MINIMAL SETS FOR UNIPOTENT FLOWS ARE COMPACT

HARVEEN KAUR (ADVISED BY DR. AMIR MOHAMMADI)

Abstract. The aim of this paper is to prove that the minimal sets for unipotent flows are compact with the help of a theorem by G.A. Margulis. We begin by introducing the reader to an important example from topological dynamics which leads to the property of equi-distribution. Heading towards abstract topological dynamics, we look at the space of lattices in which Mahler's compactness criterion will help forming a compact set that a minimal set for a unipotent flow will return to. Non-divergence of horocycle flows will play an important role in the proving the aim.

Contents

1. Introduction 3
2. Linear maps 4
 2.1. Eigenvalues 5
3. Rotations of a circle 6
 3.1. Uniform Distribution for Intervals 8
 3.2. Uniform Distribution for functions 10
4. An application of Density and Uniform Distribution 14
 4.1. Distribution of First Digits of Powers 14
5. Abstract topological dynamics 16
6. Lattices 20
 6.1. Important facts about lattices 21
 6.2. Coset space of Lattices X_2 24
 6.3. Convergence of lattices 25
 6.4. Reduced basis 26

Date: June 6, 2018.

6.5. Mahler's compactness criterion 28
7. Non-divergence of unipotent flows in $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$ 29
Acknowledgements 34
References 34

1. Introduction

Topological dynamics involves the study of asymptotic behavior which requires direct methods that do not rely on prior explicit calculation of solutions. George Birkhoff was the founder of the field and the object of study in the field is a topological dynamical system. A dynamical system is a system in which a function describes the time dependence of a point in a geometrical space. Examples include mathematical models that describe number of rabbits that can be bred from a pair of rabbits in a year, flow of water in a pipe, swinging of a pendulum etc. Topological dynamical system is a topological space with continuous transformations of the space.

Given a flow on a compact metric space, minimal sets always exist (§5). Minimal sets are, informally speaking, "irreducible" objects in topological space. If the ambient space is compact, minimal sets may exist, but they may or may not be compact.

The conclusion of this paper relies on a famous theorem by G. A. Margulis stated as follows:

Theorem 1.1. *Let N be a group and let there be a continuous action of N on a locally compact space Z. Suppose that there exists an open subset $V \subset Z$ with the following properties:*

a) closure $\overline{V}$ is compact;

b) for each $z \in V$ and each $g \in N$, the 'semi orbit' $\{g^n z | n \geq 0\}$ does not tend to infinity ;

c) $NV = Z$.

Then Z is compact.

In this paper we will introduce and study basic properties of the space of uni-modular lattices i.e. discrete subgroups of $\mathbb{R}^2$ so that the quotient has volume 1. This space is identified with the homogeneous space $X_2 := \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$. Then we will consider a natural flow on X_2, more explicitly, the action is given by left multiplication with the subgroup

$$U := \left\{ u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} : t \in \mathbb{R} \right\}$$

As is often the case we will refer to this action as the *horocycle flow* on X_2. The space X_2 is not compact; nice compact subsets of X_2 are described by Mahler's compactness criterion (§6.5). We will prove a strong non-divergence result for the horocycle flow and use that in combination with Theorem 1.1 to prove the following:

Theorem 1.2. *Any minimal set for the action of U on X_2 is compact.*

We organize the paper in the following way to maximize smooth understanding of the reader while achieving our goal. We start by introducing the reader to linear dynamics (§2), especially an important example of rotations of a circle (§3). We explore the properties of orbits of points of a circle which strengthen our understanding of rotations of circle. We supplement our learning with various propositions around the dynamics of rotations of circle.

To see an application of rotations of circle in action, we give an example from number theory (§4). Henceforth, we talk about important notions in abstract dynamics which lead us into the statement of Theorem 1.1 (§5). After introducing and proving Theorem 1.1, we make a detour into detailed discussion about space of lattices (§6). Giving brief introduction about lattices and their properties, we talk about the space of lattices, $X_2 := \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$, whose discussion leads us into an example of a compact set in the space. The final section of the paper takes together the elements from the previous sections and formulates the conclusion of the paper which relies on non-divergence of unipotent flows (§7).

We first try to understand linear maps as simple dynamical systems.

2. Linear maps

The aim of this section is to provide readers preliminary information about dynamics in linear maps given by asymptotic behavior under iteration. The solutions of such systems are governed by the type and magnitude of eigenvalues. Eigenvalues play a very important role in practically any dynamical system. In trying to solve for $Ax = \lambda x$, where we let A to be a 2×2 invertible matrix and x an eigenvector, we mean that the subspace spanned by x is preserved by A. Therefore,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix}$$

gives us the equation $\lambda^2 - (a+d)\lambda + ad - bc$ which must equal to 0 for solutions to exist.

2.1. Eigenvalues.

In case of two distinct real eigenvalues λ and μ, $Ax = \lambda x$ and $Ay = \mu y$ are solved to get $\lambda \neq \mu$, which can be used as basis of matrix B obtained by diagonalization of matrix consisting of eigenvectors x and y. A linear map of $\mathbb{R}^2$ is called *hyperbolic* if absolute value of one of the eigenvalues is greater than 1 and the magnitude of the other lies in $(-1, 1)$.

In case of one real eigenvalue λ, the matrix B looks like $\binom{\lambda\, s}{0\, \lambda}$ for some $s \neq 0$. A linear map is called *parabolic* if it is conjugate to $\binom{\lambda\, 1}{0\, \lambda}$.

