{S}$ be the closure in $\Omega^{(2)}$ of the orbit of $\phi(\mu) = (\mu^2(n))_{n\geq 1}$ under σ. Since $\mu^2(n)$ is the indicator function of the set of squarefree numbers, we call $(\mathcal{S}, \sigma)$ the squarefree flow.

Naturally, the dynamical structure of $\mathcal{S}$ is strongly tied to the statistical properties of squarefree numbers. In this connection, Sarnak uses the well-known squarefree sieve (cf. [Tsa]) to produce a certain ergodic measure ν of zero entropy on $\Omega^{(2)}$ whose support is $\mathcal{S}$ (for the definitions of terms from ergodic theory see [Wal]). This allows

him in turn to prove the following: define the *support* of a sequence $x = (x_n) \in \Omega^{(2)}$ to be the set $\mathrm{supp}(x) \subset \mathbb{N}$ of those n such that $x_n = 1$. We call a subset $A \subset \mathbb{N}$ *admissible* if for every prime p the reduction $A \bmod p^2$ is a proper subset of $\mathbb{Z}/p^2\mathbb{Z}$. We then call a sequence $(x_n) \in \Omega^{(2)}$ admissible if its support is an admissible subset of $\mathbb{N}$. As, for every prime p, the support of $(\mu^2(n))_{n\geq 1}$ is missing the residue class $0 \bmod p^2$, we see that μ^2 is admissible in this sense.

Let $\mathcal{A} \subset \Omega^{(2)}$ be the set of all admissible sequences. It's easy to show that $\mathcal{A}$ is a closed, shift invariant subset of $\Omega^{(2)}$. Using the aforementioned measure ν (and in particular the fact that its support is $\mathcal{S}$), Sarnak is able to show that in fact $\mathcal{A} = \mathcal{S}$. Though this may seem surprising at first, it arises quite naturally from the shape of the measure ν, which in turn is constructed from the main term of the squarefree sieve formula.

The ergodicity of ν implies that the subset $\mathcal{A}_1$ of $\mathcal{A}$ consisting of sequences whose support omits only one residue class $\bmod p^2$ for every p has full ν-measure. It's then not hard to show that the map which sends such a sequence to the point in $\mathcal{K} := \prod_p \mathbb{Z}/p^2\mathbb{Z}$ whose p-th coordinate is the residue class omitted by the sequence $\bmod p^2$ intertwines the shift on $\mathcal{A}$ with the translation $T_{(-1,-1,-1,\dots)}$ on $\mathcal{K}$ by the element $(-1,-1,-1,\dots)$ (note that this map is Borel measurable but not continuous). Moreover, the pushforward of ν under this map is precisely the mass one Haar measure m on the compact group $\mathcal{K}$, so we have a factor map of *measure-preserving* dynamical systems $(\mathcal{A}_1, \sigma, \nu) \to (\mathcal{K}, T_{(-1,-1,-1,\dots)}, m)$. Subsequently, Cellarosi and Sinai [CS] used spectral theory techniques to prove that this map is in fact an isomorphism of measure-preserving dynamical systems (it is however very far from being an isomorphism of *topological* dynamical systems).

While this tells the whole story for the measure ν, there are other invariant measures on $\mathcal{A}$ that should be taken into account. Specifically, as stated by Sarnak (and proven here), the system $(\mathcal{A}, \sigma)$ has topological entropy $(6/\pi^2)\log 2$ (it's no coinci-

dence that $6/\pi^2$ is the density of the squarefree numbers in $\mathbb{N}$!). The variational principle (e.g. [Wal] 8.2) states that the topological entropy of a topological dynamical system is the supremum of the measure entropies over all its invariant probability measures, and it is known that any subshift system possesses at least one *measure of maximal entropy*, i.e. an invariant probability measure whose measure entropy equals the topological entropy of the system. Thus there exists such a measure for $\mathcal{A}$.

While (ergodic) measures of maximal entropy always exist for irreducible subshifts (this is not the case for all topological dynamical systems), the number of such measures may be arbitrarily large, even infinite (see [Hay] for examples). In keeping with the general principle in ergodic theory that a scarcity of measures is more meaningful than an abundance of them, it is an important problem to determine when there is only one measure of maximal entropy on a given topological dynamical system.

The aim of the first part of this dissertation is to prove the following.

Theorem. *The squarefree flow $\mathcal{A}$ possesses a unique measure of maximal entropy.*

The essence of our proof is the construction of a Borel measurable map (eq. (1.2.3))

$$\iota : \left(\prod_p \mathbb{Z}/p^2\mathbb{Z} \right) \times \Omega^{(2)} \to \mathcal{A}$$

such that $\iota \circ T = \sigma \circ \iota$, where T is a certain skew-product transformation on $\left(\prod_p \mathbb{Z}/p^2\mathbb{Z} \right) \times \Omega^{(2)}$. This map codifies the idea of constructing a sequence with admissible support by inserting zeros into an arbitrary sequence in $\Omega^{(2)}$ along the residue classes described by an element of $\prod_p \mathbb{Z}/p^2\mathbb{Z}$.

Once this factor map has been constructed, the problem effectively reduces to proving the intrinsic ergodicity of the skew-product transformation T. Measures of maximal entropy for precisely such skew-products have been studied in great detail in [MarNew]. As their results are much more general than the case treated here, we have opted to give a direct explication of their proof in the case of interest to us, where

the base of the skew product is uniquely ergodic with entropy zero. The key ideas of this argument - namely, inducing the skew product to a subset where it decomposes as a direct product, then using Abramov's formula to compare entropies - are entirely borrowed from their work. Our contribution lies in the use of the factor map ι above to compare the complicated system $\mathcal{A}$ with a more transparent skew-product system.

This construction yields an explicit formula for the effect of the unique measure of maximal entropy on the cylinder set corresponding to any word in $\mathcal{A}$, and by using this formula we are able to show in Section 1.3 that it is not a Gibbs measure, in contrast to many familiar classes of intrinsically ergodic systems.

We note that this method also applies to the $\mathscr{B}$-free shifts studied in [ELR], which are defined as follows. Fix an infinite set $\mathscr{B} = \{b_1, b_2, b_3, \dots\} \subset \mathbb{N}^*$ satisfying the following properties:

1. $\forall 1 \leq r < r'$, b_r and $b_{r'}$ are relatively prime
2. $\sum_{r \geq 1} b_r^{-1} < \infty$.

Integers with no factors in $\mathscr{B}$ are called $\mathscr{B}$-free. We define the characteristic function

$$\nu(n) = \begin{cases} 0 & \text{if } n \text{ is not } \mathscr{B}\text{-free} \\ 1 & \text{if } n \text{ is } \mathscr{B}\text{-free.} \end{cases}$$

The $\mathscr{B}$-free flow X is then the closure of the point $(\nu(n))_{n \geq 1} \in \Omega^{(2)} = \{0,1\}^{\mathbb{N}^*}$ under the left shift, and can be identified as the set of sequences the reduction of whose support modulo each b_r is a proper subset of $\mathbb{Z}/b_r\mathbb{Z}$ ([ELR] Cor. 4.2). Thus the squarefree flow is the $\mathscr{B}$-free flow for $\mathscr{B} = \{p^2 : p \text{ prime}\}$, and as no special properties of squares of primes are used in our argument for intrinsic ergodicity aside from their mutual relative primality, the proof carries over to the general case essentially verbatim. In contrast, our argument regarding the Gibbs property is quite specific to squares of primes, and it is unclear whether this may be extended to general sets $\mathcal{B}$.

The preprint [KLW], which was prepared simultaneously and independently from our work, presents a detailed study of invariant measures for $\mathscr{B}$-free systems, including a proof of our main theorem for all such systems. Their approach, while substantially different from ours, is also based on relating an invariant measure on the squarefree flow to the Haar measure on a compact monothetic group and the $(1/2, 1/2)$ Bernoulli measure.

Möbius randomness and linear disjointness

The overall behavior of the Möbius function is captured quantitatively by the summatory function

$$S(N) := \sum_{n \leq N} \mu(n)$$

as $N \to \infty$. Sign fluctuations in $\mu(n)$, resulting from randomness in the distribution of prime numbers, create cancellations in $S(N)$ that reflect specific number-theoretic phenomena. For instance, the prime number theorem is elementarily equivalent to the estimate $S(N) = o(N)$, while the Riemann hypothesis is equivalent to the power-savings $S(N) = O_\epsilon(N^{1/2+\epsilon})$ for all $\epsilon > 0$.

Sarnak has formulated a less quantitative description of Möbius randomness from this standpoint that seeks to measure correlations of $\mu(n)$ with simpler functions, rather than its intrinsic cancellations [S]. This simultaneously gives a rigorous grounding to the well-known "Möbius randomness heuristic" [IK], which states that the sum

$$\sum_{n \leq N} \mu(n) a_n$$

should be asymptotically small whenever a_n is a "reasonable" sequence of complex numbers, and generalizes more precise results along these lines. These include Vino-

gradov's estimate for exponential sums

$$\sum_{n\leq N} \mu(n)\exp(2\pi i\alpha n) \ll_A \frac{N}{\log^A N} \text{ for all } A > 0,$$

and the generalization thereof to nilmanifolds due to Green-Tao [GT].

Sarnak's conjecture on the randomness of the Möbius function is set in the context of deterministic topological dynamical systems. It asserts that if X is a compact metric space and $T : X \to X$ a continuous surjection such that $h_{\text{top}}(X, T) = 0$, then for every $x \in X$ and every continuous function $f \in C(X)$, we have

$$\frac{1}{N}\sum_{n=0}^{N-1} \mu(n) f(T^n x) \to 0 \text{ as } N \to \infty.$$

Thus, the notion of a reasonable sequence of complex numbers in this setting is that of an observable sequence from a topological dynamical system of zero entropy.

This statement deliberately avoids demanding a rate of convergence, as these are currently available only for relatively simple systems. The fundamental nuance in this formulation is that x may be *any* point whatsoever in the space X. If we relax the convergence to merely hold almost everywhere with respect to some T-invariant measure on X, the result follows from the spectral theory of the Koopman operator for the system and Bourgain's version of the ergodic theorem [B2].

This conjecture has been verified for certain rank-one transformations [B1, ALR], skew-products on tori [LS], Kronecker systems and rotations on nilmanifolds [Dav, GT], and horocycle flows on surfaces of constant negative curvature [BSZ], among other cases. The latter two examples, although quite different in nature, are conceptually of the same ilk. Both are homogeneous dynamical systems of zero entropy, meaning that there is a Lie group G (a nilpotent group for the nilmanifold case, and $SL_2(\mathbb{R})$ for horocycle flows) and a lattice $\Gamma \subset G$ such that the dynamics of the system are given by a zero-entropy homeomorphism of the homogeneous space $\Gamma\backslash G$.

The most basic class of such zero-entropy homeomorphisms is obtained by allowing an Ad-unipotent element $u \in G$ to act on $\Gamma\backslash G$ by right translation. The profound results of Ratner on the topological and measure rigidity of such unipotent translations [R1, R2, R3, R4] render them a fertile testing ground for many phenomena in dynamics, and indeed her results are essential for proving the disjointness of $\mu(n)$ from horocycle flows [BSZ].

Our aim in the second part of this dissertation is to use the full generality of Ratner's theorems to prove the disjointness of the Möbius function from all unipotent translations on homogeneous spaces of connected Lie groups.

Theorem. *Let G be a real connected Lie group, $\Gamma \subset G$ a lattice, and $u \in G$ an Ad-unipotent element. Then for every $x \in \Gamma\backslash G$ and every continuous, bounded function f on $\Gamma\backslash G$, we have*

$$\frac{1}{N}\sum_{n\leq N}\mu(n)f(xu^n) \to 0 \text{ as } N \to \infty.$$

This does not quite settle Sarnak's conjecture for all zero-entropy homogeneous dynamics, as we do not account for the more general class of quasi-unipotent affine automorphisms. These are maps on $\Gamma\backslash G$ of the form $T_g \circ \alpha$, where α is an automorphism of G such that $\alpha(\Gamma) = \Gamma$, and $g \in G$ has the property that all eigenvalues of $\mathrm{Ad}\, g \circ \mathrm{d}\alpha$ are of absolute value 1 (T_g denotes right translation by g on $\Gamma\backslash G$). However, we strongly suspect that the ideas we present in proving the disjointness conjecture for unipotent translations should generalize in a fairly straightforward way to these systems.

It should be noted also that while we only prove our result for real connected Lie groups, the statement and method of proof should be roughly similar for the case of p-adic Lie groups and their products, as Ratner rigidity continues to hold in this context [R4].

One may view this result as the statement that the Möbius flow is linearly disjoint

from all deterministic homogeneous dynamical systems of the stated kind, giving a view of its complexity that fundamentally differs from our study of the squarefree flow, due to its incorporation of the sign changes of μ. Extending this statement beyond a linear expression in $\mu(n)$ seems out of reach at present; even to produce significant cancellations in the second-order correlations sum

$$\sum_{n \leq N} \mu(n)\mu(n+1)$$

presents an enormous challenge, and may be as difficult as the twin prime conjecture. The best known bound for this sum is the recent result of Matomäki and Radziwill [MR]: there exists some $\delta > 0$ such that

$$\frac{1}{N} \left| \sum_{n \leq N} \lambda(n)\lambda(n+1) \right| \leq 1 - \delta$$

for all sufficiently large N, where λ is the Liouville function $\lambda(n) = (-1)^{\Omega(n)}$, with $\Omega(n)$ the number of prime divisors of n counted with multiplicity. See also [C] for related results.

In this vein, Sarnak has shown that the Möbius disjointness conjecture follows from the well-known Chowla conjecture [T], which states that if $m > 0$ and $a_1, a_2, \ldots, a_m$ are nonnegative integers, at least one of which is odd, then

$$\sum_{n \leq N} \mu(n+1)^{a_1}\mu(n+2)^{a_2} \cdots \mu(n+m)^{a_m} = o(N).$$

We may interpret the sequence $\mu(n+1)^{a_1}\mu(n+2)^{a_2} \cdots \mu(n+m)^{a_m}$ as an observable sequence in the Möbius flow. An enriched version of the disjointness conjecture, allowing for correlations between such nonlinear expressions in μ and corresponding expressions in a deterministic sequence, is therefore unapproachable until the Chowla conjecture is known.

It is our hope that our result may apply to the study of solutions of polynomial equations in prime numbers, as in [GT]. However, such applications would require a stronger version of our theorem, in which the rate of convergence of the correlated sum is quantified. As mentioned above, no such rates are known in the general setting of Ratner's theorems, though partial results have been obtained in this direction [SU].

Chapter 1

Uniqueness of the measure of maximal entropy for the squarefree flow

Maintaining the notation of the introduction, the goal of this chapter is to prove the following.

Theorem 1.0.1. *The squarefree flow $\mathcal{A}$ possesses a unique measure of maximal entropy.*

1.1 Induced transformations

We first recall some generalities on induced transformations, as they are essential to our argument. Suppose $T : X \to X$ is a Borel isomorphism of a compact metric space, and $Y \subset X$ is a recurrent Borel subset, meaning $Y \subseteq \cup_{n=1}^{\infty} T^{-n}Y$. Let $r : Y \to \mathbb{N}$ be the return-time map, so $r(x) = \min\{n \geq 1 : T^n x \in Y\}$, and define the induced transformation $T_Y : Y \to Y$ by $T_Y x = T^{r(x)}x$. Given a T-invariant Borel probability measure μ on X such that $\mu(Y) > 0$, we may define a T_Y-invariant measure μ_Y on

Y by $\mu_Y(E) = \mu(E \cap Y)/\mu(Y)$. We then have the following well-known facts (cf. [Zwei],[Th]).

1. Suppose μ is a T-invariant Borel measure on X such that $X = \cup_{n=0}^{\infty} T^{-n}Y$ up to a set of μ-measure zero. Then μ_Y is ergodic for T_Y if and only if μ is ergodic for T.

2. We may calculate the entropy of μ_Y as $h_{\mu_Y}(T_Y)\mu(Y) = h_\mu(T)$.

3. Let ν be a T_Y-invariant Borel probability measure on Y. Then there exists a T-invariant Borel probability measure μ on X, concentrated on $\cup_{n=0}^{\infty} T^{-n}Y$, such that $\nu = \mu_Y$.

Fact (2) is Abramov's formula [Ab]. Facts (1) and (3) may be easily proven: (1) follows because for any $n \geq 1$, $\{x \in Y : r(x) = n\} \cap T^{-n}(E \cap Y) = \{r(x) = n\} \cap T^{-n}E$, so that $E \cap Y$ is invariant for T_Y whenever E is invariant for T. For (3), the measure μ can be given explicitly as

$$\mu(E) = \sum_{n \geq 0} \nu(\{r(x) > n\} \cap T^{-n}E).$$

The T-invariance of μ follows by writing

$$\mu(T^{-1}E) = \sum_{n \geq 1} \nu(\{r(x) > n\} \cap T^{-n}E) + \sum_{n \geq 1} \nu(\{r(x) = n\} \cap T^{-n}E)$$

and using the T_Y-invariance of ν. It follows that the assignment $\mu \mapsto \mu_Y$ is a bijection between T-invariant measures on X with $\mu(Y) > 0$ and T_Y-invariant measures on Y.

1.2 The Squarefree Flow

We now turn to our situation. We will be dealing with Borel measurable maps that are not continuous, and so must be somewhat careful in our definitions. Following

[MarNew], if S is a Borel measurable isomorphism of a compact metric space, we define the topological entropy

$$h(S) = \sup\{h_\mu(S) : \mu \text{ is an } S\text{-invariant Borel probability measure}\}. \tag{1.2.1}$$

When S is a homeomorphism, this agrees with the usual definition of $h_{\text{top}}(S)$ by the variational principle. If the sup is achieved by a unique measure, S is called intrinsically ergodic.

Define $G = \prod_p \mathbb{Z}/p^2\mathbb{Z}$, and define the translation function $T_1 : G \to G$ by $T_1(g) = (g_1 - 1, g_2 - 1, \dots)$. Let $G_1 = \{g \in G : g_p \equiv 1 \bmod p^2 \text{ for some } p\}$, and define the transformation $T : G \times \Omega^{(2)} \to G \times \Omega^{(2)}$ by

$$T(g, y) = \begin{cases} (T_1 g, y) & \text{if } g \in G_1 \\ (T_1 g, \sigma y) & \text{if not,} \end{cases}$$

where σ is the left shift on $\Omega^{(2)}$. T is a Borel isomorphism of $G \times \Omega^{(2)}$, but is not continuous since G_1 is open and not closed in G. Observe that we can describe T as a skew-product transformation: define the Borel measurable function $\psi = 1 - \chi_{G_1} : G \to \{0, 1\}$, where χ_{G_1} is the indicator function of G_1. Then

$$T(g, y) = (T_1 g, \sigma^{\psi(g)} y). \tag{1.2.2}$$

Let X be the squarefree flow as defined above. The key to our argument is an "insertion map" $\iota : G \times \Omega^{(2)} \to X$ that allows us to realize X in terms of a more transparent system. Let's describe the motivation for this construction. A sequence $x \in \Omega^{(2)}$ belongs to X if and only if x omits at least one arithmetic progression modulo p^2 for all primes p, i.e. if and only if there exists $a \in \{1, 2, \dots, p^2\}$ such that $x_{a+\ell p^2} = 0$ for any $\ell \geq 0$. Outside of these arithmetic progressions, however,

the entries of x may be chosen with complete freedom. Thus, suppose we begin with an element $g \in G$ and *any* sequence $x' \in \Omega^{(2)}$. To define a sequence in X from this information, we start reading the entries of x' until we reach a position congruent to one of the $g_p \bmod p^2$, where we insert a zero. We then continue to read x' from where we left off until we reach another position congruent to some (likely to be different) $g_q \bmod q^2$, where we again insert a zero, and so on. For example, let's carry out this process with respect to only one prime $p = 2$. Suppose we are given the residue class $3 \pmod{p^2 = 4}$ and the sequence $x' \in \Omega^{(2)}$ shown below. Then modifying x' as just described will yield the indicated sequence x:

$$x' = 01110110101101\cdots$$

$$x = 01\underline{0}110\underline{0}110\underline{0}101\underline{0}1010\underline{0}\cdots$$

where the underlined zeros have been inserted to force x to omit the progression $3 \bmod 4$. To formalize this, observe for instance that the fifth entry of x equals the fourth entry of x', since there is exactly one $n \leq 5$ with $n \equiv 3 \bmod 4$, and $5 - 1 = 4$. At the same time, the eigth entry of x equals the sixth entry of x', since now there are two $n \leq 8$ with $n \equiv 3 \bmod 4$, and $8 - 2 = 6$.

Thus, for each $g \in G$ define the function $\alpha_g : \mathbb{N} \to \mathbb{N}$ by

$$\alpha_g(m) = |\{1 \leq n \leq m : n \equiv g_p \bmod p^2 \text{ for some } p\}|.$$

Now, for $(g, y) \in G \times \Omega^{(2)}$, let $\iota(g, y) \in X$ be the sequence defined by

$$\iota(g,y)_m = \begin{cases} 0 & \text{if } m \equiv g_p \bmod p^2 \text{ for some } p \\ y_{m-\alpha_g(m)} & \text{if not.} \end{cases} \tag{1.2.3}$$

ι is clearly surjective, but not continuous, since one must read the entire infinite

sequence $g \in G$ before deciding whether to insert a zero at a given position m.