In case of complex conjugate eigenvalues, the matrix B looks like

$$\rho \begin{pmatrix} cos\ \theta & sin\ \theta \\ -sin\ \theta & cos\ \theta \end{pmatrix}$$

and a linear map conjugate to such matrix is called *elliptic.*

Now let's observe the asymptotic behavior of orbits of such maps:

Proposition 2.1. [1] *A linear map of* $\mathbb{R}^2$ *is eventually contracting if and only if all eigenvalues are of absolute value less than one.*

Proof. Having eigenvalues of absolute value less than 1 is a sufficient condition for the map to be eventually contracting. In the case of only one real eigenvalue,

$$B = \begin{pmatrix} \lambda & s \\ 0 & \lambda \end{pmatrix} \text{ with } 0 < 2s < 1 - |\lambda|,$$

$$\left| B \begin{pmatrix} x \\ y \end{pmatrix} \right| = \left| \begin{pmatrix} \lambda x + sy \\ \lambda y \end{pmatrix} \right| \leq (|\lambda| + s) \left| \begin{pmatrix} x \\ y \end{pmatrix} \right|$$

Similarly in the case of complex eigenvalues rotation by θ does not change the norm of any vector and the subsequent application of $\rho{*}\mathrm{Id}$ reduces their norm by a factor $\rho < 1$ if the eigenvalues have absolute value $\rho < 1$. □

In the case of distinct real eigenvalues, points approach the origin. The case of one real eigenvalue leads to *degenerate node* and complex eigenvalues lead to *spiraling behavior*. However, we do not explore further details since the aim of the paper is to discuss properties of notions in abstract dynamics. We stop here to discuss an important example from linear dynamics which will lay a foundation to our understanding of abstract dynamics later.

3. Rotations of a circle

So far, the asymptotic behavior in linear maps was either fixed or was attracted to various fixed points as time approached positive or negative infinity. The aim of this section is to introduce a fundamentally different type of behavior exhibited by rotations of circle.

We will introduce the concept of density of orbits in an irrational rotation and present results on the frequency of visits of points to the circle. We will discuss the role of *Birkhoff averaging operator* which will link the main result of this section to notions in physics. We will conclude the section by proving the *ergodicity* of irrational rotation using Stone-Weierstraβ theorem.

Note that most theorems and propositions in this section have been taken from Hasselblatt and Katok's *"A First Course in Dynamics..."*.

Rotations of circle form a fundamental example in theory of dynamical system whose behavior is not asymptotic, but *recurrent.* A circle of radius 1 in the complex plane is given by

$$\mathbb{R}/\mathbb{Z} = \{e^{2\pi i\phi} | \phi \in \mathbb{R}\} = S^1$$

Let R_α denote the rotation by the angle $2\pi\alpha$. Then

$$\begin{aligned} R_\alpha(z) &= e^{2\pi i\alpha} z \\ R_\alpha^n(z) &= R_{\alpha n}(z) = e^{2\pi i\alpha n} z \end{aligned}$$

When α is rational, the orbit of any point on a unit circle is a finite set and all orbits are periodic since for some $x, n \in \mathbb{Q}, R_{x\alpha}(z) = R_{nx\alpha}(z)$ as the iterates return back to original point they started from. However, the irrational case of α is different and gives more insight into other phenomena related to rotations of a circle. The following proposition proves that the orbits of irrational rotation which, intuitively[1] must be infinite, are in fact, dense.

Proposition 3.1 ([1])**.** *If $\alpha \notin \mathbb{Q}$, then every positive semi orbit of R_α is dense.*

Proof. In order to show that we divide the circle into finitely many closed arcs of any length less than $\varepsilon > 0$ and place infinitely many $R^n(z)$ into them for any $z \in S^1$. By Pigeonhole principle, at least two of the iterates $R^n(z)$ and $R^m(z)$ must be in one of the arcs such that

$$d(R^n(z), R^m(z)) < \varepsilon$$

[1]'Mathematics is not intuition' - Unknown

where,

$$d(x,y) = min\{|b-a| \; |b \; \epsilon \; x, a \; \epsilon \; y\}$$

Then,

$$d(R^{n-m}(z), z) < \varepsilon$$

We claim that,

Claim 3.2. *$d(R^{n-m}(z), z)$ is independent of z because if $w \; \epsilon \; S^1$, then $w = R_{w-z}(z)$.*

Proof. Consider the following,

$$\begin{aligned} d(R_\alpha^{n-m}(w), w) &= d(R_\alpha^{n-m}(R_{w-z}(z)), R_{w-z}(z) \\ &= d(R_{\alpha(n-m)+w-z}(z), R_{w-z}(z)) \\ &= d(R_{w-z}(R_\alpha^{n-m}(z)), R_{w-z}(z)) \\ &= d(R_\alpha^{n-m}(z), z) \end{aligned}$$

Therefore, n and m can be chosen independently of z. □

Now, let $\theta \in [\frac{-1}{2}, \frac{1}{2}]$ and $\theta = (n-m)\alpha \; mod \; 1$. Then

$$|\theta| < \varepsilon, R_\alpha^{n-m} = R_\theta$$

Let $N = \lfloor 1/|\theta| \rfloor + 1$, then the subset $\{R_{i\theta}(z) | i = 0, 1, ...N\}$ divides the circle into intervals of length $< |\theta|$ so $\exists \; k$ such that $k \leq N(n-m)$ such that

$$d(R_\alpha^k(z), x) < \varepsilon$$

for some $x \in S^1$.

□

Proposition 3.1 motivates the following definitions:

Definition 3.3. *A homeomorphism $f : X \to X$ is called topologically transitive if there exists a point $x \in X$ such that its orbit is dense in X.*

Definition 3.4. *A homeomorphism $f : X \to X$ is called minimal if the orbit of every point $x \in X$ is dense in X.*

Minimality implies *topological transitivity* and not vice-versa. Therefore, rotations of circle, $R_\alpha : S^1 \to S^1$ is minimal and hence, topologically transitive.

3.1. Uniform Distribution for Intervals.

Now we are ready to dive into detailed discussion of how iterates of a point on a circle visit parts of a circle by measuring their frequency of visits on parts of circle.

Definition 3.5. *Let Δ denote an arc of a circle, then*

$$F_\Delta(x,n) = \{k \in \mathbb{Z} \mid 0 \leq k \leq n, R_\alpha^k(x) \in \Delta\}$$

In words, $F_\Delta(x,n)$ is the number of times an iterate of a point x visits the arc Δ, and n is the maximum number of iterates allowed.

By Proposition 3.1, since positive semi orbit of any point x on circle is dense, this would imply that as $n \to \infty$, $F_\Delta(x,n) \to \infty$.

The relative frequency of visits is given by $\frac{F_\Delta(x,n)}{n}$.

Proposition 3.6. *If α is irrational and R_α is the rotation, let Δ, Δ' be arcs such that $l(\Delta') > l(\Delta)$, where $l(\Delta)$ denotes length of the arc Δ. Then there exists $N_0 \in N$ such that if $x \in S^1, N \geq N_0$ and $n \in \mathbb{N}$ then*

$$F_{\Delta'}(x, n+N) \geq F_\Delta(x,n)$$

Proof. Since length of arc Δ' is greater than that of arc Δ, we want to show that the number of iterates of x visiting Δ' is greater than number of iterates of x visiting Δ.

By Proposition 3.1, since positive semi orbit of Δ' is dense, we can find an N_0 such that $R_\alpha^{N_0}(\Delta) \subset \Delta'$ (iterates of x in Δ belong to Δ'). Then since $R_\alpha^n(x) \in \Delta$, this implies that $R_\alpha^{n+N_0}(x) \in \Delta'$. Since $N \geq N_0$,

$$F_{\Delta'}(x, n+N) \geq F_{\Delta'}(x, n+N_0) \geq F_\Delta(x,n).$$

□

Arcs also satisfy the additivity property:

$$F_{\Delta_1}(x,n) + F_{\Delta_2}(x,n) = F_{\Delta_1 \cup \Delta_2}(x,n)$$

Since we do not know if limits of relative frequencies exist, we consider the upper limits:

$$\overline{f_x}(A) := \limsup_{n\to\infty} \frac{F_A(x,n)}{n}$$

where A is the union of disjoint arcs of S^1.

Note that if $\bigcup_{i=1}^n A_i = S^1$, then

$$\sum_{i=1}^n \overline{f_x}(A_i) \geq 1$$

Proposition 3.6 implies that,

Corollary 3.7. *If $l(\Delta') \geq l(\Delta)$, then $\overline{f_x}(\Delta') \geq \overline{f_x}(\Delta)$.*

Proof. For any set A (of arcs), we have

$$F_A(x,n) + F_{A^c}(x,n) = n$$

(meaning that total number of iterates n of a point x on a circle lie in A or A^c).

Then dividing the equation by n and taking the limit on both sides, we get ,

$$\limsup_{n\to\infty} \frac{F_A(x,n)}{n} = 1 - \liminf_{n\to\infty} \frac{F_{A^c}(x,n)}{n}$$

This means that as we approach the greatest possible limit on the left hand side, the right hand side must approach the smallest limit possible. $\square$

Now we can formulate the main result of this section:

Proposition 3.8. *For any arc $\Delta \subset S^1$ and any point $x \in S^1$,*

$$f(\Delta) := lim_{n\to\infty} \frac{F_\Delta(x,n)}{n} = l(\Delta)$$

This means that the uniform limit of relative frequency of iterates of a point x in an arc Δ gives us the length of the arc! Such a property of the sequence $a_n := R_\alpha^n(x)$ expressed by this proposition is called *uniform distribution* or *equi-distribution.* This implies that the asymptotic frequency of visits is the same for arcs of equal length, regardless of where they are on the circle.

Proof. To prove this, we shall use the following lemma:

Lemma 3.9. *If $l(\Delta) = \frac{1}{k}$, then $\overline{f_x}(\Delta) \leq \frac{1}{k-1}$.*

Proof. Consider $k-1$ disjoint arcs, $\Delta_1, \Delta_2, ...\Delta_{k-1}$ each of length $\frac{1}{k-1}$, then $l(\Delta_i) \geq l(\Delta)$ for $i = 1, 2, ..., k-1$. Then by Proposition 3.6,

$$F_\Delta(x,n) \leq F_{\Delta_i}(x, n+N)$$

$$F_\Delta(x,n) \leq \frac{1}{k-1}(F_{\Delta_1}(x,n+N) + F_{\Delta_2}(x,n+N) + ... + F_{\Delta_n}(x,n+N))$$

Since $F_\Delta(x,n)$ obeys the additivity property,

$$F_\Delta(x,n) \leq \frac{1}{k-1}(F_{\Delta_1 \cup \Delta_2 ... \cup \Delta_n}(x, n+N))$$
$$F_\Delta(x,n) \leq \frac{1}{k-1}(F_{S^1}(x, n+N))$$

where $S^1 = \Delta_1 \cup \Delta_2 ... \cup \Delta_n$ (sum of all arcs)

$$F_\Delta(x,n) \leq \frac{1}{k-1}(n+N)$$

Taking the limits as $n \to \infty$,

$$\lim_{n\to\infty} \frac{F_\Delta(x,n)}{n} \leq \frac{1}{k-1}$$

or

$$\overline{f_x}(\Delta) \leq \frac{1}{k-1}$$

□

Now cut the the arc Δ into l disjoint sub-arcs of length $\frac{1}{k-1}$, then for each sub-arc and for $\varepsilon > 0$,

$$\overline{f}(\Delta) < \frac{1}{k-1} < l(\Delta) + \varepsilon$$
$$\overline{f}(\Delta) < l(\Delta)$$

Similarly,

$$\overline{f}(\Delta^c) < l(\Delta^c)$$
$$1 - \overline{f}(\Delta^c) \geq 1 - l(\Delta^c)$$
$$\underline{f}(\Delta) \geq l(\Delta)$$
$$\overline{f}(\Delta) < l(\Delta) < \underline{f}(\Delta)$$
$$f(\Delta) = l(\Delta)$$

as needed.