Proposition 1.2.1. *ι is Borel measurable, and we have $\iota \circ T = \sigma \circ \iota$.*

We first will prove the equivariance property, then demonstrate the measurability. To prove the desired relation, we must divide into cases dictated by G_1. Thus, suppose $g \in G_1$, and let $y \in \Omega^{(2)}$. Then we have for any $m \geq 1$

$$\iota(T(g,y))_m = \iota(T_1 g, y)_m = \begin{cases} 0 & \text{if } m+1 \equiv g_p \bmod p^2 \text{ for some } p \\ y_{m-\alpha_{T_1 g}(m)} & \text{if not.} \end{cases}$$

Observe that if $\nexists p$ with $m+1 \equiv g_p \bmod p^2$, then since $g \in G_1$, we have

$$\begin{aligned} \alpha_{T_1 g}(m) &= |\{1 \leq n \leq m : n+1 \equiv g_p \bmod p^2 \text{ for some } p\}| \\ &= |\{1 \leq n \leq m : n \equiv g_p \bmod p^2 \text{ for some } p\} \backslash \{1\}| \\ &= \alpha_g(m) - 1. \end{aligned}$$

Therefore, for such m we have

$$y_{m-\alpha_{T_1 g}(m)} = y_{m-\alpha_g(m)+1}.$$

At the same time,

$$\sigma(\iota(g,y))_m = \iota(g,y)_{m+1} = \begin{cases} 0 & \text{if } m+1 \equiv g_p \bmod p^2 \text{ for some } p \\ y_{(m+1)-\alpha_g(m+1)} & \text{if not.} \end{cases}$$

In the second case above, we clearly have $\alpha_g(m+1) = \alpha_g(m)$; therefore $y_{(m+1)-\alpha_g(m+1)} = y_{m-\alpha_g(m)+1} = \iota(T(g,y))_m$. In the first case, it's obvious that $\iota(T(g,y))_m = \sigma(\iota(g,y))_m = 0$; therefore we have equality for all $m \geq 1$ and so $\iota(T(g,y)) = \sigma(\iota(g,y))$.

Now assume $g \notin G_1$. Then

$$\iota(T(g,y))_m = \iota(T_1 g, \sigma y)_m = \begin{cases} 0 & \text{if } m+1 \equiv g_p \bmod p^2 \text{ for some } p \\ y_{m-\alpha_{T_1 g}(m)+1} & \text{if not.} \end{cases}$$

If $m \geq 1$ and $\nexists p$ with $m+1 \equiv g_p \bmod p^2$, we have

$$\begin{aligned} \alpha_{T_1 g}(m) &= |\{1 \leq n \leq m : n+1 \equiv g_p \bmod p^2 \text{ for some } p\}| \\ &= |\{1 \leq n \leq m : n \equiv g_p \bmod p^2 \text{ for some } p\}| \\ &= \alpha_g(m). \end{aligned}$$

Thus in this case, $y_{m-\alpha_{T_1 g}(m)+1} = y_{m-\alpha_g(m)+1}$. On the other hand,

$$\sigma(\iota(g,y))_m = \iota(g,y)_{m+1} = \begin{cases} 0 & \text{if } m+1 \equiv g_p \bmod p^2 \text{ for some } p \\ y_{(m+1)-\alpha_g(m+1)} & \text{if not.} \end{cases}$$

In the second case above, we clearly have $\alpha_g(m+1) = \alpha_g(m)$; therefore $y_{(m+1)-\alpha_g(m+1)} = y_{m-\alpha_g(m)+1} = \iota(T(g,y))_m$. In the first case, it's obvious that $\iota(T(g,y))_m = \sigma(\iota(g,y))_m = 0$; therefore we have equality for all $m \geq 1$ and so $\iota(T(g,y)) = \sigma(\iota(g,y))$. So this holds both for $g \in G_1$ and $g \notin G_1$, proving the desired equivariance.

To prove the measurability of ι, for each $r \geq 1$ define

$$X(r) = \{x \in \Omega^{(2)} : \text{supp}(x) \text{ omits at least one residue class } \bmod p_i^2 \text{ for } i = 1, \cdots, r\}.$$

Note that $X(r) \supset X$ for all $r \geq 1$. For each $r \geq 1$, let $G_r = \prod_{i=1}^r \mathbb{Z}/p_i^2\mathbb{Z}$, and let $\pi_r : G \times \Omega^{(2)} \to G_r \times \Omega^{(2)}$ be the product of projection $G \to G_r$ onto the first r coordinates with the identity on $\Omega^{(2)}$. We define the partial insertion map

$\phi_r : G_r \times \Omega^{(2)} \to X(r)$ as follows. For each $g \in G_r$ define the function $\alpha_g^r : \mathbb{N} \to \mathbb{N}$ by

$$\alpha_g^r(m) = |\{n \leq m : n \equiv g_{p_i} \bmod p_i^2 \text{ for some } i \in \{1, \ldots, r\}\}|.$$

Now, for $(g, y) \in G_r \times \Omega^{(2)}$, let $\phi_r(g, y) \in X(r)$ be the sequence defined by

$$\phi_r(g,y)_m = \begin{cases} 0 & \text{if } m \equiv g_{p_i} \bmod p_i^2 \text{ for some } i \in \{1, \ldots, r\} \\ y_{m-\alpha_g^r(m)} & \text{if not.} \end{cases} \tag{1.2.4}$$

ϕ_r is a surjective map from $G_r \times \Omega^{(2)}$ onto $X(r)$. Define $\iota_r : G \times \Omega^{(2)} \to X(r)$ to be the composition of π_r with the partial insertion map $\phi_r : G_r \times \Omega^{(2)} \to X(r)$; then ι_r is surjective as well.

Proposition 1.2.2. *ι_r is continuous for each $r \geq 1$, and we have $\iota_r \to \iota$ pointwise everywhere on $G \times \Omega^{(2)}$. Therefore, ι is Borel measurable.*

Proof. For a subshift $\Sigma \subset \Omega^{(2)}$ and a word w of length ℓ appearing in Σ (meaning there exists some $x \in \Sigma$ such that $w = x_1 \cdots x_\ell$), we will denote the cylinder set corresponding to w by

$$C_w = \{x \in \Sigma : x_1 \cdots x_\ell = w\}.$$

Such sets provide a base for the topology on Σ.

To show that ι_r is continuous, let $(g, y) \in G \times \Omega^{(2)}$ and let $M \geq 1$. Let $u = y_1 \cdots y_M$ be the first length M subword of y, and consider the subset

$$U = \left(\{g_1\} \times \{g_2\} \times \cdots \times \{g_r\} \times \prod_{k>r} \mathbb{Z}/p_k^2\mathbb{Z} \right) \times C_u \subset G \times \Omega^{(2)}.$$

U is clearly an open subset of $G \times \Omega^{(2)}$, and it's equally clear that the sequences $\iota_r(g, y)$ and $\iota_r(h, z)$ agree to at least M positions for any $(h, z) \in U$. As M was arbitrary, this proves that ι_r is continuous.

We now prove the pointwise everywhere convergence $\iota_r \to \iota$. Let $(g, y) \in G \times \Omega^{(2)}$, and fix a word w of length ℓ appearing in X. We will show that $\iota(g, y) \in C_w$ if and only if $\iota_r(g, y) \in C_w$ for all sufficiently large r (where 'sufficiently large' is taken relative to (g, y) since we only seek pointwise convergence). Observe that there exists an $R \geq 1$ such that

$$\{m \leq \ell : m \equiv g_r \bmod p_r^2 \text{ for some } r\} = \{m \leq \ell : m \equiv g_r \bmod p_r^2 \text{ for some } r \leq R\}.$$

Indeed, we have

$$\{r : g_r \equiv m \bmod p_r^2 \text{ for some } m \leq \ell\} = \coprod_{m \leq \ell} \{r : g_r \equiv m \bmod p_r^2\}$$

so we may choose an r from each nonempty set on the right and take R to be the maximum of these. Then for any $n \leq \ell$ we have

$$\begin{aligned} \alpha_g(n) &= \alpha_g^R(n) \\ &= |\{m \leq n : m \equiv g_r \bmod p_r^2 \text{ for some } r \leq R\}|. \end{aligned}$$

Thus if $S \geq R$ we have for any $n \leq \ell$

$$\begin{aligned} \iota(g, y)_n &= \begin{cases} 0 & \text{if } n \equiv g_r \bmod p_r^2 \text{ for some } r \\ y_{n - \alpha_g(n)} & \text{if not} \end{cases} \\ &= \begin{cases} 0 & \text{if } n \equiv g_r \bmod p_r^2 \text{ for some } r \leq R \\ y_{n - \alpha_g^R(n)} & \text{if not} \end{cases} \\ &= \iota_S(g, y)_n, \end{aligned}$$

and it follows that $\iota_S(g, y) \in C_w$ for $S \geq R$ if and only $\iota(g, y) \in C_w$, proving the

claim. □

For any infinite sequence of integers $I = (i_p) \in \prod_p \{1, \ldots, p^2 - 1\}$ (so i_p is the number of residue classes omitted mod p^2) consider the set

$$\{x \in X : \text{supp}(x) \text{ omits exactly } i_p \text{ residue classes mod } p^2 \ \forall p\}.$$

This set isn't invariant, since there may exist a sequence x in this set and a residue class h mod p^2 for some p such that $\text{supp}(x) \cap h$ mod p^2 is finite; then shifting x finitely many times will cause x to omit an extra residue class mod p^2, removing x from this set. Hence, to get an invariant set, define for each prime p the set of p-persistent sequences in X

$$S_p = \{x \in X : \text{ for all } h \in \mathbb{Z}/p^2\mathbb{Z}, \text{ if } \text{supp}(x) \cap h(p^2) \text{ isn't empty, then it is infinite}\}.$$

Observe that S_p is an invariant subset of X. To see that S_p is Borel, observe first that

$$S_p = \bigcap_{h \in \mathbb{Z}/p^2\mathbb{Z}} \{x \in X : \text{supp}(x) \cap h(p^2) \text{ is empty}\} \amalg \{x \in X : \text{supp}(x) \cap h(p^2) \text{ is infinite}\}. \tag{1.2.5}$$

For each $h \in \mathbb{Z}/p^2\mathbb{Z}$, let $N_h = \{x \in X : \text{supp}(x) \cap h(p^2) \neq \emptyset\}$. Then

$$N_h = \bigcup_{\ell=1}^{\infty} \coprod_{\substack{w \in \mathcal{W}_\ell(X) \\ \text{supp}(w) \cap h(p^2) \neq \emptyset}} C_w$$

where C_w is the cylinder set defined by w and $\mathcal{W}_\ell(X)$ is the set of all words of length ℓ that appear in the elements of X. Therefore, N_h is a Borel subset of X (in fact it's open). Hence, for each $h \in \mathbb{Z}/p^2\mathbb{Z}$, the first set comprising the disjoint union in the term corresponding to h on the right side of eq. (1.2.5) is Borel, since it's the

complement of N_h. At the same time, the second set in this disjoint union may be described as

$$\{x \in X : \mathrm{supp}(x) \cap h(p^2) \text{ is infinite}\} = \bigcap_{m=0}^{\infty} \sigma^{-mp^2} N_h.$$

This is immediate from the fact that $\sigma^{mp^2} x \in N_h$ whenever $h + Mp^2 \in \mathrm{supp}(x)$ for some $M \geq m$. This intersection is clearly a Borel set, so S_p is Borel as well.

Now set

$$X_I = \{x \in X : \mathrm{supp}(x) \text{ omits exactly } i_p \text{ residue classes } \mod p^2 \text{ for all } p \text{ and } x \in S_p \text{ for all } p\}.$$

Then this is an invariant subset of X. It's also Borel because

$$X_I = \bigcap_p \{x \in X : x \text{ omits exactly } i_p \text{ residue classes } \mod p^2\} \cap S_p.$$

We showed above that S_p is Borel; at the same time we have for any p

$$\{x \in X : x \text{ omits exactly } i_p \text{ residue classes } \mod p^2\} = \bigcup_{L \geq 1} \bigcap_{\ell \geq L} \coprod_{\substack{w \in W_\ell(X) \\ w \text{ omits exactly} \\ i_p \text{ residue classes } \mod p^2}} C_w.$$

This is clearly a Borel set, and it follows that X_I is Borel.

We will denote $X_1 = X_{(1,1,1,\dots)}$. Let $\mathcal{C}_I = \overline{X_I}$. Then $\mathcal{C}_I$ is a closed subshift of X. Also, for any $r \geq 1$ and $(i_1, \dots, i_r) \in \prod_{k=1}^{r} \{1, \dots, p_k^2 - 1\}$, define

$$X_{(i_1,\dots,i_r)}(r) = \{x \in X(r) : \mathrm{supp}(x) \text{ omits exactly } i_k \text{ residue classes } \mod p_k^2 \text{ for } k = 1, \dots, r \text{ and } x \in S_{p_k} \text{ for } k = 1, \dots, r\},$$

and let $C_{(i_1,\dots,i_r)}(r) = \overline{X_{(i_1,\dots,i_r)}(r)}$. Note that for an infinite sequence $I = (i_1, i_2, \dots)$, we have $C_I = \cap_{r=1}^{\infty} C_{(i_1,\dots,i_r)}(r)$. By the Chinese remainder theorem we have the inequalities

$$2^{m\prod_{k=1}^{r}(p_k^2-i_k)} \leq u^I_{mp_1^2p_2^2\cdots p_r^2} \leq \binom{b_1}{i_1}\binom{b_2}{i_2}\cdots\binom{p_r^2}{i_r} 2^{m\prod_{k=1}^{r}(p_k^2-i_k)},$$

where $u^I_{mp_1^2p_2^2\cdots p_r^2}$ is the number of distinct words of length $mp_1^2p_2^2\cdots p_r^2$ appearing in elements of $X_I(r)$. Therefore

$$h_{\text{top}}(\mathcal{C}_I(r)) = \log 2 \prod_{k=1}^{r}\left(1 - \frac{i_k}{p_k^2}\right)$$

and so

$$\begin{aligned} h_{\text{top}}(\mathcal{C}_I) &= \lim_{r\to\infty} h_{\text{top}}(\mathcal{C}_{(i_1,\dots,i_r)}(r)) \\ &= \lim_{r\to\infty} \log 2 \prod_{k\leq r}\left(1 - \frac{i_k}{p_k^2}\right) \\ &= \log 2 \prod_{p}\left(1 - \frac{i_p}{p^2}\right). \end{aligned}$$

Lemma 1.2.3. *Let ν be an ergodic measure on X. Then there exists a unique sequence $I \in \prod_p \{1, \dots, p^2-1\}$ such that $\nu(X_I) = 1$. In particular, any ergodic measure of maximal entropy on X is concentrated on X_1.*

Proof. Although one might be tempted to reach the desired conclusion immediately from the fact that X is the disjoint union of the X_I over all sequences I, this is invalid since there are uncountably many such sequences, so we must argue differently. Let ν be an ergodic measure on X, and let $Y = \text{supp}(\nu)$. For each $y \in Y$ and each p let $i_p(y) = p^2 - |\pi_{p^2}(\text{supp}(y))|$. Then $i_p : Y \to \mathbb{N}$ is measurable, and we have for any p

$$Y = \coprod_{i=1}^{p^2-1} Y_i(p)$$

where $Y_i(p) = \{y \in Y : i_p(y) = i\}$. Since ν is ergodic with $\nu(Y) = 1$, and the $Y_i(p)$ are invariant and mutually disjoint, there exists a unique $i(p) \in \{1, \ldots, p^2 - 1\}$ such that $\nu(Y_{i(p)}) = 1$. As this is true for all p, and since a countable intersection of sets of full measure also has full measure, we have $\nu(Y \cap X_I) = \nu\left(\cap_k Y_{i(p)}\right) = 1$ where $I = (i(p))_p$, which proves the claim. □

Consider the map $\rho : X_1 \to G$ given by sending a sequence x to the sequence $g_x \in G$, where $(g_x)_p$ is the unique residue class omitted by the support of x mod p^2. ρ is measurable; indeed, if we fix finitely many primes $p_1, \ldots, p_r$ and classes g_i mod p_i^2 for $i = 1, \ldots, r$, then the inverse image under ρ of the open set

$$\{g_1\} \times \{g_2\} \times \cdots \times \{g_r\} \times \prod_{p > p_r} \mathbb{Z}/p^2\mathbb{Z}$$

is

$$X_1 \cap \bigcap_{\ell \geq 1} \coprod_{\substack{w \in W_\ell(X) \\ \text{supp}(w) \text{ omits } g_i \bmod p_i^2 \\ \text{for } i=1,\ldots,r}} C_w$$

which is clearly a Borel set.

Let

$$Y = \{x \in X_1 : g_x \in G_1^c\} = \rho^{-1}(G_1^c).$$

Y is Borel (since it's the inverse image of the closed set G_1^c under the Borel map ρ) and recurrent. Indeed, any sequence in Y must contain infinitely many 1's by virtue of belonging to X_1. Therefore, if one shifts such a sequence sufficiently many times so that there is a 1 in its first position, then there can't be any prime p such that this shifted sequence omits 1 mod p^2 from its support.

Let $\sigma_Y : Y \to Y$ be the transformation induced by the shift σ on X, as defined in Section 1.1.

Proposition 1.2.4. *$h(Y, \sigma_Y) = \log 2$, where $h(Y, \sigma_Y)$ is defined by eq. 1.2.1.*

Proof. Let ν be a σ_Y-ergodic Borel probability measure on Y. By fact (3), there exists a σ-ergodic Borel probability measure μ on X such that $\nu = \mu_Y$. Since $\nu(Y) = 1$, we have $\mu(Y) > 0$, so by ergodicity, we have $\mu(X_1) = 1$ since $Y \subset X_1$ and X_1 is shift invariant. Therefore, $\rho_*\mu$ is a T_1-invariant probability measure on G, so by the unique ergodicity of T_1 it follows that $\rho_*\mu = m$, the mass one Haar measure on G. But we have $Y = \rho^{-1}(G_1^c)$, so $\mu(Y) = \mu(\rho^{-1}(G_1^c)) = m(G_1^c)$. Observe that

$$G_1^c = \prod_p (\mathbb{Z}/p^2\mathbb{Z})\backslash\{1\}$$

and therefore

$$\begin{aligned} m(G_1^c) &= \lim_{K\to\infty} \frac{1}{p_1^2 \cdots p_K^2} \prod_{k=1}^{K} (p_k^2 - 1) \\ &= \prod_p \left(1 - \frac{1}{p^2}\right). \end{aligned}$$

So we find

$$\mu(Y) = \prod_p \left(1 - \frac{1}{p^2}\right) \tag{1.2.6}$$

for any ergodic measure μ on X with $\mu(X_1) = 1$. Now, by Abramov's formula (fact (2) in Section 1.1) we have

$$h_{\mu_Y}(\sigma_Y)\mu(Y) = h_\mu(\sigma) \leq h_{top}(X) = \log 2 \prod_p \left(1 - \frac{1}{p^2}\right)$$

so by (1.2.6), this yields $h_{\mu_Y}(\sigma_Y) \leq \log 2$. On the other hand, there exists a measure of maximal entropy for X since it is a subshift system, and this induces a measure of entropy $\log 2$ for σ_Y by Abramov's formula once again; therefore $h(Y, \sigma_Y) = \log 2$ as claimed. □

Now, consider the product $G \times \Omega^{(2)}$ as before, with the skew-transformation T.

We would like to induce T to the subset $G_1^c \times \Omega^{(2)}$, but this is not strictly in keeping with our definitions since G_1^c isn't a recurrent subset of G. However, observe by Poincaré recurrence that since the system (G, T_1, m) is ergodic and $m(G_1^c) > 0$, there is a recurrent subset $R_1 \subseteq G_1^c$ with $m(R_1) = m(G_1^c)$, so we may unambiguously define the induced transformation $\overline{T} : R_1 \times \Omega^{(2)} \to R_1 \times \Omega^{(2)}$. Since (G, T_1) is uniquely ergodic, any T-invariant measure on $G \times \Omega^{(2)}$ must project to m on the first factor. Therefore, for any T-invariant measure μ on $G \times \Omega^{(2)}$, the induced transformation $\overline{T} : G_1^c \times \Omega^{(2)} \to G_1^c \times \Omega^{(2)}$ is well-defined μ-almost everyhwere.