□

3.2. **Uniform Distribution for functions.**

We can also define

$$F_A(x,n) := \sum_{k=0}^{n-1} \chi_A(R_\alpha^k(x))$$

where χ_A is the characteristic function of A (finite union of arcs Δ),

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Therefore, result of Proposition 3.8 can be reformulated as:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(R_\alpha^k(x)) = \int_{S^1} \chi_\Delta(\phi)d\phi$$

Considering similar expressions for functions ϕ, we define the *Birkhoff averaging operator* as follows:

Definition 3.10. *The Birkhoff averaging operator is the operator that associates to a function ϕ the Birkhoff function $\mathfrak{B}_n(\phi) := \sum_{k=0}^{n-1} \frac{\phi \circ R_\alpha^k}{n}$ given by*

$$\mathfrak{B}_n(\phi)(x) = \frac{1}{n} \sum_{k=0}^{n-1} \phi(R_\alpha^k(x))$$

Proposition 3.11. *For any function ϕ that is a uniform limit of step functions we have*

$$\lim_{n\to\infty} \mathfrak{B}_n(\phi) = \int_{S^1} \phi(\theta)d\theta$$

Proof. Let ϕ_ε be a step function such that $\phi_\varepsilon - \epsilon < \phi < \phi_\varepsilon + \epsilon$.

We want to show that $|\lim_{n\to\infty} \mathfrak{B}_n(\phi) - \int_{S^1} \phi(\theta)d\theta| < \epsilon$.

Observe that from the way we define ϕ_ε and traversing within the inequalities,

$$\int_{S^1} \phi d\theta - 2\epsilon = \int_{S^1} (\phi - \epsilon)d\theta - \epsilon \leq \int_{S^1} \phi_\varepsilon - \epsilon \; d\theta$$

$$\begin{aligned} \int_{S^1} \phi_\varepsilon - \epsilon \; d\theta = \lim_{n\to\infty} \mathfrak{B}_n(\phi_\varepsilon) - \epsilon \leq \lim_{n\to\infty} \mathfrak{B}_n(\phi + \epsilon) - \epsilon \leq \lim_{n\to\infty} \mathfrak{B}_n(\phi) \\ \leq \lim_{n\to\infty} \mathfrak{B}_n(\phi_\varepsilon) + \epsilon \leq \int_{S^1} \phi_\varepsilon d\theta + \epsilon \leq \int_{S^1} \phi d\theta + 2\epsilon \end{aligned}$$

Letting $\epsilon \to 0$ gives us the result. $\square$

Now Proposition 3.8 can be proven in the following way:

Theorem 3.12. *If α is irrational and ϕ is Riemann integrable, then*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\phi(R_\alpha^k(x)) = \int_{S^1}\phi(\theta)d\theta$$

converges uniformly in x.

Proof. Pick a partition of circle S^1 into finite number of arcs I_i. Then recall,

$$L(P,\phi) = \sum_i l(I_i)m_i \text{ where } m_i \text{ is the } \inf_{[I_i]}\phi$$

$$U(P,\phi) = \sum_i l(I_i)M_i, \text{ where } M_i \text{ is the } \sup_{[I_i]}\phi$$

We know that

$$\sum_i I_i m_i \geq m\sum I_i, \text{ where } m \text{ is the minimum value of } \phi.$$

and

$$\sum_i I_i M_i \leq M\sum I_i, \text{ where } M \text{ is the maximum value of } \phi.$$

Therefore,

$$m\sum I_i \leq \sum_i I_i m_i \leq \sum_i I_i M_i \leq M\sum I_i \tag{1}$$

Let $m_i = \phi_1$ and $M_i = \phi_2$, then

$$\sum_i I_i m_i \text{ becomes } \int_{S^1}\phi_1 d\theta$$

and

$$\sum_i I_i M_i \text{ becomes } \int_{S^1}\phi_2 d\theta$$

Since $m \leq m_i \Rightarrow \phi - \epsilon \leq \phi_1$ and, $M \geq M_i \Rightarrow \phi + \epsilon \geq \phi_2$

Together,

$$\phi - \epsilon \leq \phi_1 \leq \phi_2 \leq \phi + \epsilon$$

Therefore (1) can be re-written as,

$$\int_{S^1}\phi d\theta - \epsilon \leq \int_{S_1}\phi_1 d\theta \leq \int_{S^1}\phi_2 d\theta \leq \int_{S^1}\phi d\theta + \epsilon \tag{2}$$

Taking the first half of the inequality (2), and using the result from Proposition 3.11,

$$\int_{S^1} \phi d\theta - \epsilon \leq \int_{S_1} \phi_1 d\theta = \lim_{n\to\infty} \mathfrak{B}_n(\phi_1 + \epsilon) - \epsilon \leq \lim_{n\to\infty} \mathfrak{B}_n(\phi - \epsilon) =$$
$$\lim_{n\to\infty} \mathfrak{B}_n(\phi_1) \leq \lim_{n\to\infty} \mathfrak{B}_n(\phi) - \epsilon \leq \lim_{n\to\infty} \mathfrak{B}_n(\phi_2 + \epsilon) - \epsilon \leq \lim_{n\to\infty} \mathfrak{B}_n(\phi_2)$$

The last half of inequality can be rewritten as,

$$\int_{S^1} \phi_2 d\theta \leq \int_{S^1} \phi d\theta + \epsilon$$

By letting $\epsilon \to 0$ gives us the desired result. □

Remark 3.13. *Uniform convergence in all x is a special feature of irrational rotations, because in general dynamics this is false. For general dynamical systems, the above limit converges for most of the points for a function, but not for all.*

Remark 3.14. *The condition of Riemann integrability is essential.*

Counter-example: Consider a point x_0 and define the set A as the union of arcs of length $\frac{1}{2^{k-2}}$ centered at $R_\alpha^k(x_0)$ for $k \geq 0$. Then $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \chi_A(R_\alpha^k(x)) = 1$. However, sum of union of arcs is less than $\frac{1}{2}$. This is because χ_A is not Riemann-integrable (it is not a finite union of arcs).

We can relate the result of Theorem 3.12 to notion of averages in physics as following:

Definition 3.15. *Given a function ϕ, we call $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi(R_\alpha^k(x))$, the time average and the integral $\int_{S^1} \phi(\theta) d\theta$ is called the space average of the function ϕ.*

Therefore, we showed that for any Riemann-integrable function the time average exists for the orbit of any point x and always coincides with the space average. This property of irrational rotations is referred to as *unique ergodicity*.

Definition 3.16. *If X is a compact metric space and $f : X \to X$ a continuous map, then f is said to be uniquely ergodic if*

$$\frac{1}{n}\sum_{k=0}^{n-1} \phi(f^k(x))$$

converges to a constant uniformly for every continuous function ϕ.

We can also apply Stone-Weierstraβ theorem to prove the unique ergodicity of an irrational rotation (Theorem 3.12) in the following way:

Proof. Stone-Weierstraβ theorem says that continuous functions are uniform limits of trigonometric polynomials. Therefore, if α is irrational and $\phi = c_m(x) := e^{2\pi imx}$ is continuous function, then by the definition of $R_\alpha(x)$, $R_\alpha(x) = x + \alpha \ mod \ 1$ implies that

$$c_m(R_\alpha(x)) = e^{2\pi im(x+\alpha)} = e^{2\pi im\alpha}e^{2\pi imx} = e^{2\pi im\alpha}c_m(x) \tag{3}$$

Therefore, $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} c_m(R_\alpha^k(x))$ must approach a constant uniformly in x. With the help of (3),

$$\left|\frac{1}{n}\sum_{k=0}^{n-1} c_m(R_\alpha^k(x))\right| = \left|\frac{1}{n}\sum_{k=0}^{n-1} e^{2\pi imk\alpha}\right| \tag{4}$$

Employing the geometric series sum in (4) for $x = e^{2\pi im\alpha}(\sum_{k=0}^{n} x^k = \frac{1-x^{n+1}}{1-x})$

$$\left|\frac{1}{n}\sum_{k=0}^{n-1} e^{2\pi imk\alpha}\right| = \frac{|1 - e^{2\pi imn\alpha}|}{n|1 - e^{2\pi im\alpha}|} \tag{5}$$

Using the triangle inequality in the numerator,

$$|1 - e^{2\pi imn\alpha}| \leq |1| + |e^{2\pi imn\alpha}| \leq 1 + 1 \leq 2.$$

the expression in (5) is less than $\frac{2}{n|1-e^{2\pi im\alpha}|}$

which implies that $\frac{2}{n|1-e^{2\pi im\alpha}|} \to 0$ as $n \to \infty$.

Since Birkhoff averaging operators are linear, if $p(x) = \sum_i a_i c_i(x)$, then $\lim_{n\to\infty} \mathfrak{B}_n(p)(x)$ exists and is constant. □

4. An application of Density and Uniform Distribution

This section illustrates an application of density and uniform distribution through a particular example from number theory.

4.1. Distribution of First Digits of Powers.

Proposition 4.1. [1] *Let $k \in \mathbb{N}$ other than a power of 10 and $p \in \mathbb{N}$. Then there exists an $n \in \mathbb{N}$ such that p gives the initial digits of the decimal expansion of k^n.*

Proof. Mathematically, we need to prove that $\exists\, l \in \mathbb{N}$ for which $k^n = 10^l p + q$, given p and where $0 \leq q < 10^l$.

Given the form of k^n,

$$10^l p \leq k^n < 10^l (p+1)$$

Taking $\log_{10}$ throughout the inequality,

$$\begin{aligned} \log_{10}(10^l p) &\leq n \log_{10} k < l + \log_{10}(p+1) \\ l + \log_{10}(p) &\leq n \log_{10} k < l + \log_{10}(p+1) \\ \log_{10}(p) &\leq n \log_{10} k - l < \log_{10}(p+1) \end{aligned} \tag{6}$$

Define $m = \lfloor \log_{10} p \rfloor + 1$ to be the number of digits of p, then subtract $m-1$ throughout inequality (6),

$$0 \leq \log_{10} p - (m-1) \leq n \log_{10} k - l - (m-1) < \log_{10}(p+1) - (m-1) \leq 1 \tag{7}$$

The middle part of the inequality (7) given by

$$\begin{aligned} n \log_{10} k - l - (m-1) &= n \log_{10} k - (l + \lfloor \log_{10} p \rfloor) \\ &= n \log_{10} k - \lfloor n \log_{10} k \rfloor = \{n \log_{10} k\} \end{aligned}$$

Going back to the inequality (7) and replacing the middle part with the fractional part above,

$$\begin{aligned} \log_{10} p - (m-1) &\leq \{n \log_{10} k\} < \log_{10}(p+1) - (m-1) \\ = \log_{10} \frac{p}{10^{m-1}} &\leq \{n \log_{10} k\} < \log_{10} \frac{p+1}{10^{m-1}} \end{aligned}$$

Now we claim that $\log_{10} k$ is irrational. Suppose not, then $\log_{10} k = \frac{p}{q}$ for some $\frac{p}{q} \in \mathbb{Q}$. This implies that

$$\begin{aligned} k^q &= 10^p \\ k^q &= 2^p 5^q \end{aligned}$$

This implies that $k = 10^m$ and $n = m$, which is a contradiction to our assumption. Hence, $\log_{10} k$ must be irrational.

Therefore, $\{\{n \log_{10} k\} | n \in \mathbb{N}\}$ is dense on $[0, 1)$ by the final inequality. Specifically, it is dense on the interval $[\log_{10} \frac{p}{10^{m-1}}, \log_{10} \frac{p+1}{10^{m-1}}]$.

Recalling from Proposition 3.8, one can define the asymptotic frequency for such distribution as follows:

$$\lim_{n \to \infty} \frac{F_p^k(n)}{n} = \log_{10}(p+1) - \log_{10}(p)$$

∎

Hence, we saw how results from dynamics of rotations of circle play a role in number theory.