Since this introduces yet another system into our argument, we diagram the relations between the various systems of interest, where the diagram below is commutative:

$$
\begin{aligned}
\sigma &: X \to X \\
T &: G \times \Omega^{(2)} \to G \times \Omega^{(2)} && T(g, y) = (T_1 g, \sigma^{\chi_{G_1^c}(g)} y) \\
\iota &: G \times \Omega^{(2)} \to X && \text{insertion map} \\
\sigma_Y &: Y \to Y && \text{induced transformation of } \sigma \text{ to } Y \subset X \\
\overline{T} &: G_1^c \times \Omega^{(2)} \to G_1^c \times \Omega^{(2)} && \text{induced transformation of } T \\
\bar{\iota} &: G_1^c \times \Omega^{(2)} \to X && \text{restriction of } \iota \text{ to } G_1^c \times \Omega^{(2)} \\
\bar{\iota}|_W &: W \to Y, && \text{restriction of } \bar{\iota} \text{ to } W = \bar{\iota}^{-1}(Y).
\end{aligned}
$$

$$
\begin{array}{ccccccc}
\bar{\iota}^{-1}(Y) = W & \xrightarrow{\overline{T}_W} & W & \hookrightarrow & G \times \Omega^{(2)} & \xrightarrow{T} & G \times \Omega^{(2)} \\
\downarrow{\scriptstyle \bar{\iota}} & & \downarrow{\scriptstyle \bar{\iota}} & & \downarrow{\scriptstyle \iota} & & \downarrow{\scriptstyle \iota} \\
Y & \xrightarrow{\sigma_Y} & Y & \hookrightarrow & X & \xrightarrow{\sigma} & X
\end{array}
$$

Proposition 1.2.5. *For any T-invariant measure μ on $G \times \Omega^{(2)}$, we have $\overline{T} = \overline{T_1} \times \sigma$ μ-a.e., where $\overline{T_1} : G_1^c \to G_1^c$ is induced from T_1.*

Proof. Let $(g, y) \in G_1^c \times \Omega^{(2)}$. By definition, we have $\overline{T}(g, y) = T^{r'(g,y)}(g, y)$ where $r' : G_1^c \times \Omega^{(2)} \to \mathbb{N}$ is the return-time function. Now, if $r'(g, y) < \infty$ (this is the case

for μ-a.e. (g, y) by the above), we have

$$\begin{aligned} r'(g,y) &= \min\{n \geq 1 : T^n(g,y) \in G_1^c \times \Omega^{(2)}\} \\ &= \min\{n \geq 1 : T_1^n(g) \in G_1^c\} \\ &= r_1(g) \end{aligned}$$

where $r_1 : G_1^c \to \mathbb{N}$ is the return-time function (which is well-defined for m-a.e. $g \in G_1^c$). Observe that for $g \in G_1^c$ with $r_1(g) < \infty$ and $1 \leq n < r_1(g)$, we have $T_1^n g \in G_1$, and therefore $T^n(g, y) = (T_1^n g, \sigma y)$ by (1.2.2). In particular, $T^{r_1(g)-1}(g, y) = (T_1^{r_1(g)-1} g, \sigma y)$ and $T_1^{r_1(g)-1} g \in G_1$, so we have

$$\overline{T}(g,y) = T^{r_1(g)}(g,y) = (T_1^{r_1(g)} g, \sigma y) = (\overline{T}_1 \times \sigma)(g,y),$$

which proves the claim. □

Proposition 1.2.6. $h(G_1^c \times \Omega^{(2)}, \overline{T}) = \log 2$.

Proof. Let ν be a $\overline{T}$-invariant Borel probability measure on $G_1^c \times \Omega^{(2)}$. By Proposition 1.2.5, we have $\overline{T} = \overline{T_1} \times \sigma$ ν-a.e. Hence if we let $\eta = (\pi_2)_* \nu$, then η is a σ-invariant measure on $\Omega^{(2)}$, and by the unique ergodicity of $\overline{T_1}$, ν projects to $\overline{m}$ on the first factor. Therefore, we have the inequality (cf. [Dow] Fact 4.4.3)

$$h_\nu(\overline{T}) \leq h_{\overline{m}}(\overline{T_1}) + h_\eta(\Omega^{(2)}) \leq \log 2,$$

since $h_{\overline{m}}(\overline{T}_1) = 0$. At the same time, if we let $\nu = \overline{m} \times \omega_{(1/2,1/2)}$, where $\omega_{(1/2,1/2)}$ is the maximum entropy Bernoulli measure on $\Omega^{(2)}$, then

$$h_\nu(\overline{T}) = h_{\overline{m}}(\overline{T_1}) + h_{\omega_{(1/2,1/2)}}(\Omega^{(2)}) = \log 2,$$

so the upper bound is achieved, proving the claim.

□

We may define a map $\phi : Y \to G_1^c \times \Omega^{(2)}$ by $x \mapsto (g_x, \widehat{x})$, where $g_x \in G_1^c$ is the unique sequence of residue classes omitted by the support of x modulo the p^2 (this belongs to G_1^c by definition of Y) and $\widehat{x} \in \Omega^{(2)}$ is obtained by removing from x the zeros along these residue classes.

Proposition 1.2.7. *For any $x \in Y$, we have $r_1(g_x) < \infty$.*

Proof. If $r_1(g_x) = \infty$, then for any $n \geq 1$ there exists a prime p such that $g_{\sigma^n x} \equiv 1 \bmod p^2$. Hence, for any $n \geq 2$ there exists a p such that $g_x \equiv n \bmod p^2$. But this clearly implies that $x = 00000\cdots$, which doesn't belong Y since $Y \subset X_1$. □

Therefore, the composite $\overline{T} \circ \phi : Y \to G_1^c \times \Omega^{(2)}$ is well-defined everywhere on Y, since $\overline{T}(g, y)$ is well-defined whenever $r_1(g) < \infty$.

Proposition 1.2.8. *$\phi \circ \sigma_Y = \overline{T} \circ \phi$ everywhere on Y.*

Proof. Let $x \in Y$; then $\phi(\sigma_Y x) = \phi(\sigma^{r(x)} x) = (g_{\sigma^{r(x)}x}, \widehat{\sigma^{r(x)}x})$. We have $g_{\sigma^{r(x)}x} = T_1^{r(x)} x = \overline{T_1} x$ since the map $X_1 \to G, x \mapsto g_x$ intertwines σ with T_1. We claim that $\widehat{\sigma^{r(x)}x} = \sigma(\widehat{x})$, which will prove our claim in view of Proposition 1.2.5.

We may calculate for $m \geq 1$

$$(\widehat{\sigma^{r(x)}x})_m = (\sigma^{r(x)}x)_{\ell_m} = x_{\ell_m + r(x)},$$

where the sequence ℓ_m is defined for $m \geq 1$ by

$$\ell_1 = \min\{\ell \geq 1 : \ell \not\equiv g_p - r(x) \bmod p^2 \ \forall p\}$$
$$\ell_{m+1} = \min\{\ell > \ell_m : \ell \not\equiv g_p - r(x) \bmod p^2 \ \forall p\}.$$

At the same time, we have

$$\widehat{x}_m = x_{\ell'_m},$$

where $\ell'_1 = 1$ and $\ell'_{m+1} = \min\{\ell > \ell'_m : \ell \not\equiv g_p \bmod p^2 \ \forall p\}$. Observe that by definition of Y and r and the fact that $x \in Y$, for $1 \leq n < r(x)$ there exists a prime p_n such that $\sigma^n x$ omits the residue class 1 mod p_n^2. Therefore, x itself omits the classes $n+1 \bmod p_n^2$ for $1 \leq n < r(x)$. Thus, if $m > 1$ then $\ell'_m > r(x)$. It follows that $\ell_m + r(x) = \ell'_{m+1}$. Indeed, when $m = 1$ we have

$$\begin{aligned}
\ell'_2 &= \min\{\ell > 1 : \ell \not\equiv g_p \bmod p^2 \ \forall p\} \\
&= \min\{\ell > r(x) : \ell \not\equiv g_p \bmod p^2 \ \forall p\} \\
&= r(x) + \min\{\ell \geq 1 : \ell \not\equiv g_p - r(x) \bmod p^2 \ \forall p\} \\
&= r(x) + \ell_1
\end{aligned}$$

and now for $m > 1$ it easily results from the inductive definitions of ℓ_m and ℓ'_m. Therefore

$$\widehat{(\sigma^{r(x)}x)}_m = x_{\ell_m + r(x)} = x_{\ell'_{m+1}} = (\sigma(\widehat{x}))_m$$

which proves the claim.

□

Let $W = \overline{\iota}^{-1}(Y)$. Note that $\overline{\iota}|_W : W \to Y$ is bijective since $\overline{\iota} \circ \phi = Id_Y$. It follows that the restriction $\overline{\iota}|_W : W \to Y$ is a Borel isomorphism. Moreover, if $\phi(x) \in W$ (every element of W may be realized this way for a unique $x \in Y$) then by Proposition 1.2.8

$$\overline{\iota}\overline{T}\phi(x) = \overline{\iota}\phi\sigma_Y x = \sigma_Y x = \sigma_Y \overline{\iota}\phi(x),$$

so $\overline{\iota}|_W$ is a measurable Borel isomorphism between the systems $(W, \overline{T}|_W)$ and (Y, σ_Y):

$$\overline{\iota}|_W : (W, \overline{T}|_W) \xrightarrow{\sim} (Y, \sigma_Y).$$

In particular, $h(W, \overline{T}|_W) = h(Y, \sigma_Y) = \log 2 = h(G_1^c \times \Omega^{(2)}, \overline{T})$ where the second

equality is Proposition 1.2.4 and the last equality is Proposition 1.2.6. Since $(W, \overline{T}|_W)$ is a subsystem of $(G_1^c \times \Omega^{(2)}, \overline{T})$, it follows that:

If $(G_1^c \times \Omega^{(2)}, \overline{T})$ is intrinsically ergodic, then $(W, \overline{T}|_W)$ is intrinsically ergodic.

Indeed, if we could find two distinct measures of maximal entropy for $\overline{T}|_W$, then these could be extended to distinct measures of maximal entropy for $\overline{T}$ on $G_1^c \times \Omega^{(2)}$. On the other hand, since $(W, \overline{T}|_W)$ is Borel isomorphic to (Y, σ_Y), this implies the intrinsic ergodicity of the latter system. But (Y, σ_Y) is induced from (X, σ). Thus, if ν is a σ_Y-ergodic measure of maximal entropy $\log 2$ on Y, then by facts (1) and (3) in section 1.1 there is a σ-ergodic measure μ on X with $\nu = \mu_Y$, and then by Abramov's formula we have

$$h_\mu(X) = \mu(Y) h_\nu(\sigma_Y) = \mu(Y) \log 2 = \log 2 \prod_p \left(1 - \frac{1}{p^2}\right) = h_{\text{top}}(X)$$

where the second to last equality follows from (1.2.6), since the fact that $\mu(Y) > 0$ implies $\mu(X_1) = 1$ since μ is ergodic and $Y \subset X_1$. This shows that the map $\mu \mapsto \mu_Y$ is a bijection between σ-ergodic measures of maximal entropy on X (such a measure satisfies $\mu(Y) = 6/\pi^2 > 0$ by (1.2.6) and σ_Y-ergodic measures of maximal entropy on Y, and in particular intrinsic ergodicity for the two systems is equivalent. Combining this with the above observations, we see that

If $(G_1^c \times \Omega^{(2)}, \overline{T})$ is intrinsically ergodic, then (X, σ) is intrinsically ergodic.

Proposition 1.2.9. $(G_1^c \times \Omega^{(2)}, \overline{T})$ *is intrinsically ergodic.*

Proof. By Proposition 1.2.5, $\overline{T} = \overline{T_1} \times \sigma$ up to a set of measure zero for any invariant measure. Also, by the facts in Section 1.1, $\overline{T_1}$ is uniquely ergodic, with the zero entropy measure $\overline{m} = m_{G_1^c}$ as its unique invariant measure. Therefore, if μ is a $\overline{T}$-ergodic measure of maximal entropy $\log 2$ on $G_1^c \times \Omega^{(2)}$, then its projection onto G_1^c

must be $(\pi_1)_*\mu = \overline{m}$, while it projects to some σ-ergodic measure $\eta = (\pi_2)_*\mu$ on $\Omega^{(2)}$. As in the proof of Proposition 1.2.6, we have

$$\log 2 = h_\mu(G_1^c \times \Omega^{(2)}, \overline{T}) \leq h_{\overline{m}}(G_1^c, \overline{T_1}) + h_\eta(\Omega^{(2)}, \sigma) = h_\eta(\Omega^{(2)}, \sigma) \leq \log 2,$$

which shows that $h_\eta(\Omega^{(2)}, \sigma) = \log 2$, and therefore $\eta = \omega_{(1/2,1/2)}$, the maximum entropy Bernoulli measure on $\Omega^{(2)}$. But the system $(G_1^c, \overline{T_1}, \overline{m})$ is ergodic with entropy zero, and any such system is *disjoint* from the Bernoulli system $(\Omega^{(2)}, \sigma, \omega_{(1/2,1/2)})$: the only invariant measure μ on the product system $(G_1^c \times \Omega^{(2)}, \overline{T} \times \sigma)$ with $(\pi_1)_*\mu = \overline{m}$ and $(\pi_2)_*\mu = \omega_{(1/2,1/2)}$ is the product measure $\overline{m} \times \omega_{(1/2,1/2)}$ ([Fu] Thm. 1.2). As μ was chosen arbitrarily, it follows that this is the only measure of maximal entropy on $(G_1^c \times \Omega^{(2)}, \overline{T})$.

□

It now follows from the facts in Section 1.1 that $m \times \omega_{(1/2,1/2)}$ is the only measure of maximal entropy on the skew-product $(G \times \Omega^{(2)}, T)$, which finally gives us the following.

Theorem 1.2.10. *$\iota_*(m \times \omega_{(1/2,1/2)})$ is the only measure of maximal entropy on X.*

In fact, it's easy to see that the restriction of ι to the inverse image of X_1 in $G \times \Omega^{(2)}$ is bijective; since X_1 has full measure for $\eta = \iota_*(m \times \omega_{(1/2,1/2)})$, we see that $\iota^{-1}(X_1)$ has full measure for $m \times \omega_{(1/2,1/2)}$, and this gives the following.

Theorem 1.2.11. *ι induces an isomorphism of measure-preserving systems*

$$\iota : (G \times \Omega^{(2)}, T, m \times \omega_{(1/2,1/2)}) \xrightarrow{\sim} (X, \sigma, \eta).$$

1.3 Failure of the Gibbs property

An invariant probability measure μ supported on a closed subshift X of the full 2-shift $\Omega^{(2)}$ is said to be a Gibbs measure if there exists a constant $c > 0$ such that for any $n \geq 1$ and any word w of length n appearing in X, we have the inequality

$$\mu(C_w) \geq c \cdot e^{-n \cdot h_{\text{top}}(X)}$$

where $C_w \subset X$ is the cylinder set defined by w. In other words, the quantity $e^{|w| \cdot h_{\text{top}}(X)} \mu(C_w)$ is uniformly bounded below as w varies over all words in X. The importance of this property is the following:

Proposition 1.3.1 ([Wei] Lemma 2)**.** *Suppose X is a closed subshift of $\Omega^{(2)}$ and μ is an ergodic Gibbs measure supported on X such that $h_\mu(X) = h_{top}(X)$. Then μ is the only measure of maximal entropy on X.*

The unique measures of maximal entropy on subshifts of finite type and sofic systems are Gibbs measures. Our aim in this section is to show that the unique measure of maximal entropy on the squarefree flow X does not possess the Gibbs property, in contrast to many well-known classes of intrinsically ergodic systems.

We have shown that the measure $\eta = \iota_*(m \times \omega_{(1/2,1/2)})$ is the only measure of maximal entropy on the squarefree flow X. This allows us to give an explicit formula for $\eta(w) := \eta(C_w)$ for any word w in X, in the following way. Let $G(w) = \{g \in G : \forall p,\ g_p \notin \text{supp}(w) \bmod p^2\}$. Observe that if C_w is the cylinder defined by w, then

$$\iota^{-1}(C_w) = \{(g, y) \in G \times \Omega^{(2)} : g \in G(w) \text{ and } y_{m-\alpha_g(m)} = w_m \text{ whenever } m \leq n \text{ and } \delta_g(m) = 0\}$$

where $\delta_g(m) = 0$ if and only if there does not exist any prime p such that $m \equiv g_p \bmod p^2$. Hence, by Fubini's theorem and the definition of the Bernoulli measure

we have

$$\eta(w) = \iota_*(m \times \omega_{(1/2,1/2)})(C_w) = \int_{G(w)} 2^{-|\{m \le n:\ \delta_g(m)=0\}|} dm(g).$$

By definiton, we have for $g \in G(w)$ that $\mathrm{supp}(w) \subseteq \{m \le n :\ \delta_g(m) = 0\}$. Therefore, we have $2^{-|\{m \le n:\ \delta_g(m)=0\}|} \le 2^{-|\,\mathrm{supp}(w)|}$ for any $g \in G(w)$, so we get $\eta(w) \le 2^{-|\,\mathrm{supp}(w)|} m(G(w))$. For $r \ge 1$, define $G_r(w) = \{g \in G_r : g_i \notin \mathrm{supp}(w) \bmod p_i^2 \text{ for } i = 1, \ldots, r\}$. By expressing the Haar measure on G as the limit of the counting measures on the G_r as $r \to \infty$, we find

$$\eta(w) \le 2^{-|\,\mathrm{supp}(w)|} \lim_{r\to\infty} \frac{|G_r(w)|}{p_1^2 \cdots p_r^2}.$$

For any $r \ge 1$ we have

$$\frac{|G_r(w)|}{p_1^2 \cdots p_r^2} = \prod_{i=1}^{r} \left(1 - \frac{u(w,i)}{p_i^2}\right)$$

where $u(w,i)$ is the number of residue classes modp_i^2 defined by $\mathrm{supp}(w)$. Observe that if $p_i^2 > n$ then $u(w,i) = |\,\mathrm{supp}(w)|$. Since every term in the product is less than 1 we get

$$\frac{|G_r(w)|}{p_1^2 \cdots p_r^2} \le \prod_{\substack{1 \le i \le r \\ p_i > \sqrt{n}}} \left(1 - \frac{|\,\mathrm{supp}(w)|}{p_i^2}\right), \tag{1.3.1}$$

so letting $r \to \infty$, we find

$$\eta(w) \le 2^{-|\,\mathrm{supp}(w)|} \prod_{\substack{p \\ p > \sqrt{n}}} \left(1 - \frac{|\,\mathrm{supp}(w)|}{p^2}\right). \tag{1.3.2}$$

For each $n \ge 1$, let $w^{(n)}$ be the word of length n appearing in X defined by $w_k^{(n)} = \mu^2(k)$ for $1 \le k \le n$. Clearly, $|\,\mathrm{supp}(w^{(n)})| = Q(n)$, the number of squarefree integers

less than or equal to n. Therefore, the above inequality yields

$$\eta(w^{(n)}) \leq 2^{-Q(n)} \prod_{p>\sqrt{n}} \left(1 - \frac{Q(n)}{p^2}\right).$$

We proceed to estimate the product; call it $P(n)$. Taking logs and using the Taylor series for $\log(1-x)$ gives

$$\begin{aligned} \log P(n) &= \sum_{p>\sqrt{n}} \log\left(1 - \frac{Q(n)}{p^2}\right) \\ &\sim -Q(n) \sum_{p>\sqrt{n}} \frac{1}{p^2}. \end{aligned}$$

By Riemann-Stieltjes integration and the prime number theorem we therefore get

$$\log P(n) \sim -2\frac{Q(n)}{\sqrt{n}\log n}$$

as $n \to \infty$, so we find from the above

$$\eta(w^{(n)}) \ll 2^{-Q(n)} \exp\left(-2\frac{Q(n)}{\sqrt{n}\log n}\right)$$

as $n \to \infty$, with an absolute implied constant.

Now, write $Q(n)$ as

$$Q(n) = \frac{6}{\pi^2}n + R(n).$$

The error term $R(n)$ satisfies ([P])

$$R(n) = o\left(\frac{\sqrt{n}}{(\log N)^A}\right) \tag{1.3.3}$$

for any $A > 0$. Since $h_{\text{top}}(X) = (6/\pi^2)\log 2$ we have

$$\begin{aligned} e^{n \cdot h_{\text{top}}(X)} \eta(w^{(n)}) &= 2^{6n/\pi^2} \eta(w^{(n)}) \\ &\ll 2^{6n/\pi^2} 2^{-Q(n)} \exp\left(-2\frac{Q(n)}{\sqrt{n}\log n}\right) \\ &\ll \exp\left(-2\frac{Q(n)}{\sqrt{n}\log n} - R(n)\log 2\right) \end{aligned}$$

and using the above expression for $Q(n)$ and the error estimate (1.3.3) we see

$$e^{n \cdot h_{\text{top}}(X)} \eta(w^{(n)}) \ll \exp\left(-\frac{12}{\pi^2}\frac{\sqrt{n}}{\log n} + o\left(\frac{\sqrt{n}}{\log n}\right)\right).$$

Since the exponent tends to $-\infty$ as $n \to \infty$, we see that

$$e^{n \cdot h_{\text{top}}(X)} \eta(w^{(n)}) \to 0 \text{ as } n \to \infty.$$

Since the Gibbs property precisely states that the quantity on the left is uniformly bounded below for all n and w with length n, we have shown the following.

Proposition 1.3.2. *η is not a Gibbs measure.*

In a similar vein, the above arguments show that any $\mathscr{B}$-free shift as mentioned in the introduction and described in [ELR] carries a unique measure of maximal entropy (this is also proven in [KLW]), and one can ask whether the Gibbs property holds for this measure. Given that the argument in this section used specific properties of the squares of primes, it's unclear whether or not this should be the case in general.

Chapter 2

Möbius disjointness for homogeneous dynamics

This chapter is dedicated to proving the following result.

Theorem 2.0.3. *Let G be a real connected Lie group, $\Gamma \subset G$ a lattice, and $u \in G$ an Ad-unipotent element. Then for every $x \in \Gamma\backslash G$ and every continuous, bounded function f on $\Gamma\backslash G$, we have*

$$\frac{1}{N}\sum_{n\leq N}\mu(n)f(xu^n) \to 0 \text{ as } N \to \infty.$$

2.0.1 Necessary background

We begin by recording several powerful results as the fundamental input for our work. First, the following criterion allows us to establish Möbius disjointness by studying correlations of powers of the u-action.