5. Abstract topological dynamics

This section aims at explaining abstract dynamics which is developed in the context of *flows*. Understanding *group actions* and *flows* lead to the notion of a *minimal set* upon which one of the famous theorems of G. A. Margulis is stated. The theorem gives conditions under which a space is compact. To illustrate the importance of the theorem, we discuss an example of a space to which it does not apply to. Subconsciously, we understand how irrational rotations play an important role in abstract dynamics.

Irrational rotations serve as the starting point for a number of generalizations coming up next. For instance, one can look at the circle as a compact abelian group in which rotations can act like group multiplication:

$$L_{g_0} : G \to G \text{ given by the action } L_{g_0}g = g_0g,\ \forall g \in G$$

The *orbit* of any element is a cyclic subgroup and to define the notion of a dense orbit, we introduce *topological group* to be a group with a metric for which every L_{g_0} is a homeomorphism and taking inverses is continuous.

To define a flow, let's define what a group action is:

A *group action* of a group G on set X is a function $f : G \times X \to X$ satisfying the following two properties:

$\forall x \in X, \forall g, h \in G$ and e_G is the identity element in G:

$$f(e_G, x) = x$$

$$f(gh, x) = f(g, f(h, x))$$

If G is a topological group and X is a topological space, then we can talk about continuous group actions $f : G \times X \to X$. A *flow* on X is a group action of $\mathbb{R}$ on X. It is a continuous mapping, $g : X \times \mathbb{R} \to X$.

If (X, g) is a flow and $x \epsilon X$ then we define the *orbit* of x to be the set $S = \{tx \mid t \epsilon X\}$. In simple words, it is the collection of all elements of the space X to which an element $x \in X$ has been moved by the group action of elements of G.

Informally, a *minimal set* is the 'irreducible' object of our compact metric space X. Mathematically, it is a *non-empty, closed* and *invariant* set such that none of its proper subsets exhibit the same properties.

We claim that every compact metric space contains minimal sets. We prove it the following way:

Theorem 5.1. *Let X be a compact metric space and $f : X \to X$ be a homeomorphism. If $Y \subset X$ which is closed, invariant and non-empty such that $\Omega = \{f^n(y) : y \in Y\}$ given by a partial ordering $Y_1 < Y_2 \Leftrightarrow Y_2 \subseteq Y_1$, then $\exists\ Y_\alpha$ such that Y_α is a minimal set.*

Proof. Since X is a compact metric space and Y is a closed subset of X then by compactness of X,

$$\bigcap_{i=1}^{\infty} Y_i = Y_\alpha$$

where Y_i are the closed subsets of X.

We claim that,

Claim 5.2. $Y_\alpha \in \Omega$

Proof. Y_α is closed since it lies in the intersection of the closed subsets as said above.

Y_α is non-empty because if not, then $\exists\ n$ such that $Y_n = \phi$, but this is a contradiction since Y_i is non-empty for all i.

Want to show that Y_α is invariant or if $f : Y_\alpha \to Y_\alpha$ then if $y \in Y_\alpha \Rightarrow f(y) \in Y_n$,

$$f(y) \in Y_n, \forall n \Rightarrow f(y) \in Y_\alpha$$

□

Therefore, by Zorn's Lemma which states that a partially ordered set with the property that every chain in it has an upper bound, then the set contains at least one maximal element, implies that $\exists\ Z \in \Omega$, which is maximal. Hence, Y_α is a minimal set. □

The next theorem enumerates conditions for compact metric spaces exhibiting compact minimal sets:

Theorem 5.3 (Margulis [3]). *Let N be a group and let there be a continuous action of N on a locally compact space Z. Suppose that there exists an open subset $V \subset Z$ with the following properties:*

a) closure $\overline{V}$ is compact;

b) for each $z \in V$ and each $g \in N$, the 'semi orbit' $\{g^n z | n \geq 0\}$ does not tend to infinity ;

c) $NV = Z$.

Then Z is compact.

Proof. Let the group N be the direct product of $\mathbb{R}^k$ and $\mathbb{Z}^l$. $N = \mathbb{R}^k \times \mathbb{Z}^l$. Then N is a group consisting of elements with coordinates $(t_1, t_2, ..., t_n)$, where $n = k + l$.

$$t_i \in \mathbb{R} \text{ if } 1 \leq i \leq k$$

$$t_i \in \mathbb{Z} \text{ if } k + 1 \leq i \leq k + l$$

So, if g is an element of N, then

$$g = (t_1(g), t_2(g), ..., t_n(g))$$

and we define the norm of g to be,

$$\|g\| = |t_1(g)| + |t_2(g)| + ... + |t_n(g)|$$

Balls, $\{g \in N | \|g\| \leq r\}, r \geq 0$ are compact.

Given that $NV = Z$ implies that for every $z \in Z$, one can find a $\phi(z) \in N$ such that $\phi(z)z \in \overline{V}$ and if $\|g\| < \|\phi(z)\|$, then $gz \notin \overline{V}$. This means that $\phi(z)$ is the smallest element of the group which brings some z in the closure of V.

Now, assume by way of contradiction that Z is not compact. Then there exists a sequence $\{z_m\} \subset Z$ such that $\|\phi(z_m)\| \to \infty$ (the coordinates will push $z_m s$ apart in Z).

Passing to a subsequence $\{z_{m_j}\}$, for some $i, 1 \leq i \leq m$,

$$\lim_{m \to \infty} |t_i(\phi(z_m))| = \infty$$

and the sign of all such coordinates is the same.

Let h be an element of N such that $|t_i(h)| = 1$ and $t_j(h) = 0$ if $j \neq i, 1 \leq j \leq n$ (for instance h looks like $(-1, 0, ...0)$). Clearly the sign of $t_i(h)$ differs from sign of $|t_i(\phi(z_m))|$. So $t_i(h)$ plays the role of bringing back the elements back which $t_i(\phi(z_m))$ have pushed apart.

Now, if $0 \leq r \leq |t_i(\phi(z_m))|$, then

$$\|h^r \phi(z_m)\| = \|h^r\| + \|\phi(z_m)\| = -r + \|\phi(z_m)\|$$

Consequently,

$$\|gh^r \phi(z_m)\| \leq \|g\| + \|h^r \phi(z_m)\| \leq \|g\| + \|\phi(z_m)\| - r$$

This implies that,

$gh^r\phi(z_m)z_m \notin \overline{V}$ if $\|g\| < r \leq |t_i(\phi(z_m))|$ since $\|gh^r\phi(z_m)\| < \|\phi(z_m)\|$.

Since $\overline{V}$ is compact, the subsequence $\{\phi(z_m)z_m\}$ has a limit, say, $v \in \overline{V}$.

Since V is open and $\lim_{m\to\infty} |t_i(\phi(z_m))| = \infty$ and $gh^r\phi(z_m)z_m \notin \overline{V}, gh^r v \notin V$ if $\|g\| < r$. Why?

Suppose not, if $gh^r v \in V$ then there exists a neighborhood of $gh^r v$ in V such that $gh^r\phi(z_m)z_m \in V$ as v is the limit of the sequence $\phi(z_m)z_m$. However, $gh^r\phi(z_m)z_m \notin \overline{V}$ and hence, $gh^r\phi(z_m)z_m \notin V$, which is a contradiction.

Given the condition (b) in the theorem, if $h \in N$ and $z \in V$, the sequence $\{h^r z\}$ does not tend to infinity. Therefore, $\exists$ a sequence $\{r_j\}$ of positive integers such that,

$$\lim_{j\to\infty} r_j = \infty \text{ and } \lim_{j\to\infty} h^{r_j} v = z \text{ for some } z \in Z,$$

which implies that $g(h^{r_j} v) \to g(z)$, however, since $gh^r v \notin V \Rightarrow gz \notin V$ for every $g \in N$. But this is a contradiction to condition (c) of the theorem which states that $NV = Z$.

Hence, Z must be compact. □

Remark 5.4. *The above theorem shows that if a space is locally compact, then given the above three properties of such a space, the space is compact. However, if one of the properties breaks down, the conclusion of the theorem is no longer true.*

Let's see a counter-example:

We know that $\mathbb{R}$ (real line) is locally compact and not compact. This means that at least one of the conditions of the theorem must fail. We notice that if we introduce the group action $\{n + m\sqrt{2} \mid n, m \in \mathbb{Z}\}$, then we claim that since this set is dense (we prove this later), the second condition of the theorem is no longer true.

For instance, if we consider the action of 1 on 0, then one can see that 0 never returns and its semi-orbit becomes infinite under the group action of 1. However, the orbits for other points become dense to irrationality of $\sqrt{2}$.

We prove the density of the set mentioned before:

Claim 5.5. *The set $S = \{n + m\sqrt{2} \mid n, m \in \mathbb{Z}\}$ is dense.*

Proof. First, we will prove that the set $R_\alpha = \{m\sqrt{2} \mid m \in \mathbb{Z}\}$ is dense in $(0, 1]$.

Let $\varepsilon > 0$ and pick an integer m such that $m > \frac{1}{\varepsilon}$.

Divide the interval $[0, 1]$ into m sub intervals of length $\frac{1}{m}$, then two numbers from the set $\{\{\sqrt{2}\}, \{2\sqrt{2}\}, ..., \{(m+1)\sqrt{2}\}\}$ must lie in the same sub-interval by the Pigeonhole principle. In other words, $\exists\, i, j \in \mathbb{Z}, 1 \leq i < j \leq m + 1$, such that

$$0 < |\{i\sqrt{2}\} - \{j\sqrt{2}\}| < \frac{1}{m} \tag{8}$$

Now, for any $y \in [0, 1], \exists\ 0 \leq k \leq m - 1$ such that $y \in [\frac{k}{m}, \frac{k+1}{m}]$

This implies that by inequality (8), for some q,

$$q\{(i-j)\sqrt{2}\} \in [\frac{k}{m}, \frac{k+1}{m}]$$
$$|y - q\{(j-i)\sqrt{2}\}| < \frac{1}{m} < \varepsilon$$

Hence, R_α is dense.

Now, any number $z \in \mathbb{R}$ can be written as $z = a + r$, where $a \in \mathbb{Z},\ 0 \leq r < 1$

Since R_α is dense in $[0, 1]$,

$$|r - \{b\sqrt{2}\}| < \epsilon$$
$$| - b\sqrt{2} + \lfloor b\sqrt{2} \rfloor + z - a| < \epsilon$$
$$a - \lfloor b\sqrt{2} \rfloor + b\sqrt{2} \in R_\alpha$$

which is dense in $\mathbb{R}$. □

This counterexample is essentially another proof for Proposition 3.1.

To see how Theorem 5.3 plays a role in establishing compactness of minimal sets, we take a detour to study the space of lattices which will provide us with tools to approach our goal.

6. Lattices

This section acquaints the reader to lattices and their properties. Once we learn properties of lattices, we will see properties of the coset space $X_2 := \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$ of lattices. We will discuss the notion of converging sequence of lattices and conditions governing them, which will help us in proving the Mahler's compactness criterion. Proving Mahler's criterion gives us a necessary tool towards proving our main goal.

Definition 6.1. *In a locally compact group, a lattice is a discrete co-compact subgroup with the property that the quotient space has finite invariant measure. In special case of subgroups of $\mathbb{R}^n$ a lattice in $\mathbb{R}^n$ is a subgroup of the additive group $\mathbb{R}^n$ which is isomorphic to the additive group $\mathbb{Z}^n$, and which spans the real vector space $\mathbb{R}^n$.*

We will consider the geometric notion of lattice given by periodic subset of points.