Theorem 2.0.4 (The disjointness criterion, [BSZ] Thm. 2)**.** *Let $F : \mathbb{N} \to \mathbb{C}$ with $|F| \leq 1$ and let ν be a multiplicative function with $|\nu| \leq 1$. Let $\tau > 0$ be small and*

assume that for all primes $p, q \leq e^{1/\tau}, p \neq q$, *we have that for* M *large enough*

$$\left| \sum_{m \leq M} F(pm)\overline{F(qm)} \right| \leq \tau M. \tag{2.0.1}$$

Then for N *large enough,*

$$\left| \sum_{n \leq N} \nu(n)F(n) \right| \leq 2\sqrt{\tau \log(1/\tau)}N. \tag{2.0.2}$$

In fact, the conclusion of this statement can be drawn even when the inequality (2.0.1) is violated on a small set of primes, whose size may increase as $\tau \to 0$. We will have more to say on this later.

Thus, our aim is to analyze the correlation limits

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(xu^{\lambda n})f(xu^{\mu n}) \tag{2.0.3}$$

where $x \in X, f \in C_b(X)$, and $\lambda, \mu \in \mathbb{N}$. Ratner's uniform distribution theorem allows us to describe such limits in terms of ergodic measures arising naturally from subgroups of G. If $x \in X$, we say that the closure $\overline{\{xu^k : k \in \mathbb{Z}\}}$ of the u-orbit of x is homogeneous if there is a closed subgroup $H \subset G$ such that $u \in H, \widetilde{x}H\widetilde{x}^{-1} \cap \Gamma$ is a lattice in $\widetilde{x}H\widetilde{x}^{-1}$, where $\Gamma\widetilde{x} = x$, and $\overline{\{xu^k : k \in \mathbb{Z}\}} = xH$.

Theorem 2.0.5 (Uniform distribution of unipotent trajectories - [R2] Theorem B and Corollary A (1))**.** *Let* G *be a connected Lie group,* $\Gamma \subset G$ *a lattice, and* $u \in G$ *a unipotent element. Then every orbit of* u *acting on* $\Gamma\backslash G$ *is homogeneous and uniformly distributed in its closure. More precisely, there is a closed subgroup* $H \subset G$

such that $\overline{\{xu^k : k \in \mathbb{Z}\}}$ *is homogeneous with respect to* H *as above, and we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(xu^n) = \int_X f \, \mathrm{d}\nu_H$$

for any bounded continuous function f *on* X*, where* ν_H *is an* H*-invariant Borel probability measure supported on* xH*.*

Moreover, if H *is the smallest closed subgroup of* G *with respect to which the closure of the* u*-orbit of* x *is homogeneous, then the action of* u *on* (xH, ν_H) *is ergodic.*

A probability measure of the form appearing in the above theorem is called *algebraic*. Applying this theorem to the action of the unipotent element (u^λ, u^μ) in the group $G \times G$, we have that the limit (2.0.3) exists for any $\lambda, \mu > 0$, and is given by

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(xu^{\lambda n}) f(xu^{\mu n}) = \int_{\Gamma\backslash G\times\Gamma\backslash G} (f \otimes f)(x, y) \, \mathrm{d}\nu(x, y) \tag{2.0.4}$$

where ν is an algebraic ergodic probability measure on the system $(X \times X, u^\lambda \times u^\mu)$.

Our proof proceeds by considering the various possibilities for the subgroup H appearing in Theorem 2.0.5, relative to which the u-orbit of the point x is homogeneous. In section 2.2, we consider the case in which H is semisimple, where we use geometric arguments to show that the correlation limit measure appearing in (2.0.4) must be the trivial joining of the Haar measure of H with itself, for all but finitely many primes $p \neq q$. In section 2.3, we consider the case that H is any connected Lie group satisfying the conditions of Theorem 2.0.5, and use the Levi decomposition to combine our work in the semisimple case with the known case of Möbius disjointness for nilmanifolds.

We begin by outlining our method in the case $G = SL_2(\mathbb{R})$. While this case has already been addressed in [BSZ], we provide this concrete example in order to

illustrate our general approach in the case of semisimple groups.

2.1 The case $G = SL_2(\mathbb{R})$

Suppose that $\Gamma \subset SL_2(\mathbb{R})$ is an arithmetic lattice, and let $x \in \Gamma\backslash G$. By a theorem of Dani, the closure of the u-orbit of x is one of a finite set, a circle, or all of $\Gamma\backslash G$. In the first two cases, the disjointness of $\mu(n)$ from the u orbit of x is well-known to be true (by Dirichlet's theorem on primes in arithmetic progressions in the first case, and by a theorem of Vinogradov in the second [Dav]).

Thus, we suppose that the u orbit of x is uniformly distributed with respect to the Haar measure ν_G. Then we have by Theorem 2.0.5

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n\leq N} f(xu^{pn})\overline{f(xu^{qn})} = \int_{\Gamma\backslash G\times\Gamma\backslash G} f\otimes \overline{f}\, \mathrm{d}\lambda,$$

where λ is an ergodic joining of $(\Gamma\backslash G, u^p, \nu_G)$ with $(\Gamma\backslash G, u^q, \nu_G)$. We are clearly finished if $\lambda = \nu_G \times \nu_G$, since by the prime number theorem, the statement of Möbius disjointness is unaffected if we subtract $\int_{\Gamma\backslash G} f\, \mathrm{d}\nu_G$ from f.

Thus, suppose that there exists a nontrivial joining of $(\Gamma\backslash G, u^p, \nu_G)$ with $(\Gamma\backslash G, u^q, \nu_G)$. By Ratner's joinings theorem ([R1], or see Theorem 2.3.2 below), there exists some $\beta \in \mathrm{Com}(\Gamma)$ such that

$$\beta u \beta^{-1} = u^{p/q},$$

and

$$\widetilde{x}\beta\widetilde{x}^{-1} \in \mathrm{Com}(\Gamma),$$

where $\Gamma\widetilde{x} = x$. It follows that $\widetilde{x}\beta\widetilde{x}^{-1}$ belongs to $P \cap \mathrm{Com}(\Gamma)$, where P is the parabolic subgroup $N_G(\widetilde{x}U\widetilde{x}^{-1})$ (U being the one-parameter unipotent subgroup generated by u). The fact that the u-orbit of x equidistributes with respect to ν_G implies that, on realizing P as the stabilizer $\mathrm{stab}_G(z)$ of a point $z \in \mathbb{P}^1(\mathbb{R})$, we have $z \notin \mathbb{P}^1(\mathbb{Q})$.

Hence, we will be finished if we can show the following.

Proposition 2.1.1. *Let $P \subset G$ be a proper parabolic subgroup, so that $P = P_z$ is the stabilizer of some $z \in \mathbb{P}^1(\mathbb{R})$, and suppose that $z \notin \mathbb{P}^1(\mathbb{Q})$. Let $\chi : P \cap Com(\Gamma) \to \mathbb{R}^*_{>0}$ be a real character. Then $\chi(P \cap Com(\Gamma)) \cap \mathbb{Q}^* = \{1\}$.*

Letting χ be the real character of $N_G(U)$ defined by

$$\gamma u \gamma^{-1} = u^{\chi(\gamma)},$$

which then determines a character of any conjugate of P, it follows that in the setting above, we must in fact have $p = q$, i.e. there are no nontrivial joinings of $(\Gamma\backslash G, u^p, \nu_G)$ with $(\Gamma\backslash G, u^q, \nu_G)$ if $p \neq q$. Theorem 2.0.4 then implies Theorem 2.0.3.

We divide into cases according to whether Γ is a uniform lattice in G.

2.1.1 $\Gamma\backslash G$ is compact

In this case, Γ is commensurable with a unit group in a quaternion division algebra D defined over a totally real number field K [We].

Consider the 4×4 matrix representation

$$\psi : D \to M_4(\mathbb{R})$$

$$\alpha \mapsto x_0 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + x_1 \begin{pmatrix} 0 & 1 & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & a & 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ b & 0 & 0 & 0 \\ 0 & -b & 0 & 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -a & 0 \\ 0 & b & 0 & 0 \\ -ab & 0 & 0 & 0 \end{pmatrix}.$$

Let D_1 be the group of unit quaternions; this is identified with an algebraic K-subgroup of $SL_4(\mathbb{R})$ under this representation. We thereby obtain an algebraic $\mathbb{Q}$-group $G := \mathrm{Res}_{K/\mathbb{Q}}(D_1)$; this is now a subgroup of $SL_m(\mathbb{R})$ for some $m \geq 4$. Γ

is isomorphic over K to a lattice commensurable with $G_{\mathbb{Z}}$, and so we have a K-isomorphism

$$\mathrm{Com}(\Gamma) \cong \mathrm{Com}(G_{\mathbb{Z}}).$$

Up to scalar multiples of the identity, $\mathrm{Com}(G_{\mathbb{Z}})$ is simply $G_{\mathbb{Q}}$.

Now, suppose that P is a parabolic subgroup of $SL_2(\mathbb{R})$ defined over $\mathbb{R}$, and identify this with a parabolic subgroup of G (we will still use P to denote this group). Let H be the Zariski closure of $P \cap G_{\mathbb{Q}}$ inside $SL_m(\mathbb{R})$ (i.e. the smallest zero locus of a set of polynomials with real coefficients containing $P \cap G_{\mathbb{Q}}$). Observe that H is defined over $\mathbb{Q}$, since it consists of rational matrices. Let $\chi : H \to \mathbb{R}^*_{>0}$ be a character. If P is conjugate to the group of unipotent upper triangular matrices in $SL_2(\mathbb{R})$, then χ must be trivial, so we assume instead that P is conjugate to the group of all upper triangular matrices. Consequently, if $g \in P$ is conjugate to an upper-triangular matrix with diagonal entries $\lambda, \mu = \lambda^{-1}$, then $\chi(g) = \chi(\mathrm{diag}(\lambda, \mu))$.

Since $H \subset P$, it follows from this that χ is determined by its behavior on a maximal $\mathbb{Q}$-algebraic torus S in H. Since Γ is a uniform lattice in $SL_2(\mathbb{R})$, so too is $G_{\mathbb{Z}}$ in $G_{\mathbb{R}}$, and therefore G is $\mathbb{Q}$-anisotropic by the theorem of Borel and Harish-Chandra. This implies that G contains no nontrivial $\mathbb{Q}$-split tori, so this is of course true of H as well.

Therefore, S must be $\mathbb{Q}$-anisotropic, so the restriction $\chi|_S$ is defined over a number field F with $d := [F : \mathbb{Q}] > 1$. Since $S_{\mathbb{Q}}$ consists of rational matrices, we have $\chi(S_{\mathbb{Q}}) \subset F^*$. Thus, if we let $N_{F/\mathbb{Q}} : F^* \to \mathbb{Q}^*$ be the norm map, the composite $N_{F/\mathbb{Q}} \circ \chi : S_{\mathbb{Q}} \to \mathbb{Q}^*$ is well-defined. However, this composite is a $\mathbb{Q}$-character of $S_{\mathbb{Q}}$ (since it's fixed by the action $\chi^\sigma(g) = \sigma(\chi(\sigma^{-1}(g)))$ of $\mathrm{Gal}(F/\mathbb{Q})$), and therefore must be trivial (as otherwise it could be extended to a nontrivial $\mathbb{Q}$-character of the $\mathbb{Q}$-anisotropic torus S).

It follows that if $\chi(s) \in \mathbb{Q}^*$ for some $s \in S_{\mathbb{Q}}$, then

$$N_{F/\mathbb{Q}}(\chi(s)) = 1 = \chi(s)^d$$

and since $d > 1$, this implies $\chi(s) = 1$. Therefore, $\chi(S_{\mathbb{Q}}) \cap \mathbb{Q}^* = \{1\}$, and from the above it follows that $\chi(P \cap G_{\mathbb{Q}}) \cap \mathbb{Q}^* = \{1\}$ as well.

These arguments immediately extend from $P \cap G_{\mathbb{Q}}$ to $P \cap \mathrm{Com}(\Gamma)$: if $B \in \mathrm{Com}(\Gamma)$, then B^n belongs to $D_{\mathbb{Q}}$ for some $n \geq 1$. Therefore, we get $\chi(B^n) = 1$ if B^n belongs to a subgroup of H on which χ is trivial, and it follows in this case that $\chi(B) = 1$ since χ is a character.

2.1.2 $\Gamma\backslash G$ is noncompact

In this case, Γ is commensurable with $SL_2(\mathbb{Z})$, and thus its commensurator is

$$\mathrm{Com}(\Gamma) = \left\{ \frac{1}{(\det A)^{1/2}} A : A \in GL_2(\mathbb{Q}) \right\}.$$

Suppose $P \subset G$ is a parabolic subgroup stabilizing a point z in $\mathbb{P}^1(\mathbb{R})\backslash\mathbb{P}^1(\mathbb{Q})$. Let $\chi : P \cap \mathrm{Com}(\Gamma) \to \mathbb{R}^*$ be a real character. We first restrict our attention to $P \cap SL_2(\mathbb{Q})$. This consists of those rational matrices that stabilize z; in particular, this group is trivial if z is not quadratic, so we assume there exist $a, b, c \in \mathbb{Z}$ such that $az^2 + bz + c = 0$, with $(a, b, c) = 1$ and $d = b^2 - 4ac > 0$. Then we have

$$P \cap \mathrm{Com}(\Gamma) = \left\{ \begin{pmatrix} \frac{t+bu}{2} & au \\ -cu & \frac{t-bu}{2} \end{pmatrix} : t^2 - du^2 \in \mathbb{Q}^+, t, u \in \mathbb{Q} \right\}.$$

The Zariski closure of $P \cap SL_2(\mathbb{Q})$ is defined over $\mathbb{Q}$. Thus, an argument similar to that in the cocompact case implies that it suffices to show that P does not contain a $\mathbb{Q}$-split torus. Suppose to the contrary that it does contain a $\mathbb{Q}$-split torus S. This

torus is conjugate via an element of $GL_2(\mathbb{Q})$ to the diagonal subgroup of $SL_2(\mathbb{R})$, so S stabilizes a point in $\mathbb{P}^1(\mathbb{Q})$. The unipotent radical U of P contracts S, i.e. we have

$$a^{-t}ua^t \to e \text{ as } t \to \infty \text{ for all } u \in U, \{a^t\} \subset S.$$

Thus if $\{a^t\} \subset S$ is a one-parameter subgroup such that the geodesic ray $\{a^t y\}_{t=0}^{\infty}$ converges to z for some $y \in \mathbb{H}$, then we have for any $u \in U$

$$\lim_{t\to\infty} d_{\mathbb{H}}(ua^t y, a^t y) = \lim_{t\to\infty} d(a^{-t}ua^t y, y) < \infty$$

since the set $\{a^{-t}ua^t : t \geq 0\}$ is bounded. It follows that U and therefore all of P fixes the same boundary point as S. But this contradicts the hypothesis that P stabilizes $z \in \mathbb{P}^1(\mathbb{R})\backslash\mathbb{P}^1(\mathbb{Q})$.

2.2 The general semisimple case

Let G be the real locus of a connected, semisimple algebraic group of noncompact type with finite center defined over $\mathbb{Q}$ with $G \subset SL_n(\mathbb{R})$ for some n, Γ an irreducible lattice in G, and u an Ad-unipotent element of G, acting by right translation on the homogeneous space $X = \Gamma\backslash G$. We aim to prove the following.

Theorem 2.2.1. *For any point $x \in X$ whose u-orbit is equidistributed relative to the G-invariant Borel probability measure ν_G on X and any continuous, bounded function f on X, we have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu(n) f(xu^n) = 0. \tag{2.2.1}$$

Let $u = \exp v$ where $v \in \mathfrak{g}$ is nilpotent, and define the one-parameter unipotent

subgroup

$$U = \{\exp(tv) : t \in \mathbb{R}\} = \left\{u^t\right\}.$$

Applying Ratner's uniform distribution theorem (Theorem 2.0.5) to the action of (u^p, u^q) on $\Gamma\backslash G \times \Gamma\backslash G$, where $p \neq q$ are primes, we have

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n\leq N} f(xu^{pn})\overline{f(xu^{qn})} = \int_{\Gamma\backslash G\times\Gamma\backslash G} f \otimes \overline{f}\, \mathrm{d}\nu,$$

where ν is an ergodic algebraic joining of $(\Gamma\backslash G, u^p, \nu_G)$ with $(\Gamma\backslash G, u^q, \nu_G)$.

Let $P = N_G(U)$ and $P_\xi = \xi P \xi^{-1}$. By Ratner's joinings theorem [R1], if $\nu \neq \nu_G \times \nu_G$, then there exists $\beta \in G$ with

$$\beta u \beta^{-1} = u^{p/q}$$

(so $\beta \in P$) such that

$$\delta := \xi\beta\xi^{-1} \in \mathrm{Com}(\Gamma) \cap P_\xi.$$

The definition of P implies that there is a character $\chi : P \to \mathbb{R}^*_{>0}$ determined by

$$\alpha u \alpha^{-1} = u^{\chi(\alpha)} \text{ for } \alpha \in P = N_G(U).$$

This character is then also well-defined on any conjugate of P, and we see from the above that

$$\chi(\delta) = p/q.$$

In order to apply Theorem 2.0.4, we will show that the existence of such a δ forces the pair (p, q) to belong to a finite set determined by x. Since we saw above that any nontrivial joining of u^p with u^q gives us such a δ, this implies the following.

Theorem 2.2.2. *For the character χ described above, the intersection*

$$\chi(Com(\Gamma) \cap P_\xi) \cap \left\{ \frac{p}{q} : p, q \text{ are prime} \right\}$$

is finite. Therefore, if $x \in \Gamma\backslash G$ is u-generic for the Haar measure ν_G, then there are at most finitely many pairs of primes $p \neq q$ such that there is a nontrivial joining of $(\Gamma\backslash G, u^p, \nu_G)$ with $(\Gamma\backslash G, u^q, \nu_G)$ which is supported on the closed orbit $\overline{\{(xu^{pn}, xu^{qn})\}}_{n\in\mathbb{N}}$.

The proof of this is divided according to the nature of the lattice Γ. Recall that a lattice Γ is called *arithmetic* if there exist

1. A closed, connected, semisimple subgroup G' of some $SL_m(\mathbb{R})$ such that G' is defined over $\mathbb{Q}$,

2. Compact, normal subgroups K and K' of G and G', respectively, and

3. an isomorphism $\phi : G/K \to G'/K'$,

such that $\phi(\overline{\Gamma})$ is commensurable to $\overline{G'_{\mathbb{Z}}}$, where $\overline{\Gamma}$ and $\overline{G'_{\mathbb{Z}}}$ are the images of Γ and $G'_{\mathbb{Z}}$ in G/K and G'/K', respectively ([WM2] Def. 5.18).

For nonuniform lattices in the kinds of groups on which we'll focus, there is no need for the compact subgroups arising in this definition:

Proposition 2.2.3 ([WM2] Cor. 5.27)**.** *Assume Γ is an irreducible, arithmetic, nonuniform lattice in the semisimple group G, which is connected and has no compact factors. Then, possibly after replacing G by an isogenous group, there is an embedding of G in some $SL_m(\mathbb{R})$ such that G is defined over $\mathbb{Q}$ and Γ is commensurable to $G_{\mathbb{Z}}$.*

First, we may easily prove Theorem 2.2.2 in the case that Γ is a non arithmetic lattice.

Proof of Thm 2.2.2 when Γ *is not an arithmetic lattice in* G. We have the following theorem of Margulis:

Theorem 2.2.4 ([Mar])**.** *Suppose* Γ *is an irreducible lattice in the semisimple Lie group* G. *Then* Γ *is arithmetic if and only if* $Com(\Gamma)$ *is dense in* G. *Moreover, if* Γ *is nonarithmetic, then* $Com(\Gamma)$ *is a lattice in* G *as well.*

Hence, we have in this case that $\mathrm{Com}(\Gamma)$ is a lattice in G. In particular, $\mathrm{Com}(\Gamma) \cap P_\xi$ is finitely generated as an abstract group, and therefore so too is $\chi(\mathrm{Com}(\Gamma) \cap P_\xi)$. It follows that the set $\{p/q \in \chi(\mathrm{Com}(\Gamma) \cap P_\xi) : p, q \text{ are prime}\}$ is finite. □

We now turn to proving Thm. 2.2.2 when Γ is arithmetic. We first reduce this to a statement about $\mathbb{Q}$-split tori inside $\mathrm{Com}(\Gamma) \cap P_\xi$. Recall that $P = N_G(U)$. This is a parabolic subgroup of G since it contains the normalizer of a maximal unipotent subgroup of G containing U, which is a Borel subgroup of G (cf. [Rag] 12.8). We therefore have the real Langlands decomposition ([W] pg. 81): we may write $P = MAN$ where M is reductive, A is an $\mathbb{R}$-split torus, and N is the unipotent radical of P. Moreover, if we let $L = MA$ and $X(L)$ the group of continuous real characters of L, then

$$M = \bigcap_{\chi \in X(L)} \ker(\chi^2).$$

Thus, the restriction of our character χ to M is necessarily trivial, and the same is true of the restriction of χ to N since this group is unipotent. Let $H = G_\mathbb{Q} \cap P_\xi = M_\mathbb{Q} A_\mathbb{Q} N_\mathbb{Q}$. Since $\chi|_M$ is trivial by the above remarks, it follows that $\chi(H) = \chi(A_\mathbb{Q})$, and therefore

$$\chi(G_\mathbb{Q} \cap P_\xi) \subset \chi(A_\mathbb{Q}).$$

Let $T \subset P_\xi$ be a maximal $\mathbb{Q}$-torus such that $A_\mathbb{Q} \subset T$. There exist $\mathbb{Q}$-tori $S, E \subseteq T$, respectively $\mathbb{Q}$-split and $\mathbb{Q}$-anisotropic, such that $T = SE$ and the intersection $S \cap E$ is finite ([PR] pg. 53). We now make the following key assumption:

Suppose that for the character χ discussed above, we have $\chi(S) = 1$.

This is the fundamental point that will occupy our discussion below. We have by the above $\chi(G_{\mathbb{Q}} \cap P_\xi) \subset \chi(A_{\mathbb{Q}}) \subset \chi(T_{\mathbb{Q}}) = \chi(S_{\mathbb{Q}})\chi(E_{\mathbb{Q}}) = \chi(E_{\mathbb{Q}})$, using the assumption that $\chi(S) = \{1\}$. Let $K/\mathbb{Q}$ be the smallest number field over which E splits and let $d = [K : \mathbb{Q}]$; observe that $d > 1$ since E is $\mathbb{Q}$-anisotropic. Let $E_{\mathbb{Q}} = E \cap G_{\mathbb{Q}}$. χ is defined over K (as is any character of E), and we have $\chi(E_{\mathbb{Q}}) \subset K^*$ since $E_{\mathbb{Q}}$ consists of matrices with rational entries. Therefore, letting $N_{K/\mathbb{Q}}$ be the norm map, the composite $\psi = N_{K/\mathbb{Q}} \circ \chi : E_{\mathbb{Q}} \to \mathbb{Q}^*$ is well-defined.

Observe that $E_{\mathbb{Q}}$ has no nontrivial $\mathbb{Q}$-characters; indeed, it is Zariski dense in E, so such a character could be extended to a $\mathbb{Q}$-character on all of E, contradicting its $\mathbb{Q}$-anisotropy. But ψ is a $\mathbb{Q}$-character of $E_{\mathbb{Q}}$ (indeed, the $\mathbb{Q}$-characters are precisely those characters invariant under the action

$$\psi^\sigma(g) = \sigma(\psi(\sigma^{-1}g))$$

of $\mathrm{Gal}(K/\mathbb{Q})$), and therefore must be trivial.

Suppose that $\chi(h) \in \mathbb{Q}$ for some $h \in E_{\mathbb{Q}}$. Then $N_{K/\mathbb{Q}}(\chi(h)) = \chi(h)^d$. But $N_{K/\mathbb{Q}}(\chi(h)) = \psi(h) = 1$ by the above, so $\chi(h)^d = 1$. Since $d > 1$ and χ only takes positive real values, it follows that $\chi(h) = 1$. This shows

$$\chi(E_{\mathbb{Q}}) \cap \mathbb{Q}^* = \{1\}.$$

Since $\chi(G_{\mathbb{Q}} \cap P_\xi) \subset \chi(E_{\mathbb{Q}})$, this yields Thm. 2.2.2, so long as our main assumption above is true. Hence, the proof of Theorem 2.2.2 reduces to the following.

Theorem 2.2.5. *Suppose $S \subset P_\xi$ is a $\mathbb{Q}$-split torus, where the U-orbit of $x = \Gamma\xi$ is generic with respect to the Haar measure ν_G on X. Then $\chi(S) = \{1\}$.*

The proof of this may be divided into two cases.

Γ is an arithmetic lattice in G and $\Gamma\backslash G$ is compact

After replacing G by the group G' arising in the definition of arithmetic lattices, we may assume that Γ is commensurable with $G_{\mathbb{Z}}$ and the quotient $G_{\mathbb{Z}}\backslash G_{\mathbb{R}}$ is compact. Hence, the theorem of Borel and Harish-Chandra ([BHC] Thm. 11.6) shows that G is a $\mathbb{Q}$-anisotropic group, meaning it contains no nontrivial $\mathbb{Q}$-split torus. The same is therefore of course also true of $\mathrm{Com}(\Gamma)$, which simply equals $G_{\mathbb{Q}}$ up to scalar multiples of the identity. Hence, the intersection $G_{\mathbb{Q}} \cap P_\xi$ contains no nontrivial $\mathbb{Q}$-split torus, and Theorem 2.2.5 follows immediately in this case.

Γ is an arithmetic lattice in G and $\Gamma\backslash G$ is not compact

In this case, we may assume by Proposition (2.2.3) that Γ is commensurable with $G_{\mathbb{Z}}$. This allows us to restrict our attention to rational matrices by the following.

Proposition 2.2.6. *Let S be a subgroup of $Com(\Gamma)$ and $\chi : S \to \mathbb{R}^*_{>0}$ a character. If χ is trivial on $S \cap G_{\mathbb{Q}}$, then χ is trivial on S.*

Proof. Since Γ is commensurable with $G_{\mathbb{Z}}$, Prop. 4.6 on pg. 206 of [PR] yields $\mathrm{Com}(\Gamma) = \pi^{-1}((G/N)_{\mathbb{Q}})$, where N is the largest invariant $\mathbb{Q}$-subgroup of compact type. Since G is assumed to have noncompact type, N is the finite subgroup generated by the centers of the $\mathbb{Q}$-simple factors of G. Since N is finite, we see that if $g \in S \subset \mathrm{Com}(\Gamma)$, then $g^d \in G_{\mathbb{Q}}$ for some $d \in \mathbb{N}_{>0}$. But then $\chi(g^d) = \chi(g)^d = 1$ since χ is assumed to be trivial on $S \cap G_{\mathbb{Q}}$; as χ only takes positive real values, it follows that $\chi(g) = 1$. □

In order to prove Thm. 2.2.5 in this case (when $\Gamma\backslash G$ is non-compact), we will use the fundamental relation between parabolic subgroups of G and the boundary of the symmetric space $Y = G/K$, described by the following.

Proposition 2.2.7 ([GJT] Prop. 3.8)**.** *A closed subgroup $P \subset G$ is parabolic if and only if there exists a point $z \in Y(\infty)$ such that $P = stab_G(z)$.*

We now introduce several notions from [AM] in order to relate behavior at the boundary of Y to $\mathbb{Q}$-split tori in G. Let $y_0 \in Y$ be any point.

Definition 2.2.8. *We call $\eta \in Y(\infty)$ a horospherical limit point for Γ if every open horoball based at η intersects the orbit $\Gamma \cdot y_0$.*

This definition is independent of the choice of y_0. Let's illustrate this notion for the case $G = SL_2(\mathbb{R})$ and $\Gamma \subset G$ being any nonuniform arithmetic lattice. We claim that the horospherical limit points for Γ acting on $G/K = \mathbb{H}$ are precisely the points $\mathbb{P}^1(\mathbb{R}\backslash\mathbb{Q})$. Indeed, suppose that $\eta \in \mathbb{P}^1(\mathbb{R})$ is *not* a horospherical limit point for Γ. We may identify η with the point at infinity by some element of G, under which the horoballs based at η become simply the vertical regions $\{z = x + iy \in \mathbb{H} : y > t\}$ for $t \in \mathbb{R}_{>0}$. Thus, the assumption that η is not a horospherical limit point for Γ implies that for any basepoint $y_0 \in \mathbb{H}$, there is some $T \in \mathbb{R}_{>0}$ such that the orbit Γy_0 is contained in the region $\{x + iy : 0 < y \leq T\}$.

Observe that η belongs to $\mathbb{P}^1(\mathbb{Q})$ if and only if its stabilizer $P = \mathrm{stab}_G(\eta) \subset G$ contains a $\mathbb{Q}$-split $\mathbb{Q}$-torus. Suppose to the contrary that P contains a $\mathbb{Q}$-anisotropic $\mathbb{Q}$-torus A. Then η is on the boundary of Ay_0 for some $y_0 \in \mathbb{H}$. By the above, the orbit $(A \cap \Gamma)y_0$ is bounded below a fixed vertical height in $\mathbb{H}$. However, since A is $\mathbb{Q}$-anisotropic, its group of $\mathbb{Z}$-points $A_{\mathbb{Z}} = A \cap SL_2(\mathbb{Z})$ is a cocompact lattice in A, and therefore so too is $A \cap \Gamma$ since Γ is commensurable with $SL_2(\mathbb{Z})$. It follows that the orbit Ay_0 is also vertically bounded in $\mathbb{H}$; but this contradicts the hypothesis that η lies on the boundary of Ay_0. Therefore, P contains a $\mathbb{Q}$-split torus, so $\eta \in \mathbb{P}^1(\mathbb{Q})$.

For the converse, suppose $\eta \in \mathbb{P}^1(\mathbb{Q})$. Then its stabilizer P contains a $\mathbb{Q}$-split torus S, and we have the Iwasawa decomposition $G = NSK$ where N is the unipotent radical of P. If η is a horospherical limit point for Γ, then it also is one for $SL_2(\mathbb{Z})$, so we may find a sequence of elements $\gamma_i \in SL_2(\mathbb{Z})$ such that, writing $\gamma_i = n_i s_i k_i$ under the Iwasawa decomposition, we have $\alpha(s_i) \to \infty$, where α is the root of S in

G obtained from the root of the usual $\mathbb{Q}$-split torus defined by

$$\alpha \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} = t^2.$$

Since the stabilizer P of η is defined over $\mathbb{Q}$, the $\mathbb{Z}$ points $N_{\mathbb{Z}}$ of its unipotent radical N form a cocompact subgroup of N, so there is a compact subset $C \subset N$ such that we may write any element of N as $n = n'c$ where $n' \in N_{\mathbb{Z}}$ and $c \in C$. Therefore, for the sequence γ_i above we can write

$$\gamma_i = n_i' c_i s_i k_i,$$

and since $n_i \in N_{\mathbb{Z}} \subset SL_2(\mathbb{Z})$, we may simply relabel as needed to see that there is a sequence $\gamma_i \in SL_2(\mathbb{Z})$ such that $\gamma_i = c_i s_i k_i$, where $\alpha(s_i) \to \infty$ and c_i belongs to the fixed compact set $C \subset N$. Now, since $\alpha(s_i) \to \infty$, we have $s_i^{-1} n s_i \to e$ as $i \to \infty$ for any $n \in N$, where e is the identity in G. Hence, since $C \subset N$ is compact, a diagonal argument applied to the sequence $s_i^{-1} c_i s_i$ shows that, by passing to a subsequence if necessary, we have $s_i^{-1} c_i s_i \to e$ as $i \to \infty$. Therefore, if i is sufficiently large,

$$s_i^{-1} \gamma_i = s_i^{-1} c_i s_i k_i \approx k_i.$$

Since $k_i \in K$ and K is compact, we may therefore assume by passing to another subsequence that the sequence $s_i^{-1} \gamma_i$ converges to some $g \in G$. Since S is $\mathbb{Q}$-split, there is some $h \in GL_2(\mathbb{Q})$ such that we may write

$$s_i = h \begin{pmatrix} t_i & 0 \\ 0 & t_i^{-1} \end{pmatrix} h^{-1}$$

where $t_i \to \infty$ as $i \to \infty$. Write

$$h^{-1}g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then, since $s_i^{-1}\gamma_i$ converges to g, we have for all sufficiently large i

$$\gamma_i \approx s_i g = h \begin{pmatrix} t_i & 0 \\ 0 & t_i^{-1} \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = h \begin{pmatrix} t_i a & t_i b \\ t_i^{-1} c & t_i^{-1} d \end{pmatrix}.$$

It follows that for all sufficiently large i, every entry of the rightmost matrix must be close to the corresponding entry of $h^{-1}\gamma_i$. Since $h^{-1} \in GL_2(\mathbb{Q})$, there is an integer M such that $Mh^{-1} \in GL_2(\mathbb{Z})$; then the bottom row of $Mh^{-1}\gamma_i$ is close to $(Mt_i^{-1}c, Mt_i^{-1}d)$ for all large i. But c, d and M are fixed, whereas $t_i \to \infty$; this implies that the bottom row of the integer matrix $Mh^{-1}\gamma_i$ consists only of zeros when i is large enough, a contradiction.

The ideas we've illustrated in the case of $SL_2(\mathbb{Z}) \subset SL_2(\mathbb{R})$ may be generalized to yield the following theorem. For any $\mathbb{R}$-split torus $S \subset G$, there exists a point $y \in Y$ such that Sy is a flat in Y (i.e. a complete, totally geodesic flat submanifold). We say that this flat is $\mathbb{Q}$-split if S is defined over $\mathbb{Q}$ and $\mathbb{Q}$-split.

Theorem 2.2.9 ([AM] Thm. 1.3)**.** *A point $\eta \in Y(\infty)$ is a horospherical limit point for Γ if and only if η does not lie on the boundary of any $\mathbb{Q}$-split flat.*

Recall that $P = N_G(U)$. Therefore, P is a parabolic subgbroup of G (cf. [Rag] 12.8), and so there exists $z \in Y(\infty)$ such that $P = \mathrm{stab}_G(z)$. We then have $P_\xi = \xi P \xi^{-1} = \mathrm{stab}_G(\xi(z))$, and we make the following claim.

Proposition 2.2.10. *Suppose $\xi \in G$ has the property that the $U-$orbit of the point $\Gamma\xi \in X = \Gamma\backslash G$ is generic with respect to the Haar measure on G. Let*

- *P_ξ be the parabolic subgroup of G defined above,*

- *$A \subset P_\xi$ be a maximal $\mathbb{R}$-split torus,*

- *$y \in Y$ be a point such that Ay is a flat in Y,*

- *$\{a^t\}$ be a nontrivial one-parameter subgroup of A such that $\{a^t\} \not\subset C_G(U)$,*

- *$\eta \in Y(\infty)$ be the endpoint of the geodesic ray $\{a^t y\}_{t=0}^{\infty}$, and*

- *$\eta' \in Y(\infty)$ be the endpoint of the geodesic ray $\{a^{-t} y\}_{t=0}^{\infty}$.*

Then at least one of η or η' is a horospherical limit point for Γ.

Proof. We must show that either any open horoball in Y based at η intersects the orbit Γy, or that this holds for any open horoball based at η'. The open horoballs based at η may be parametrized by $\mathbb{R}_{>0}$ as follows: let A^+ denote the Weyl chamber of A containing the ray $\{a^t\}_{t=0}^{\infty}$, and let

$$N = \{u \in G : c^k u c^{-k} \to e \text{ as } k \to \infty \text{ for all } c \text{ in the interior of } A^+\}.$$

N is a maximal unipotent subgroup of G, and we claim that $\xi U \xi^{-1} \subset N$ (under a certain assumption on $\{a^t\}$ to be discussed momentarily). By definition, we have $P = N_G(U)$, and therefore $P_\xi = N_G(\xi U \xi^{-1})$. Hence, for our character χ discussed above, we have for $u_\xi \in \xi U \xi^{-1}$ and c in the interior of A^+

$$c^k u_\xi c^{-k} = u_\xi^{\chi(c)^k}.$$

Observe that since $\{a^t\} \not\subset C_G(U)$, we have either $\chi(c) < 1$ or $\chi(c) > 1$ for all c in the interior of A^+. Thus, the above equation shows that $c^k u_\xi c^{-k} \to e$ assuming $\chi(c) < 1$. If this is not the case, then we may replace $\{a^t\}$ by $\{a^{-t}\}$ to send the Weyl chamber A^+ to its opposite, and thereby will obtain the desired conclusion for η' in place of η. Observe that χ yields a root of the Lie algebra of A; since $\chi(c) < 1$ is implied by the condition that this root be positive on the Weyl chamber containing c^{-1}, we see that either $\xi U \xi^{-1} \subset N$, or that this holds when we replace $\{a^t\}$ by $\{a^{-t}\}$ (and change N accordingly). As the desired conclusion is symmetric with respect to this replacement, we may assume without loss of generality that we're in the former situation.

Now, let $A_\perp$ denote the complement in A of $\{a^t\}$ with respect to the Killing form. Then for any $s \in \mathbb{R}_{\geq 0}$, the set

$$\mathcal{O}_s = \bigcup_{t>s} N A_\perp a^t y \subset Y$$

is an open horoball based at η, and any such horoball may be written in this form ([AM] Lemma 2.5).

For any $s \in \mathbb{R}$, the set $V_s = \cup_{t>s} N_\xi A_\perp a^t \tilde{y} K \subset G$ is open in G, where $\tilde{y} \in G$ is chosen so that its coset in G/K equals y. Now, since the U orbit of $\Gamma\xi$ is dense in $\Gamma\backslash G$, the set $\Gamma\xi U$ is dense in G. Therefore, the set $U\xi^{-1}\Gamma$ is also dense in G, and so too is $\xi U \xi^{-1} \Gamma \tilde{y}$. Hence, the latter set must intersect the open set V_s, so there exists some $t > s$ such that

$$N A_\perp a^t \tilde{y} K \cap \xi U \xi^{-1} \Gamma \tilde{y} \neq \emptyset.$$

But by the preceding discussion, we have $\xi U \xi^{-1} \subset N$; thus the above implies

$$N A_\perp a^t \tilde{y} K \cap \Gamma \tilde{y} \neq \emptyset$$

and since $t > s$, this is equivalent to $\mathcal{O}_s \cap \Gamma y \neq \emptyset$, proving the claim.

□

We are now in a position to prove Thm 2.2.5.

Proof of Thm 2.2.5. Suppose $S \subset P_\xi$ is a nontrivial $\mathbb{Q}$-split torus, but that $\chi(S) \neq \{1\}$. Choose $y \in Y$ such that Sy is a flat in Y. Since $\chi(S) \neq \{1\}$, the definition of χ implies there is a one-parameter subgroup $\{a^t\} \subset S$ such that $\{a^t\} \not\subset C_G(U)$. Let $\eta \in Y(\infty)$ be the endpoint of the geodesic ray $\{a^t y\}_{t=0}^{\infty}$. Then by Prop. 2.2.10, either η or η' (which is also on the boundary of Sy) is a horospherical limit point for Γ (note that S is contained in a maximal $\mathbb{R}$-split torus in P_ξ, which is also a maximal $\mathbb{R}$-split torus in G since P_ξ is parabolic). But this contradicts Thm. 2.2.9. □

2.3 The general case

We now consider all possibilities for the group H appearing in Theorem 2.0.5, rather than only semisimple ones. Indeed, this theorem yields a closed subgroup $C \subset G \times G$ such that $\{(u^{pt}, u^{qt})\}_{t\in\mathbb{R}} \subset C$, $\Gamma_C^\xi := \widetilde{\xi}^{-1}(\Gamma \times \Gamma)\widetilde{\xi} \cap C$ is a lattice in C (where $\widetilde{\xi} = (\xi, \xi) \in G \times G$), and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(xu^{pn}) f(xu^{qn}) = \int_{X\times X} f \otimes f \, \mathrm{d}\nu_C \tag{2.3.1}$$

for any bounded continuous function f on X, where ν_C is a C-invariant probability measure supported on $(x, x)C$. At the same time, there exists a closed subgroup $H \subset G$ such that $\{u^t\}_{t\in\mathbb{R}} \subset H, \Gamma_H^\xi := \xi^{-1}\Gamma\xi \cap H$ is a lattice in H and

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(xu^n) = \int_X f \, \mathrm{d}\nu_H$$

for any bounded continuous function f on X, where ν_H is an H-invariant probability measure supported on xH. It follows that ν_C is a joining of the systems (X, u, ν_H) and $(X, u^{p/q}, \nu_H)$. Observe that each of these systems is conjugate (in the measure-theoretic sense) to the corresponding system on the smaller homogeneous space $\Gamma_H^\xi \backslash H$; this is provided by the map

$$\Gamma\backslash G \to \Gamma_H^\xi \backslash H$$
$$\Gamma\xi h \mapsto \Gamma_H^\xi h,$$

which is well-defined and defined ν_H-almost everywhere, since this measure is supported on xH. The image of the measure ν_C, which lives on the product $X \times X$, under the product of this conjugacy with itself, is then a joining of the systems $(\Gamma_H^\xi \backslash H, u, \nu_H)$ and $(\Gamma_H^\xi \backslash H, u^{p/q}, \nu_H)$. We will refer to this joining as λ, and denote $X_H^\xi := \Gamma_H^\xi \backslash H$. Observe that the systems (X_H^ξ, u, ν_H), $(X_H^\xi, u^{p/q}, \nu_H)$, and $(X_H^\xi \times X_H^\xi, u \times u^{p/q}, \lambda)$ are all ergodic.

Now, since the identity component H^0 has finite index in H, it does no harm to assume that H is connected. Furthermore, on passing to the universal cover, we may assume that H is simply connected as well. This is justified by the obvious fact that the statement of Theorem 2.0.3 holds for $\Gamma\backslash H$ as soon as it holds for any topological cover thereof.

We now consider the real Levi decomposition of the connected, simply connected real Lie group H [PR]: we may write H as a semi-direct product $H = L \rtimes R$, where L is a maximal semisimple subgroup of H, and R is its solvable radical (the maximal connected normal solvable subgroup of H). R is the kernel of the projection $p : H \to L$, and both L and R are connected and simply connected. In order to relate the unipotent flow on $\Gamma\backslash H$ to one on a homogeneous space of L, we require the following fact.

Proposition 2.3.1 ([Sta2],[Wi]). *Suppose that H is a connected Lie group with Levi decomposition $H = L \rtimes R$, and $\Gamma \subset H$ is a lattice such that $\Gamma\backslash H$ supports an ergodic unipotent translation. Then L is totally non-compact (i.e. it has no nontrivial compact factors).*

It follows from this and Wang's theorem ([Sta] Theorem E.11) that the projection Γ_L of the lattice $\Gamma_H^\xi \cap H$ into L is a lattice Γ_L^ξ in L, and $R \cap \Gamma_H^\xi$ is a cocompact lattice in R. Let ν be the image of λ under the product projection $X_H^\xi \times X_H^\xi \to \Gamma_L^\xi\backslash L \times \Gamma_L^\xi\backslash L$.

ν is an ergodic joining of the systems $(\Gamma_L^\xi\backslash L, u_L, \nu_L)$ and $(\Gamma_L^\xi\backslash L, u_L^{p/q}, \nu_L)$, where $u_L = p(u) \in L$ and ν_L is the L-invariant Borel probability measure on $\Gamma_L^\xi\backslash L$. Moreover, ν is supported on the closure of the $(u_L, u_L^{p/q})$-orbit of a point (x_L, x_L), where the u_L orbit of x_L in $\Gamma_L^\xi\backslash L$ is uniformly distributed with respect to ν_L (indeed, the topological density of the u-orbit of ξ in $\Gamma\backslash H$ implies the same of its image under the continuous surjection $\Gamma\backslash H \to \Gamma_L^\xi\backslash L$).

It now follows from our analysis in the previous section (specifically, Theorem 2.2.2) that either $\nu = \nu_L \times \nu_L$, or the pair (p, q) belongs to a finite set determined by the point x. The latter case has no effect on the disjointness criterion (Theorem 2.0.4), so we may assume in what follows that

$$\nu = \nu_L \times \nu_L. \tag{2.3.2}$$

Thus, we have an ergodic joining λ of $(\Gamma_H^\xi\backslash H, u, \nu_H)$ and $(\Gamma_H^\xi\backslash H, u^{p/q}, \nu_H)$ whose projection to $\Gamma_L^\xi\backslash L \times \Gamma_L^\xi\backslash L$ is the product measure $\nu_L \times \nu_L$. In order to lift this information to the level of the group H, where the Levi decomposition $H = L \rtimes R$ may be directly used, we need the full generality of Ratner's joinings theorem.

To state this, we require some terminology. Suppose that H a connected Lie group, Γ a lattice in H, $u_1, u_2 \in H$ unipotent elements, and ν the H-invariant Borel probability measure on $X = \Gamma\backslash H$. Let λ be an ergodic algebraic joining of (X, u_1, ν)

with (X, u_2, ν) and $\Lambda = \mathrm{stab}_{H\times H}(\lambda)$, so that there exists some $x(\lambda) \in X \times X$ such that $\lambda(x(\lambda)\Lambda) = 1$ (by Ratner's measure classification theorem - [R1] Theorem 1). Then the subgroups $\Lambda_1, \Lambda_2 \subset H$ defined by

$$\Lambda_1 = \{h \in H : (h, e) \in \Lambda\}, \qquad \Lambda_2 = \{h \in H : (e, h) \in \Lambda\} \tag{2.3.3}$$

are closed normal subgroups of H. For $z \in X$, let

$$\xi_\lambda(z) = \{y \in X : (z, y) \in x(\lambda)\Lambda\}.$$

This is called the z-fiber of λ.

Theorem 2.3.2 ([R1] Theorem 2)**.** *Let all notation and assumptions be as in the above paragraph. Then there is a $c \in H$ and a continuous, surjective homomorphism $\alpha : H \to H/\Lambda_2$ with kernel Λ_1 and $\alpha(u_1) = u_2\Lambda_2$, such that*

$$\xi_\lambda(\Gamma h) = \{\Gamma c\beta_i\alpha(h) : i = 1, \ldots, n\}$$

for all $h \in H$, where the intersection $\Gamma_0 = \alpha(\Gamma) \cap c\Gamma c^{-1}\Lambda_2$ is of finite index in $\alpha(\Gamma)$ and $c\Gamma c^{-1}\Lambda_2, n = |\Gamma_0\backslash\alpha(\Gamma)|$, and $\alpha(\Gamma) = \{\Gamma_0\beta_i : i = 1, \ldots, n\}$.

Recall that, in our setting, λ is an ergodic joining of (X_H, u^p, ν_H), with (X_H, u^q, ν_H), and its pushforward to $X_L \times X_L$ under the map induced by the projection $p \times p : H \times H \to L \times L$ is the product measure $\nu_L \times \nu_L$. We now process Ratner's description of joinings in the following way.

Proposition 2.3.3. *Let $\Lambda_1, \Lambda_2 \subseteq H$ be the normal subgroups defined in eq. (2.3.3). Then we have $p(\Lambda_1) = p(\Lambda_2) = L$.*

Proof. By Ratner's measure classification theorem, we have $\lambda(x_0\Lambda) = 1$ for some $x_0 \in X_H \times X_H$, and it follows that $(\pi\times\pi)_*\lambda((\pi\times\pi)(x_0\Lambda)) = 1$. Clearly, $(\pi\times\pi)(x_0\Lambda) =$

$((\pi \times \pi)x_0)(p \times p)\Lambda$. Since $(\pi \times \pi)_*\lambda = \nu_L \times \nu_L$, it follows that we must have $((\pi \times \pi)x_0)(p \times p)\Lambda = X_L \times X_L$, which yields $(p \times p)\Lambda = L \times L$.

Given $h \in H$, we write $\alpha(h) = \widetilde{h}\Lambda_2$, where $\alpha : H \to H/\Lambda_2$ is the surjective homomorphism of Ratner's joinings theorem above. Since Λ_2 is a closed, normal subgroup of H, $H/\Lambda_2 := H_2$ is a connected Lie group in its own right, and we let $p_2 : H_2 \to L/p(\Lambda_2)$ be the projection induced by $p : H \to L$. Observe that we may identify $L/p(\Lambda_2) := L_2$ as a Levi subgroup of H_2.

Now, let $h \in H$, and let $x \in \xi_\lambda(\Gamma h)$ be a point in the Γh-fiber of λ. Then by Ratner's joinings theorem, there exists an index $i \in \{1, \ldots, n\}$ such that $x \in \Gamma c\beta_i\alpha(h) = \Gamma c\beta_i\widetilde{h}\Lambda_2$. It follows that

$$\pi_L(\xi_\lambda(\Gamma h)) = \{\pi_L(\Gamma c\beta_i\widetilde{h}\Lambda_2) : i = 1, \ldots, n\} = \{\Gamma_L c_L \delta_i p_2(\widetilde{h}\Lambda_2) : i = 1, \ldots, n\}.$$

where $c_L = \pi_L(c)$ and $\delta_i = p(\beta_i)$. Now, observe that $\alpha(R)$ is a connected, normal, solvable subgroup of H_2, and therefore that p_2 is trivial on $\alpha(R)$ since L_2 is semisimple, being a quotient of the semisimple group L. It follows that if we let $h = \ell r$ with $\ell \in L$ and $r \in R$, then

$$\begin{aligned} p_2(\alpha(h)) &= p_2(\alpha(\ell r)) \\ &= p_2(\alpha(\ell)\alpha(r)) \\ &= p_2(\alpha(\ell)) \\ &= p(\widetilde{\ell})p(\Lambda_2). \end{aligned}$$

Therefore,

$$\pi_L(\xi_\lambda(\Gamma h)) = \{\Gamma_L c_L \delta_i p(\widetilde{\ell})p(\Lambda_2) : i = 1, \ldots, n\}.$$

On the other hand, we have from above that $(p \times p)(\Lambda) = L \times L$, from which it follows that $\pi_L(\xi_\lambda(\Gamma h)) = X_L$ for any $h \in H$. Indeed, if we let $w \in X_L$, then there is some

$z \in X_H$ such that $(\pi \times \pi)(\Gamma h, z) = (y, w)$ and $(\Gamma h, z) \in \mathrm{supp}(\lambda) = x_0\Lambda$. It's then clear that $\pi(z) = w$ and $z \in \xi_\lambda(\Gamma h)$. We have shown that

$$\{\Gamma_L c_L \delta_i p(\widetilde{\ell}) p(\Lambda_2) : i = 1, \ldots, n\} = X_L,$$

and since $p(\Lambda_2)$ is a closed subgroup of L while $p(\widetilde{\ell}) \in L$ and $n \in \mathbb{N}$ are fixed, it follows that $p(\Lambda_2) = L$, as desired (otherwise, we would have written X_L as a finite union of lower-dimensional submanifolds). We immediately get the same conclusion for Λ_1, since by Ratner's joinings theorem, the groups H/Λ_1 and H/Λ_2 are isomorphic (under α). Hence, the triviality of the Levi factor $L/p(\Lambda_2)$ of H_2 implies that $L/p(\Lambda_1)$ must be trivial as well.

□

The following easy lemma will allow us to use this proposition to relate the action of (u^p, u^q) to the action of (u_L^p, u_L^q), where $u_L = p(u) \in L$, and to that of (u_R^p, u_R^q), where $u_R \in R$ is the unique element such that $u = u_L u_R$.

Lemma 2.3.4. *Let H be a connected Lie group with Levi decomposition $H = L \ltimes R$. If $N \subset H$ is a closed, connected, normal subgroup such that $p(N) = L$, then $L \subset N$.*

Proof. Since N is closed in H, it is a connected Lie group in its own right, and thus admits a Levi decomposition $N = S \ltimes Q$, where Q is the solvable radical of N. Observe that $R \cap N$ is a connected, normal, solvable subgroup of N, and therefore $R \cap N \subseteq Q$ since Q is the largest such subgroup. At the same time, since $p|_N : N \to L$ is surjective, the image $p(Q) \subset L$ is a normal, connected, solvable subgroup of L, and therefore must be trivial since L is semisimple. It follows that $Q \subset R = \ker(p)$, and therefore $Q \subseteq R \cap N$. Since we already had the reverse containment, this shows that $Q = R \cap N$.

It follows that $N = S \ltimes (R \cap N)$, with S a Levi subgroup of N. Therefore, we have $p(N) = p(S)$, and since $p(N) = L$ by hypothesis, we see that $\dim S \geq \dim L$. Since

L is a maximal semisimple subgroup of H, and since any two such subgroups are conjugate in H, it follows that S too is a Levi subgroup of H. Since N is normal in H, it follows that N contains all Levi subgroups of H, and in particular $L \subset N$. $\square$

Combining this lemma with the proposition before it, we find that $L \subset \Lambda_i$ for $i = 1, 2$. In particular, we have $u_L^p \in \Lambda_1$, and $u_L^q \in \Lambda_2$. Let $u_R^{(p)} = u_L^{-p} u^p$ and $u_R^{(q)} = u_L^{-q} u^q$. Then $u_R^{(p)}, u_R^{(q)} \in R$, and both are unipotent elements. We find

$$(u_L^p, u_L^q),\ (u_R^{(p)}, u_R^{(q)}) \in \Lambda = \mathrm{stab}_{H \times H} \lambda. \tag{2.3.4}$$

The presence of $(u_R^{(p)}, u_R^{(q)})$, rather than a pair of genuine powers of u_R, is unappealing, and can be easily rectified. Indeed, it is easy to see that

$$\begin{aligned} u_R^{(p)} &= \prod_{k=0}^{p-1} u_L^{-k} u_R u_L^k \\ &= u_R^p \cdot \prod_{k=0}^{p-1} u_R^{-1} u_L^{-k} u_R u_L^k. \end{aligned}$$

Since Λ_i is normal in H and $L \subset \Lambda_i$, we have $u_L^k \in \Lambda_i$ and $u_R^{-1} u_L^{-k} u_R \in \Lambda_i$ for all $k \in \mathbb{Z}, i = 1, 2$. Therefore

$$\prod_{k=0}^{p-1} u_R^{-1} u_L^{-k} u_R u_L^k \in \Lambda_1 \text{ and } \prod_{k=0}^{q-1} u_R^{-1} u_L^{-k} u_R u_L^k \in \Lambda_2,$$

and so we find

$$\left(\prod_{k=0}^{p-1} u_R^{-1} u_L^{-k} u_R u_L^k, \prod_{k=0}^{q-1} u_R^{-1} u_L^{-k} u_R u_L^k \right) \in \Lambda.$$

But, on combining this with the above expression for $u_R^{(p)}$ and (2.3.4), we see that we also have

$$(u_R^p, u_R^q) = (u_R^{(p)}, u_R^{(q)}) \cdot \left(\prod_{k=0}^{p-1} u_R^{-1} u_L^{-k} u_R u_L^k, \prod_{k=0}^{q-1} u_R^{-1} u_L^{-k} u_R u_L^k \right)^{-1} \in \Lambda.$$

We pause to formally record these points.

Proposition 2.3.5. *Let all notation be as above for the connected, simply connected Lie group H. Then the following holds for all but finitely many pairs of primes $p \neq q$: if λ is an ergodic joining of the systems (X_H, u^p, ν_H) and (X_H, u^q, ν_H) which is supported on the closure of the (u^p, u^q) orbit of $(\Gamma e, \Gamma e)$, then we have both*

$$(u_L^p, u_L^q) \in stab_{H\times H}\lambda$$

and

$$(u_R^p, u_R^q) \in stab_{H\times H}\lambda.$$

The short exact sequence

$$1 \to R \to H \xrightarrow{p} L \to 1$$

gives rise to a fibration

$$1 \to X_R \to X_H \xrightarrow{\pi} X_L \to 1,$$

hence also a fibration

$$1 \to X_R \times X_R \to X_H \times X_H \xrightarrow{\pi\times\pi} X_L \times X_L \to 1.$$

For every $y \in X_L$, we may identify the fiber $\mathcal{F}_y := \pi^{-1}(\{y\})$ with $X_R = \Gamma_R^\xi \backslash R$: fix once and for all a Borel measurable section $\sigma : X_L \to L$ of the projection $L \to X_L$, the existence of which is guaranteed by the results of [Mac]. This section need not be differentiable or even continuous, and the reader may rightly worry that this will cause problems in our analysis, since Ratner's theorems only apply as stated to continuous functions, and moreover we will be using aspects of harmonic analysis on

the nilmanifold X_R that only apply to continuous functions. However, the reason we are safe to work with a section that is only measurable is that we have, at this point, already employed the continuity of f and the genericity of the point x for ν_H to obtain the joining λ. All of our work will now proceed relative to λ and the various Haar measures involved, at which level the distinction between continuity and mere measurability is not harmful.

That said, the continuity of f (and in fact its differentiablility, which we will assume shortly), will be essential to our arguments involving harmonic analysis, since we will be examining the fine properties of f in the X_R fibers mentioned above. The point is that it is only the properties of f in ν_L-almost all of these fibers that are relevant, once we take averages against the various pertinent measures. We will be clear in what follows at which points we are exploiting the continuity or differentiability of f, versus the ability to restrict our attention to properties that only hold almost everywhere with respect to some invariant measure of interest.

We will write $\sigma(y) = \widetilde{y} \in L$ from here on. Then the map

$$\begin{aligned} \varphi_y : X_R &\to \mathcal{F}_y \\ \Gamma_R r &\mapsto \Gamma r \widetilde{y} \end{aligned}$$

is clearly well-defined and is a homeomorphism, which explicitly identifies the fibers of π_L with X_R.

The following useful fact, due to Witte Morris, will allow us to apply an inductive argument to prove Möbius disjointness, along the lines of [GT] and [Z].

Proposition 2.3.6 ([Wi] Proposition 2.6)**.** *Suppose that H is a connected Lie group and $\Gamma \subset H$ is a lattice, such that $\Gamma\backslash H$ supports an ergodic unipotent translation. Then $radH$ is nilpotent.*

Since we are beginning with the assumption that $\Gamma\backslash H$ supports an ergodic unipo-

tent translation, we see that R is nilpotent. Our argument will proceed by induction on the step d of R; thus the base case $d = 1$ is that in which R is abelian.

The following version of the Mautner phenomenon, due to Ratner, allows us to reduce to the case in which the action of u_R on X_R is ergodic.

Proposition 2.3.7 ([R2] Proposition 1.5)**.** *Let H be a connected, simply connected Lie group whose radical is nilpotent and which has no compact semisimple factors, and let Γ be a lattice in H. Let U be a connected unipotent subgroup of H and N the smallest closed connected normal subgroup of H such that $U \subset N$ and $N \cap \Gamma$ is a lattice in N. Then ν_H-almost every ergodic component of the action of U on $(\Gamma\backslash H, \nu_H)$ is an N-orbit.*

Taking $U = \{\exp(tv_R) : t \in \mathbb{R}\}$, where $v_R = \log u_R$, the ergodic components of the action of U on (X_H, ν_H) agree with those of the action of u_R. Since $u_R \in R$, R is a closed, connected, normal subgroup of H by construction, and $\Gamma \cap R$ is a lattice in R by the above remarks, we see that $N \subset R$, where N is the subgroup of the proposition. Replacing H with the smaller group $L \ltimes N$, we may assume that ν_H-a.e. ergodic component of the action of u_R on (X_H, ν_H) is an R-orbit. In particular, we may assume that the action of u_R on X_R is ergodic (for the Haar measure ν_R), and that the ergodic decomposition of ν_H for the u_R action agrees with the disintegration of ν_H over $\pi : X_H \to X_L$:

$$\nu_H = \int_{X_L} \nu_y \, \mathrm{d}\nu_L(y),$$

where ν_y is supported on $\mathcal{F}_y$ and satisfies $(\varphi_y^{-1})_* \nu_y = \nu_R$, for ν_L-a.e. $y \in X_L$.

This ergodicity now implies the following.

Proposition 2.3.8. *For ν_L-a.e. $y \in X_L$, the action of $\widetilde{y} u_R \widetilde{y}^{-1}$ on X_R is ergodic relative to ν_R, and therefore uniquely and totally ergodic.*

Proof. By Proposition 2.3.6, R is nilpotent. Thus we may apply a theorem of Leon Green ([AGH]), which asserts that a rotation on a nilmanifold is ergodic if and only if

its projection to the associated horizontal torus is ergodic. Thus, it suffices to assume that X_R is a torus. Let

$$E = \left\{ \ell \in L : \text{the action of } \ell u_R \ell^{-1} \text{ on } (X_R, \nu_R) \text{ is not ergodic} \right\}.$$

By the ergodicity criterion for a rotation on a torus, we have

$$E = \bigcup_{\substack{\chi \in \widehat{X}_R \\ \chi \neq \chi_0}} E_\chi$$

where

$$E_\chi = \left\{ \ell \in L : \chi(\ell u_R \ell^{-1}) = 1 \right\}$$

and χ_0 is the trivial character. Observe that E_χ is an analytic subvariety of L. Moreover, no E_χ contains the identity element of L, since u_R acts ergodically. Therefore, each E_χ is a *proper* analytic subvariety of L, and thus has Haar measure 0 in L. Since $\widehat{X_R}$ is countable, it follows that the Haar measure of E is 0 as well. But it's then immediate that

$$\nu_L \left(\left\{ y \in X_L : \text{the action of } \widetilde{y} u_R \widetilde{y}^{-1} \text{ on } (X_R, \nu_R) \text{ is not ergodic} \right\} \right) = 0,$$

as desired.

Finally, the statements of unique and total ergodicity are true in general for ergodic nilrotations (cf. [Fra] page 10). □

Notice that for any $r, r' \in R$ we have

$$\varphi_y(\Gamma_R r r') = \Gamma r r' \widetilde{y} = \Gamma r \widetilde{y} (\widetilde{y}^{-1} r' \widetilde{y}) = \varphi_y(\Gamma_R r) \widetilde{y}^{-1} r' \widetilde{y}, \tag{2.3.5}$$

and $\widetilde{y}^{-1} r \widetilde{y} \in R$ since R is normal in H.

Now observe that for any point $y \in X_L$ and any $r \in R$, we have $\Gamma r\widetilde{y}u^p = \Gamma r(\widetilde{y}u^p u_L^{-p}\widetilde{y}^{-1})\widetilde{y}u_L^p$, and $\widetilde{y}u^p u_L^{-p}\widetilde{y}^{-1} \in R$ because $p(u) = u_L = p(u_L)$ and $R = \ker p$. Since this point clearly belongs to the fiber of π over $\Gamma_L\widetilde{y}u_L^p$, we see that

$$\varphi_y^{-1}\left(\Gamma r\widetilde{y}u_R^p\right) = \varphi_y^{-1}\left(\Gamma r\left(\widetilde{y}u_R^p\widetilde{y}^{-1}\right)\widetilde{y}\right) = \Gamma_R r\left(\widetilde{y}u_R^p\widetilde{y}^{-1}\right). \tag{2.3.6}$$

Let

$$\lambda = \int_{X_L\times X_L} \lambda_{(y,z)}\,\mathrm{d}(\nu_L\times\nu_L)(y,z) \tag{2.3.7}$$

be the disintegration of λ with respect to $\pi\times\pi : X_H\times X_H \to X_L\times X_L$, where we've used the assumption that $(\pi\times\pi)_*\lambda = \nu_L\times\nu_L$. For $\nu_L\times\nu_L$-almost every pair $(y,z)\in X_L\times X_L, \lambda_{(y,z)}$ is a Borel probability measure on $X_H\times X_H$ supported on the fiber $\mathcal{F}_{(y,z)} = (\pi\times\pi)^{-1}\{(y,z)\} = \mathcal{F}_y\times\mathcal{F}_z$.

Observe that (u^p,u^q) maps $\mathcal{F}_{(y,z)}$ to $\mathcal{F}_{(yu_L^p, zu_L^q)}$. Since λ is (u^p,u^q)-invariant, we have for any $(y,z)\in X_L\times X_L$

$$(u^p,u^q)_*\lambda_{(y,z)} = \lambda_{(yu_L^p,zu_L^q)}. \tag{2.3.8}$$

We showed above (Proposition 2.3.5) that the Levi components (u_L^p,u_L^q) and (u_R^p,u_R^q) also belong to Λ, the stabilizer of λ. Since R acts trivially on X_L, we therefore have

$$\left(u_R^p,u_R^q\right)_*\lambda_{(y,z)} = \lambda_{(y,z)}$$

for $(\nu_L\times\nu_L)$-a.e. (y,z), so almost every $\lambda_{(y,z)}$ is invariant under the action of (u_R^p,u_R^q) on $\mathcal{F}_{(y,z)}$. It follows from (2.3.6) that the image measure $\rho_{(y,z)} = (\varphi_{(y,z)}^{-1})_*\lambda_{(y,z)}$ on $X_R\times X_R$ is invariant for the action of $(\widetilde{y}u_R^p\widetilde{y}^{-1},\widetilde{z}u_R^q\widetilde{z}^{-1})$.

We now show that we may assume without loss of generality that almost every $\rho_{(y,z)}$ is *ergodic* for this action. Naturally, we would like to demonstrate this by means of the ergodic decomposition of each of the (u_R^p,u_R^q)-invariant measures $\lambda_{(y,z)}$. The

problem with this idea is that the measurable partition of $X_H \times X_H$ to which this decomposition gives rise a priori depends on (y, z), creating complications in the order of logic should we attempt to average with respect to λ. However, this (y, z) dependence may be removed in the following way.

Since λ is ergodic with respect to the action of (u^p, u^q), it is algebraic by Ratner's measure classification theorem, and thus the probability space $(X_H \times X_H, \lambda)$ may be identified with the homogeneous space (X_Λ, ν_Λ), where $\Lambda = \mathrm{stab}_{H\times H}\lambda$ as above, $X_\Lambda = (\Lambda \cap \Gamma \times \Gamma)\backslash\Lambda$, and ν_Λ is a Λ-invariant Borel probability measure on X_Λ. Since $(u_R^p, u_R^q) \in \Lambda$ by Proposition 2.3.5, we have by Proposition 2.3.7 that λ-a.e. ergodic component of the action of (u_R^p, u_R^q) on $(X_H \times X_H, \lambda)$ is an N-orbit, where N is the smallest closed connected normal subgroup of Λ such that $(u_R^p, u_R^q) \in N$ and $N \cap \Gamma \times \Gamma$ is a lattice in N.

We clearly have $N \subset \Lambda \cap R \times R$; thus, if $(x_1, x_2) \in X_H \times X_H$, then the N-orbit $(x_1, x_2)N$ is contained in the fiber $\mathcal{F}_{(y_1,y_2)}$ where $y_i = \pi_L(x_i)$. It follows that the measurable partition determined by the ergodic decomposition of each $\lambda_{(y,z)}$ is given by the partition of $X_H \times X_H$ into N-orbits, which does not depend on (y, z) (note, however, that the relevant ergodic measures on the atoms of this partition do vary with $\lambda_{(y,z)}$).

To summarize, there is a measurable partition $\xi = \{C(x, x') : (x, x') \in X_H \times X_H\}$ of $X_H \times X_H$ into (u_R^p, u_R^q)-invariant subsets $C(x, x') \subset X_H \times X_H, (x, x') \in C(x, x')$, a probability measure P on $X_H \times X_H/\xi$, and ergodic (u_R^p, u_R^q)-invariant Borel probability measures $\{\lambda_{(y,z)}^C : C \in \xi, (y, z) \in X_L \times X_L\}$ such that

$$\lambda_{(y,z)}(E) = \int_{X_H\times X_H/\xi} \lambda_{(y,z)}^C(E \cap C)\,\mathrm{d}P(C)$$

for every measurable subset $E \subset X_H \times X_H$. It follows that

$$\begin{aligned}\int_{X_H\times X_H} f\otimes f\,\mathrm{d}\lambda &= \int_{X_L\times X_L}\int_{X_H\times X_H} f\otimes f\,\mathrm{d}\lambda_{(y,z)}\,\mathrm{d}(\nu_L\times\nu_L)(y,z)\\ &= \int_{X_L\times X_L}\int_{X_H\times X_H/\xi}\int_{X_H\times X_H} f\otimes f\,\mathrm{d}\lambda^C_{(y,z)}\,\mathrm{d}P(C)\,\mathrm{d}(\nu_L\times\nu_L)(y,z)\\ &= \int_{X_H\times X_H/\xi}\int_{X_L\times X_L}\int_{X_H\times X_H} f\otimes f\,\mathrm{d}\lambda^C_{(y,z)}\,\mathrm{d}(\nu_L\times\nu_L)(y,z)\,\mathrm{d}P(C).\end{aligned} \tag{2.3.9}$$

This freedom to change the order of integration is contingent on the independence of the partition ξ from (y,z), which is the reason for our arguments above.

Our proof of Möbius disjointness in the case that R is abelain will make the crucial assumption that f is in $C^1(X_H)$. This causes no loss of generality at the level of our proof of Theorem 2.0.3, since it should be clear (and we will show) that we can prove this theorem for arbitrary continuous functions once we have it for differentiable functions, by uniformly approximating the former class. This differentiability will enter our analysis by allowing us to control the decay of Fourier coefficients when we use harmonic analysis in the fibers of $\pi_L\times\pi_L$ (which, we recall from above, have been explicitly identified with $X_R\times X_R$).

Observe that since the fiber measure $\lambda^C_{(y,z)}$ is ergodic for (u_R^p,u_R^q) for $\nu_L\times\nu_L$-a.e. $(y,z)\in X_L\times X_L$ and P-a.e. $C\in X_H\times X_H/\xi$, (2.3.6) implies that $\rho^C_{(y,z)} = (\varphi_y^{-1}\times\varphi_z^{-1})_*\lambda^C_{(y,z)}$ is ergodic for $(\widetilde{y}u_R^p\widetilde{y}^{-1},\widetilde{z}u_R^q\widetilde{z}^{-1})$ (with the same caveats on (y,z) and C).

Given a continuous, bounded function $f\in C_b(X_H)$ and $y\in X_L$, we may define a continuous, bounded function f_y on X_R by setting $f_y(\Gamma_R r) = f(\varphi_y(\Gamma_R r)) = f(\Gamma r\widetilde{y})$. We may now summarize the above discussion with the following.

Proposition 2.3.9. *Let H be a connected, simply connected real Lie group, $\Gamma\subset H$ a lattice, and $u\in H$ an Ad-unipotent element, such that the u-orbit of the identity coset $x=\Gamma e\in X_H$ is uniformly distributed with respect to ν_H. Let $H = L\rtimes R$ be*

the real Levi decomposition of H. Then the following holds for all but finitely many pairs of primes $p \neq q$: for any ergodic joining λ of the systems (X_H, u^p, ν_H) and (X_H, u^q, ν_H) which is supported on the orbit closure $\overline{\{(x,x)(u^{pn}, u^{qn})\}_{n\in\mathbb{N}}}$, there exist a probability space (Z, P) and a family $\left\{\rho^C_{(y,z)} : (y,z) \in X_L \times X_L, C \in Z\right\}$ of Borel probability measures on $X_R \times X_R$ such that for any continuous, bounded function f on $\Gamma\backslash H$, we have

$$\int_{X_H\times X_H} f \otimes f \, d\lambda = \int_Z \int_{X_L\times X_L} \int_{X_R\times X_R} f_y \otimes f_z \, d\rho^C_{(y,z)} \, d(\nu_L \times \nu_L)(y,z) \, dP(C).$$

Moreover, for $(\nu_L \times \nu_L)$-a.e. $(y,z) \in X_L \times X_L$ and P-a.e. $C \in Z$, $\rho^C_{(y,z)}$ is ergodic for the action of $(\widetilde{y}u_R^p\widetilde{y}^{-1}, \widetilde{z}u_R^q\widetilde{z}^{-1})$ on $X_R \times X_R$.

Our interest in this theorem arises from (2.3.1), which tells us that the left side of the equation in the theorem governs the behavior of correlations of u^p and u^q, and thereby is connected to Möbius disjointness by Theorem 2.0.4. We now show how this works in practice, beginning with the case that R is abelian.

2.3.1 The case of abelian R

We apply the formula of Proposition 2.3.9:

$$\int_{X_H\times X_H} f\otimes f \, \mathrm{d}\lambda = \int_Z \int_{X_L\times X_L} \int_{X_R\times X_R} f_y \otimes f_z \, \mathrm{d}\rho^C_{(y,z)} \, \mathrm{d}(\nu_L\times\nu_L)(y,z) \, \mathrm{d}P(C). \quad (2.3.10)$$

For each $y \in X_L$, f_y is a continuous function on the compact abelian group X_R, and therefore admits a Fourier expansion

$$f_y(u) = \sum_{\chi\in\widehat{X}_R} \widehat{f}_y(\chi)\chi(u).$$

We assume furthermore that $f \in C^1(X_H)$, so that $f_y \in C^1(X_R)$ and this series converges pointwise. This causes no loss of generality for our proof of Möbius disjointness,

as we will show below.

Since $\rho^C_{(y,z)}$ is a probability measure, the bounded convergence theorem now allows us to exchange the order of sums and integrals for the expression inside the integral over Z:

$$\begin{aligned}
&\int_{X_L\times X_L}\int_{X_R\times X_R} f_y\otimes f_z\,\mathrm{d}\rho^C_{(y,z)}\,\mathrm{d}(\nu_L\times\nu_L)(y,z)\\
&\quad=\int_{X_L\times X_L}\int_{X_R\times X_R}\sum_{\chi,\psi\in\widehat{X}_R}\widehat{f}_y(\chi)\widehat{f}_z(\psi)\chi\otimes\psi\,\mathrm{d}\rho^C_{(y,z)}\mathrm{d}(\nu_L\times\nu_L)(y,z)\\
&\quad=\sum_{\chi,\psi\in\widehat{X}_R}\int_{X_L\times X_L}\widehat{f}_y(\chi)\widehat{f}_z(\psi)\int_{X_R\times X_R}\chi\otimes\psi\,\mathrm{d}\rho^C_{(y,z)}\,\mathrm{d}(\nu_L\times\nu_L)(y,z). \qquad (2.3.11)
\end{aligned}$$

By Theorem 2.3.9, $\rho^C_{(y,z)}$ is ergodic for a unipotent translation on $X_R\times X_R$, and therefore is algebraic by Ratner's measure classification theorem, meaning it is supported on an orbit of $\mathrm{stab}_{(R\times R)}\,\rho^C_{(y,z)} := S^C_{(y,z)}$. Moreover, the point (r_C, s_C) on whose $S^C_{(y,z)}$ orbit the measure $\rho^C_{(y,z)}$ is supported does not depend on (y,z). This is because, following our discussion above, the ergodic decomposition of almost every $\lambda_{(y,z)}$ for the action of (u_R^p, u_R^q) on $X_H\times X_H$ equals the partition into N-orbits, where N is determined solely by λ. Thus, the points (r_C, s_C) parametrize these N-orbits in the fibers of $\pi_L\times\pi_L$, which we have identified with $X_R\times X_R$. As we have $N\subset R\times R$, the points $(r_C, s_C)\in X_R\times X_R$ are determined by the partition of $X_R\times X_R$ into N-orbits (each C is an N-orbit, and we have $(r_C, s_C)\in C$), with no dependence on (y,z).

Consequently, we have

$$\int_{X_R\times X_R}\chi(r)\psi(s)\,\mathrm{d}\rho^C_{(y,z)} = \begin{cases}\chi(r_C)\psi(s_C) & \text{if } (\chi\otimes\psi)|_{S^C_{(y,z)}} \text{ is trivial}\\ 0 & \text{else.}\end{cases}$$

Since $S^C_{(y,z)}$ contains $(\widetilde{y}u_R^p\widetilde{y}^{-1}, \widetilde{z}u_R^q\widetilde{z}^{-1})$, and since $\rho^C_{(y,z)}$ is ergodic for the action

thereof, we see that

$$\int_{X_R\times X_R} \chi\otimes\psi\,\mathrm{d}\rho^C_{(y,z)} = \begin{cases} \chi(r_C)\psi(s_C) & \text{if } \chi(\widetilde{y}u_R^p\widetilde{y}^{-1})\psi(\widetilde{z}u_R^q\widetilde{z}^{-1})=1 \\ 0 & \text{if not.}\end{cases}$$

Therefore, (2.3.10) and (2.3.11) together yield the following formula:

$$\int_{X_H\times X_H} f\otimes f\,\mathrm{d}\lambda = \sum_{\chi,\psi\in\widehat{X}_R}\int_Z \chi(r_C)\psi(s_C)\int_{X_L^{p,q}(\psi,\chi)}\widehat{f}_y(\chi)\widehat{f}_z(\psi)\,\mathrm{d}(\nu_L\times\nu_L)(y,z)\,\mathrm{d}P(C) \tag{2.3.12}$$

where

$$X_L^{p,q}(\chi,\psi)=\left\{(y,z)\in X_L\times X_L : \chi(\widetilde{y}u_R^p\widetilde{y}^{-1})\psi(\widetilde{z}u_R^q\widetilde{z}^{-1})=1\right\}.$$

Now, introduce the set

$$L^{p,q}(\psi,\chi)=\left\{(\ell_1,\ell_2)\in L\times L : \chi(\ell_1u_R^p\ell_1^{-1})\psi(\ell_2u_R^q\ell_2^{-1})=1\right\}.$$

Observe that $L^{p,q}(\psi,\chi)$ is an analytic subvariety of $L\times L$, since the group operations in H and the characters ψ,χ on R are analytic. However, any analytic subvariety of $L\times L$ which is not equal to $L\times L$ has Haar measure 0 (it is well-known that in any system of local coordinates the Haar measure is absolutely continuous with respect to the Lebesgue measure). Since $X_L^{p,q}(\psi,\chi)\subseteq(\pi\times\pi)(L^{p,q}(\psi,\chi))$, where $\pi: L\to X_L$ is the projection, we see that we have the following dichotomy: either

- $L^{p,q}(\psi,\chi)=L\times L$, so $X_L^{p,q}(\psi,\chi)=X_L\times X_L$ and $(\nu_L\times\nu_L)(X_L^{p,q}(\psi,\chi))=1$, or

- $L^{p,q}(\psi,\chi)\neq L\times L$, and then $(\nu_L\times\nu_L)(X_L^{p,q}(\psi,\chi))=0$.

Consequently, (2.3.12) becomes

$$\int_{X_H\times X_H} f\otimes f\,\mathrm{d}\lambda = \sum_{\substack{\chi,\psi\\ L^{p,q}(\psi,\chi)=L\times L}} \widehat{f}(\chi)\widehat{f}(\psi)\int_Z \chi(r_C)\psi(s_C)\,\mathrm{d}P(C) \tag{2.3.13}$$

where

$$\widehat{f}(\chi) = \int_{X_L} \widehat{f_y}(\chi)\,\mathrm{d}y.$$

Observe, however, that if $L^{p,q}(\psi,\chi) = L\times L$ then we have in particular $\chi(u_R^p)\psi(u_R^q) = 1$, which implies by the ergodicity of u_R acting on X_R that $\chi^p = \psi^{-q}$. As this will turn out to be the central point of our argument below, we write (2.3.13) as

$$\int_{X_H\times X_H} f\otimes f\,\mathrm{d}\lambda = \sum_{\substack{\chi,\psi\\ \chi^p=\psi^{-q}}}{}' \widehat{f}(\chi)\widehat{f}(\psi)\int_Z \chi(r_C)\psi(s_C)\,\mathrm{d}P(C), \tag{2.3.14}$$

where the primed summation is restricted to those pairs of characters such that $L^{p,q}(\psi,\chi) = L\times L$.

We are now in a position to prove Möbius disjointness in the abelian case, yielding the base of induction on the step of the nilradical of H.

Proof of Theorem 2.0.3 for the case of abelian R. First, we may suppose without loss of generality that $\int_{X_H} f\,\mathrm{d}\nu_H = 0$. Indeed, for any constant c we have

$$\sum_{n\le N}\mu(n)f(xu^n) = \sum_{n\le N}\mu(n)(f(xu^n)-c) + c\sum_{n\le N}\mu(n) = \sum_{n\le N}\mu(n)(f(xu^n)-c) + o(N)$$

by the prime number theorem. Since $x\mapsto f(x)-c$ remains a continuous, bounded function on X_H, all of our work above applies to this function in place of f, with $c = \int_{X_H} f\,\mathrm{d}\nu_H$.

Furthermore, we may assume that f belongs to $C^1(X_H)$. Indeed, if f is an arbitrary continuous function, and f_k a sequence of C^1 functions which converge to f

uniformly, then for any $\epsilon > 0$ we have for $N \geq 1$ and all sufficiently large k

$$\frac{1}{N}\sum_{n\leq N}\mu(n)f(xu^n) = \frac{1}{N}\sum_{n\leq N}\mu(n)f_k(xu^n) + O(\epsilon).$$

If Möbius disjointness is known for C^1 functions, we may choose k so large that the first term above is smaller than ϵ as well, making it clear that limit of the expression on the left as $N \to \infty$ must be zero.

Now, notice that

$$\begin{aligned}\widehat{f}(\chi_0) &= \int_{X_L} \widehat{f_y}(\chi_0)\,\mathrm{d}y \\ &= \int_{X_L}\int_{X_R} f(\Gamma r\widetilde{y})\,\mathrm{d}r\,\mathrm{d}y \\ &= \int_{X_H} f\,\mathrm{d}\nu_H \\ &= 0.\end{aligned}$$

Therefore, the expression on the right side of (2.3.14) becomes

$$\sum_{\substack{\chi,\psi\\ \chi^p=\psi^{-q}}}{}' \widehat{f}(\chi)\widehat{f}(\psi)\int_Z \chi(r_C)\psi(s_C)\,\mathrm{d}P(C) = \sum_{\substack{\chi,\psi\neq\chi_0\\ \chi^p=\psi^{-q}}}{}' \widehat{f}(\chi)\widehat{f}(\psi)\int_Z \chi(r_C)\psi(s_C)\,\mathrm{d}P(C).$$

Since R is simply connected and abelian, it is isomorphic to $\mathbb{R}^k$, where $k = \dim R$. Therefore X_R is isomorphic to the torus $\mathbb{T}^k$, and $\widehat{X_R}$ may be identified with $\mathbb{Z}^k$, from which it is easily seen that

$$\sum_{\substack{\chi,\psi\neq\chi_0\\ \chi^p=\psi^{-q}}}{}' \widehat{f}(\chi)\widehat{f}(\psi)\int_Z \chi(r_C)\psi(s_C)\,\mathrm{d}P(C) = \sum_{\chi\neq\chi_0}{}' \widehat{f}(\chi^{-q})\widehat{f}(\chi^p)\int_Z \chi^{-q}(r_C)\chi^p(s_C)\,\mathrm{d}P(C).$$

By using the bounded convergence theorem to move the sum past the integral over

Z and the integral over X_L in the definition of $\widehat{f}(\chi)$, we find in all

$$\begin{aligned}\int_{X_H\times X_H} f\otimes f\,\mathrm{d}\lambda &= \sum_{\chi\neq\chi_0}{}' \widehat{f}(\chi^{-q})\widehat{f}(\chi^p)\int_Z \chi^{-q}(r_C)\chi^p(s_C)\,\mathrm{d}P(C)\\ &= \int_Z\int_{X_L\times X_L}\sum_{\chi\neq\chi_0}{}' \chi^{-q}(r_C)\chi^p(s_C)\widehat{f}_y(\chi^{-q})\widehat{f}_z(\chi^p)\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}P(C).\end{aligned} \tag{2.3.15}$$

For each $y\in X_L$, the functions $\chi\mapsto\widehat{f}_y(\chi^{-q})$ and $\chi\mapsto\widehat{f}_y(\chi^p)$ belong to $\ell^2(\widehat{X}_R)$ by Plancherel's theorem. The same is then true of the functions $\chi\mapsto\widehat{f}_y(\chi^{-q})\chi^{-q}(r_C)$ and $\chi\mapsto\widehat{f}_y(\chi^p)\chi^p(s_C)$ since $|\chi|\equiv 1$. Thus, for each $y,z\in X_L$ and $C\in Z$, we have by the Cauchy-Schwartz inequality

$$\left|\sum_{\chi\neq\chi_0}{}' \chi^{-q}(r_C)\chi^p(s_C)\widehat{f}_y(\chi^{-q})\widehat{f}_z(\chi^p)\right|^2 \leq \sum_{\chi\neq\chi_0}{}' \left|\widehat{f}_y(\chi^{-q})\right|^2 \sum_{\chi\neq\chi_0}{}' \left|\widehat{f}_z(\chi^p)\right|^2,$$

and the sums on the right side both converge. Note that this estimate does not depend on C. Hence, on combining this with (2.3.15) and using the triangle inequality to eliminate the integral over Z, we get the estimate

$$\left|\int_{X_H\times X_H} f\otimes f\,\mathrm{d}\lambda\right| \leq \int_{X_L\times X_L}\left(\sum_{\chi\neq\chi_0}\left|\widehat{f}_y(\chi^{-q})\right|^2\sum_{\chi\neq\chi_0}\left|\widehat{f}_z(\chi^p)\right|^2\right)^{1/2}\mathrm{d}y\,\mathrm{d}z, \tag{2.3.16}$$

where the positivity allows us to remove the prime on the summation.

We now use the standard decay of Fourier coefficients for C^1 functions. According to this, since each f_y is differentiable on X_R, we have for each $m=(m_1,\ldots,m_k)\in\mathbb{Z}^k$ (now identifying $\widehat{X_R}$ with $\mathbb{Z}^k$, so that $\widehat{f}_y(m)$ makes sense)

$$|\widehat{f}_y(m)| \leq c_k\frac{\max(\|f_y\|_{L^1},\max_{i=1,\ldots,k}\|\frac{\partial}{\partial x_i}f_y\|_{L^1})}{1+|m|},$$

where $c_k>0$ depends only on k, and $(x_1,\ldots,x_k)$ are the coordinates on $\mathbb{T}^k$. Let

$M(f_y) = \max(\|f_y\|_{L^1}, \max_{i=1,\dots,k} \|\frac{\partial}{\partial x_i} f_y\|_{L^1})$. Then we see

$$\begin{aligned}\sum_{m\neq 0} \left|\widehat{f_y}(qm)\right|^2 &\leq c_k^2 M(f_y)^2 \sum_{m\neq 0} \frac{1}{(1+|qm|)^2} \\ &= \frac{c_k^2 M(f_y)^2}{q^2} \sum_{m\neq 0} \frac{1}{(q^{-1}+|m|)^2} \\ &\leq C_k M(f_y)^2 q^{-2},\end{aligned}$$

where C_k is c_k^2 multiplied by the value of the convergent sum $\sum_{m\neq 0\in\mathbb{Z}^k} |m|^{-2}$.

Plugging this inequality into (2.3.16), we find

$$\left|\int_{X_H\times X_H} f\otimes f \,\mathrm{d}\lambda\right| \leq \frac{C_k^2}{pq} \int_{X_L\times X_L} M(f_y)M(f_z)\,\mathrm{d}y\,\mathrm{d}z. \tag{2.3.17}$$

Our only remaining obstacle is to show that the integral on the right side of this expression converges. Observe that

$$\|f_y\|_{L^1} = \int_{X_R} |f(\Gamma r\widetilde{y})|\,\mathrm{d}\nu_R \leq \|f\|_\infty,$$

so $M(f_y) \leq \max(\|f\|_\infty, \max_{i=1,\dots,k} \|\frac{\partial}{\partial x_i} f_y\|_{L^1(X_R)})$. At the same time, since $\pi : X_H \to X_L$ is a surjective submersion, we may introduce local coordinates on X_H around x in such a way that the coordinates of x are $(x_1,\dots,x_k,y_1,\dots,y_s)$, where $(x_1,\dots,x_k)$ are coordinates on $\mathbb{T}^k$ and $(y_1,\dots,y_s)$ are the coordinates of $y=\pi(x)\in X_L$. In this system of local coordinates, we have

$$\frac{\partial}{\partial x_i} f_y(\Gamma_R r) = \frac{\partial}{\partial x_i} f(x),$$

where $x = \Gamma r\widetilde{y}$. Since f is assumed to be differentiable, we have $(\partial/\partial x_i)f \in L^1(X_H)$;

thus, we find

$$\begin{aligned}\int_{X_L} \left\| \frac{\partial}{\partial x_i} f_y \right\|_{L^1(X_R)} \mathrm{d}y &= \int_{X_L} \int_{X_R} \left| \frac{\partial}{\partial x_i} f_y(\Gamma_R r) \right| \mathrm{d}\nu_R \, \mathrm{d}y \\ &= \int_{X_H} \left| \frac{\partial}{\partial x_i} f(x) \right| \mathrm{d}x \\ &= \left\| \frac{\partial}{\partial x_i} f \right\|_{L^1(X_H)}.\end{aligned}$$

Consequently, we have

$$\begin{aligned}\int_{X_L} M(f_y) \, \mathrm{d}y &\leq \max \left(\|f\|_\infty, \max_{i=1,\ldots,k} \int_{X_L} \left\| \frac{\partial}{\partial x_i} f_y \right\|_{L^1(X_R)} \mathrm{d}y \right) \\ &= \max \left(\|f\|_\infty, \max_{i=1,\ldots,k} \left\| \frac{\partial}{\partial x_i} f \right\|_{L^1(X_H)} \right) < \infty.\end{aligned}$$

Thus, (2.3.17) finally yields

$$\left| \int_{X_H \times X_H} f \otimes f \, \mathrm{d}\lambda \right| \leq \frac{1}{pq} C_f, \tag{2.3.18}$$

where $C_f > 0$ only depends on f, not on p or q. Putting things together, we see that if $\tau > 0$ and $p, q > (\tau C_f)^{-1/2}$, then

$$\left| \int_{X_H \times X_H} f \otimes f \, \mathrm{d}\lambda \right| < \tau.$$

Since $(\tau C_f)^{-1/2}$ is vanishingly small compared to $e^{1/\tau}$, Theorem 2.0.4 therefore implies that

$$\sum_{n \leq N} \mu(n) f(x u^n) = o(N),$$

as desired.

□

2.3.2 The case of general R

We now prove Theorem 2.0.3 when R, the radical of H, is not necessarily abelian (recall that R is nilpotent by Proposition 2.3.6). Our argument proceeds by induction on d, the step of the nilpotent group R, with the base case $d = 1$ carried out above. In order to mimic the proof in the abelian case, we will use harmonic analysis on the nilpotent group R, and specifically the notions of "vertical Fourier analysis" ([GT],[Zo]). Let R be a nilpotent group of nilpotency class d (i.e. its descending central series has length $d+1$), and $\Gamma \subset R$ a lattice. A continuous function g on the nilmanifold $\Gamma\backslash R$ is called a vertical character if there exists a character χ of the compact abelian group $\Gamma \cap R_d\backslash R_d = \Gamma_d\backslash R_d$ such that

$$g(vr_d) = g(v)\chi(\Gamma_d r_d) \text{ for all } v \in \Gamma\backslash R, r_d \in R_d.$$

Given any continuous function g on $\Gamma\backslash R$ and character $\chi \in \widehat{\Gamma_d\backslash R_d}$, we define

$$g^\chi(v) := \int_{\Gamma_d\backslash R_d} g(vr_d)\overline{\chi(r_d)}\,\mathrm{d}r_d.$$

Then g^χ is a vertical character of $\Gamma\backslash R$, with χ the associated character of $\Gamma_d\backslash R_d$, and we have the vertical Fourier series

$$g = \sum_{\chi \in \widehat{\Gamma_d\backslash R_d}} g^\chi \text{ in } L^2(\Gamma\backslash R),$$

with pointwise convergence if e.g. g is differentiable.

We now apply this expansion in the setting of Theorem 2.0.4, assuming $f \in C^1(X_H)$. By (2.3.1), we have for any $p \neq q$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n\leq N} f(xu^{pn})\overline{f(xu^{qn})} = \int_{X_H \times X_H} f \otimes \overline{f}\,\mathrm{d}\lambda \tag{2.3.19}$$

for an appropriate joining λ. We prove by induction on d that there is a constant $C_f > 0$ depending only on f such that

$$\left| \int_{X_H \times X_H} f \otimes \overline{f} \, \mathrm{d}\lambda \right| \leq \frac{1}{pq} C_f.$$

By Proposition 2.3.9, we have for all but finitely many $p \neq q$ that (using the notation therein)

$$\begin{aligned} \int_{X_H \times X_H} f \otimes f \, \mathrm{d}\lambda &= \int_Z \int_{X_L \times X_L} \int_{X_R \times X_R} f_y \otimes f_z \, \mathrm{d}\rho^C_{(y,z)} \, \mathrm{d}(\nu_L \times \nu_L)(y,z) \, \mathrm{d}P(C) \\ &= \sum_{\chi,\psi \in \widehat{\Gamma_d \backslash R_d}} \int_Z \int_{X_L \times X_L} \int_{X_R \times X_R} f_y^\chi \otimes \overline{f_z^\psi} \, \mathrm{d}\rho^C_{(y,z)} \, \mathrm{d}(\nu_L \times \nu_L)(y,z) \, \mathrm{d}P(C), \end{aligned} \tag{2.3.20}$$

where the exchange of sums and integrals is justified by the pointwise convergence of the vertical Fourier series and the bounded convergence theorem, since each f_y is differentiable on X_R.

Now, the triviality of χ_0 means that for each $y \in X_L$, the vertical character $f_y^{\chi_0}$ on X_R is R_d-invariant; thus it passes to a continuous function on the nilmanifold $\Gamma \backslash R / R_d$. The nilpotent group R/R_d has strictly lower step than R, and thus we may apply the inductive hypothesis to assert that

$$\left| \int_Z \int_{X_L \times X_L} \int_{X_R \times X_R} f_y^{\chi_0} \otimes \overline{f_z^{\chi_0}} \, \mathrm{d}\rho^C_{(y,z)} \, \mathrm{d}(\nu_L \times \nu_L)(y,z) \, \mathrm{d}P(C) \right| \leq \frac{1}{pq} C_f, \tag{2.3.21}$$

where C_f depends only on f.

It remains to deal with

$$\sum_{\chi,\psi}{}' \int_Z \int_{X_L \times X_L} \int_{X_R \times X_R} f_y^\chi \otimes \overline{f}_z^\psi \, \mathrm{d}\rho_{(y,z)} \, \mathrm{d}(\nu_L \times \nu_L)(y,z) \, \mathrm{d}P(C),$$

where the primed summation runs over pairs of characters χ, ψ, at least one of which

is nontrivial.

Following [Z], we now need the following fundamental point.

Lemma 2.3.10. *Let R_d be the last term in the lower central series for the nilpotent group R, and let the measures $\rho_{(y,z)}$ on $X_R \times X_R$ be as in Proposition 2.3.9. For $\nu_L \times \nu_L$-a.e. $(y,z) \in X_L \times X_L$ and P-a.e. $C \in Z$, the support of $\rho^C_{(y,z)}$ is invariant under the action of $R_d \times R_d$.*

Proof. By Proposition 2.3.8, the action of $\widetilde{y}u_R\widetilde{y}^{-1}$ on (X_R, ν_R) is ergodic for ν_L-a.e. $y \in X_L$, and also uniquely and totally ergodic. Therefore, the action of $\widetilde{y}u_R^p\widetilde{y}^{-1}$ is ergodic, as is that of $\widetilde{z}u_R^q\widetilde{z}^{-1}$ for ν_L-a.e. $z \in X_L$. It follows from [Z2] that the orbit $\{(\Gamma_R\widetilde{y}u_R^{pn}\widetilde{y}^{-1}, \Gamma_R\widetilde{z}u_R^{qn}\widetilde{z}^{-1}) : n \in \mathbb{Z}\}$ is equidistributed in $Y_{(y,z)} = \{\Gamma_R\widetilde{y}r^p\widetilde{y}^{-1}, \Gamma_R r'\widetilde{z}r^q\widetilde{z}^{-1} : r \in R, r' \in [R,R]\}$ for $\nu_L \times \nu_L$-a.e. (y,z). Therefore, the support of $\rho^C_{(y,z)}$ equals $(r_C, s_C)Y_{(y,z)}$ for $\nu_L \times \nu_L$-a.e. $(y,z) \in X_L \times X_L$ and P-a.e. $C \in Z$. However, R_d is connected and therefore, being abelian, is divisible. Moreover, as a term in the lower central series for R, it is invariant under every automorphism of R, and in particular invariant under conjugation by $\widetilde{y}$ and $\widetilde{z}$, from which the claim now follows. □

Hence, for any $r_d, s_d \in R_d \times R_d$ and any $\chi, \psi \in \widehat{\Gamma_d\backslash R_d}$, at least one of which is nontrivial, we have

$$\begin{aligned}\int_{X_R\times X_R} f_y^\chi \otimes \overline{f}_z^\psi \,\mathrm{d}\rho^C_{(y,z)} &= \int_{X_R\times X_R} f_y^\chi(rr_d)\overline{f}_z^\psi(ss_d)\,\mathrm{d}\rho^C_{(y,z)}(r,s)\\ &= \chi(\Gamma_d r_d)\psi(\Gamma_d s_d)\int_{X_R\times X_R} f_y^\chi \otimes \overline{f}_z^\psi \,\mathrm{d}\rho^C_{(y,z)},\end{aligned}$$

and it follows that

$$\int_{X_R\times X_R} f_y^\chi \otimes \overline{f}_z^\psi \,\mathrm{d}\rho^C_{(y,z)} = 0.$$

Combining this with (2.3.20), we see that

$$\int_{X_H \times X_H} f \otimes \overline{f}\, \mathrm{d}\lambda = \int_Z \int_{X_L \times X_L} \int_{X_R \times X_R} f_y^{\chi_0} \otimes \overline{f_z^{\chi_0}}\, \mathrm{d}\rho_{(y,z)}^C\, \mathrm{d}(\nu_L \times \nu_L)(y,z)\, \mathrm{d}P(C),$$

and thus, by (2.3.21),

$$\left| \int_{X_H \times X_H} f \otimes \overline{f}\, \mathrm{d}\lambda \right| \leq \frac{1}{pq} C_f.$$

Theorem 2.0.4 now yields

$$\sum_n \mu(n) f(xu^n) = o(N),$$

proving Theorem 2.0.3. □

Bibliography

[Ab] L.M. Abramov, *Entropy of induced automorphisms*, Dokl. Akad. Nauk. SSSR **128** (1959), 647 – 650.

[AGH] L. Auslander, L. Green and F. Hahn, *Flows on homogeneous spaces*, Annals of Mathematics Studies **53**, Princeton University Pres (1963).

[ALR] E. El Abdalaoui, M. Lemanczyk, and T. de la Rue, *On spectral disjointness of powers for rank-one transformations and Mbius orthogonality*, Journal of Functional Analysis **266** No. 1 (2014), 284 – 317.

[AM] G. Avramidi and D. Witte Morris, *Horospherical limit points of finite-volume locally symmetric spaces*, New York J. Math **20** (2014), 353 – 366.

[B1] J. Bourgain, *On the correlation of the Moebius function with random rank-one systems*, Journal d'Analyse Mathmatique **120** No. 1 (2013), 105 – 130.

[B2] J. Bourgain *An approach to pointwise ergodic theorems* L.N.M., 1317, (1988), 204 – 223 .

[BHC] A. Borel and Harish-Chandra, *Arithmetic subgroups of algebraic groups*, Ann. Math **75** no. 3 (1962), 485 – 535.

[Bre] J. Brezin, *Harmonic analysis on compact solvmanifolds*, Lecture Notes in Math., Vol. 602, Springer-Verlag (1977).

[BSZ] J. Bourgain, P. Sarnak, and T. Ziegler, *Disjointness of Mobius from horocycle flows*, preprint (2011).

[C] J. Cassaigne et. al., *On finite pseudorandom binary sequences III: The Liouville function, I*, Acta Arithmetica LXXXVII.4 (1999), 367–390.

[CS] F. Cellarosi and Y. Sinai, *Ergodic properties of square-free numbers*, J. Eur. Math. Soc., **15** 4 (2013), 1343–1374.

[CT] V. Climenhaga and D. Thompson, *Intrinsic ergodicity beyond specification: β-shifts, S-gap shifts, and their factors*, Israel J. Math **192** (2012), 785 – 817.

[Dav] H. Davenport, *Multiplicative number theory*, Graduate Texts in Mathematics, Vol. 74, Springer-Verlag (1980).

[Dow] T. Downarowicz, *Entropy in Dynamical Systems*, Cambridge University Press (2011).

[Ei] M. Einsiedler, E. Lindenstrauss, P. Michel, and A. Venkatesh, *The distribution of closed geodesics on the modular surface, and Duke's theorem*, L'Enseignement Mathematique **2** 58 (2012), 249–313.

[ELR] E. El Abdalaoui, M. Lemanczyk, and T. De La Rue, *A dynamical point of view on the set of $\mathscr{B}$-free integers*, preprint, http://arxiv.org/abs/1311.3752 (2013).

[Fra] N. Frantzikinakis, *Equidistribution of sparse sequences on nilmanifolds*, Journal d'Analyse Mathmatique **109** No. 1 (2009), 353–395.

[FFR] L. Flaminio, G. Forni, and F. Rodriguez Hertz, *Invariant distributions for homogeneous flows*, preprint, arXiv:1303.7074 (2013).

[Fu] H. Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in diophantine approximation*, Mathematical Systems Theory **1** No. 1 (1967), 1 – 49.

[GT] B. Green and T. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. Math. 175 **2** (2012), 541–566.

[GJT] Y. Guivarc'h, L. Ji and J.C. Taylor, *Compactifications of symmetric spaces*, Birkhauser (1998).

[Hall] R.R. Hall, *The distribution of square-free numbers*, J. Reine Angew. Math **394** (1989), 107 – 117.

[Hay] N. Haydn, *Phase transitions in one-dimensional subshifts*, preprint, http://arxiv.org/abs/1111.6227 (2011).

[IK] H. Iwaniec and E. Kowalski, *Analytic number theory*, AMS Colloquium Publications **53** (2004).

[Kec] A. Kechris, *Classical Descriptive Set Theory*, Springer-Verlag (1995).

[KLW] J. Kulaga-Przymus, M. Lemanczyk, B. Weiss, *On invariant measures for $\mathcal{B}$-free systems*, preprint, arxiv.org/abs/1406.3745 (2014).

[MarNew] B. Marcus, S. Newhouse, *Measures of maximal entropy for a class of skew products*, Springer Lecture Notes in Math, **729** (1978), 105 – 125.

[LS] J. Liu and P. Sarnak, *The Möbius function and distal flows*, arXiv:1303.4957 (2014).

[Mac] G. W. Mackey, *Induced representation of locally compact groups I*, Ann. of Math., **55** (1952), 101–139.

[Mar] G. Margulis, *Discrete subgroups of semisimple Lie groups*, Springer-Verlag (1991).

[MR] K. Matomäki and M. Radziwill, *Multiplicative functions in short intervals*, arxiv:1501.04585 (2015).

[P] J. Pintz, *On the distribution of square-free numbers*, J. London Math Soc **28** (1983), 401 – 405.

[PR] V. Platonov and A. Rapinchuk, *Algebraic groups and number theory*, Academic Press (1994).

[Rag] M.S. Raghunathan, *Discrete subgroups of Lie groups*, Springer-Verlag (1972).

[R1] M. Ratner, *On measure rigidity of unipotent subgroups of semisimple groups*, Acta. Math **165** (1990), 229 – 309.

[R2] _____, *Raghunathan's topological conjecture and distributions of unipotent flows*, Duke Math. J. **63** No. 1 (1991), 235 – 280.

[R3] _____, *Horocycle flows, joinings and rigidity of products*, Ann. Math. **118** No. 2 (1983), 277 – 313.

[R4] _____, *Raghunathan's conjectures for cartesian products of real and* $\mathfrak{p}$*-adic Lie groups*, Duke Math J. **77** No. 2 (1995), 275 – 382.

[S] P. Sarnak, *Three lectures on the Möbius function, randomness and dynamics*, IAS Lecture Notes, http://publications.ias.edu/sites/default/files/MobiusFunctionsLectures(2).pdf (2009).

[Sta] A. Starkov, *Dynamical systems on homogeneous spaces*, AMS (2000).

[Sta2] A. Starkov, *An ergodic decomposition for homogeneous flows*, Math USSR Izvestiya **31** No. 3 (1988), 503 – 525.

[SU] P. Sarnak and A. Ubis, *The horocycle flow at prime times*, arXiv:1110.0777v4 (2011).

[T] T. Tao, *The Chowla conjecture and the Sarnak conjecture*, blog post, https://terrytao.wordpress.com/2012/10/14/the-chowla-conjecture-and-the-sarnak-conjecture/ (2012).

[Tsa] K.M. Tsang, *The distribution of r-tuples of squarefree numbers*, Mathematika **32** (1985), 265–275.

[Th] M. Thaler, *Transformations on* $[0,1]$ *with infinite invariant measures*, Israel J. Math, **46** (1983), 67–96.

[W] G. Warner, *Harmonic analysis on semi-simple Lie groups I*, Springer (1972).

[Wal] P. Walters, *An Introduction to Ergodic Theory*, Springer (1982).

[We] A. Weil, *Algebras with involution and the classical groups*, J. Indian Math. Soc. **24** (1960) 589 – 623.

[Wei] B. Weiss, *Subshifts of finite type and sofic systems*, Monats. Math. **77** (1973), 462 – 474.

[Wi] D. Witte, *Zero-entropy affine maps on homogeneous spaces*, American J. Math. **109** No. 5 (1987), 927 – 961.

[WM1] D. Witte Morris, *Ratner's theorems on unipotent flows*, The University of Chicago Press (1995).

[WM2] D. Witte Morris, *Introduction to Arithmetic Groups*, arxiv.org/src/math/0106063, (2014).

[Z] T. Zeigler, *A soft proof of orthogonality of Möbius to nilflows*, available online at http://tx.technion.ac.il/ tamarzr/soft-green-tao.pdf.

[Z2] T. Zeigler, *A non-conventional ergodic theorem for a nil-system*, Ergodic Theory and Dynamical Systems **25** No. 4 (2005), 1357–1370.

[Zo] P. Zorin-Kranich, *Ergodic theorems for polynomials in nilpotent groups*, PhD Thesis, Korteweg-de Vries Institute for Mathematics (2013).

[Zwei] R. Zweimüller, *Invariant measures for general(ized) induced transformations*, Proc. Amer. Math. Soc. **133** (2005), 2283–2295.