Given n linearly independent vectors $b_1, ..., b_n \in \mathbb{R}^n$, the lattice generated by them is defined as,

$$\Lambda(b_1, ..., b_n) = \sum_{i=1}^{n} x_i b_i,\ x_i \in \mathbb{Z}$$

The vectors $b_1, ..., b_n$ form the basis of the lattice. For instance, the lattice generated by $(1, 0)^T$ and $(0, 1)^T$ is the lattice of all integer points $(\mathbb{Z}^2)$.

Note that different bases can generate the same lattice, for instance $(1, 1)^T$ and $(2, 1)^T$ generate $\mathbb{Z}^2$ as well.

6.1. **Important facts about lattices.**

A lattice Λ is a *discrete, additive* subgroup of $\mathbb{R}^n$ with a finite quotient space $\frac{R^n}{\Lambda}$.

Discrete means that $\forall x \in \Lambda, \exists\ r_x > 0$ such that $B(x, r_x) \cap \Lambda = x$, meaning the neighborhood of a ball of positive radius r contains exactly one element(vector) of the lattice.

Additive means that if $x, y \in \Lambda, x - y \in \Lambda$.

Therefore, by the additivity property of lattices, $\exists\ \epsilon > 0$, such that $\forall x \neq y \in \Lambda, \|x - y\| > \epsilon$. This gives us another definition of discreteness.

Definition 6.2. *A matrix U is Uni-modular if $det(U) = \pm 1$.*

Definition 6.3 ([5])**.** *For any lattice Λ and a basis $(b_1, b_2, ..., b_n)$, we define a fundamental domain to be the set*

$$\{(b_1, b_2, ..., b_n)x : x \in \mathbb{R}^n, 0 \leq x_i < 1\} = \mathfrak{D}.$$

A *fundamental domain* must not contain any integral lattice points except the origin as implied by the definition and the volume of $\mathfrak{D}$ is not unique. For the purpose of this paper, we work with uni-modular lattices whose volume of the torus $(\frac{R^n}{\Lambda})$ is 1, which is equivalent to saying that the determinant of the uni-modular lattice is 1 (Volume of the fundamental domain of a lattice equals the determinant of the lattice).

Let's look at the following cases for *discreteness* of Λ:

Case 1: $\Lambda \subset \mathbb{R}^2$ is discrete if $\exists\ v \in \mathbb{R}^2$ with a minimum norm $\|v\|$ such that $\Lambda = \{mv \mid m \in \mathbb{Z}\}$.

Proof. We want to show that $\mathfrak{D} \cap \Lambda = \{v\}$.

By way of contradiction, let w be another vector such that $w = \mathfrak{D} \cap \Lambda$. Then since Λ is an additive subgroup, $v - w \in \Lambda$, but $\|v - w\| < \|v\|$ which is a contradiction since $\|v\|$ is minimum by assumption. □

Note that in this case Λ is not a lattice since the *fundamental domain* does not exist because it is a one dimensional vector space with infinite co-volume.

Let's look at the second case:

Case 2: $\Lambda \subset \mathbb{R}^2$ is discrete if $\exists\ \{v_1, v_2\} \in \mathbb{R}^2$, such that $\Lambda = \{mv_1 + nv_2 \mid m, n \in \mathbb{Z}\}$, where v_1, v_2 are linearly independent.

Proof. We want to show that $\mathfrak{D} \cap \Lambda = \{0\}$

Since v_1, v_2 are linearly independent, then any vector in lattice $x = mv_1 + nv_2$ for some $m, n \in \mathbb{Z}$.

The only vector with integral combination in $\mathfrak{D}$ by definition is the zero vector, hence x must be 0. Therefore, $\mathfrak{D} \cap \Lambda = \{0\}$ and Λ is discrete. □

Not every set of linearly independent vectors can generate a lattice. Therefore, using the following lemma we determine what vectors form the basis of the lattice and the fact that $\mathfrak{D} \cap \Lambda = \{0\}$:

Lemma 6.4. *Let Λ be a lattice and $B = (b_1, ..., b_n)$ be a set of linearly independent lattice vectors. Then*

$$\mathfrak{D} \cap \Lambda = \{0\} \iff B \textit{ forms a basis of } \Lambda.$$

Proof. '$\Rightarrow$':

If $\mathfrak{D} \cap \Lambda = \{0\}$, then since B is a set of linearly independent vectors, then any lattice point, say x, can be written as a linear combination of them,

$$x = \sum_{i=1}^{n} y_i b_i,\ y_i \in \mathbb{R}$$

Since lattices are *discrete, additive* subgroups of $\mathbb{R}^n$, this implies that

$$z = \sum_{i=1}^{n} y_i b_i - \sum_{i=1}^{n} \lfloor y_i \rfloor b_i \text{ for some } z \in \Lambda$$

Now $z \in \Lambda$ and $z \in \mathfrak{D}$ by assumption,

$$z \in \Lambda \cap \mathfrak{D} = \{0\} \Rightarrow z = \{0\}$$
$$\sum_{i=1}^{n} y_i b_i = \sum_{i=1}^{n} \lfloor y_i \rfloor b_i \Rightarrow y_i = \lfloor y_i \rfloor \Rightarrow y_i \in \mathbb{Z}$$

This implies that x is an integral combination of B, hence B is a basis.

'$\Leftarrow$':

If B forms a basis of Λ, then all lattice points are the integral linear combination of vectors in B. By the definition of *fundamental domain*, the only integral combination of basis vectors in $\mathfrak{D}$ is the zero vector. Hence, $\Lambda \cap \mathfrak{D} = \{0\}$ □

Remark 6.5. *A fundamental domain is a subset of the space of lattice which holds in itself the images of every vector in the lattice. The following lemma shows why each vector in a lattice can be mapped into the fundamental domain.*

Lemma 6.6. *Let Λ be a lattice generated by the basis $B = \{v_1, v_2\}$ such that $\Lambda = \{mv_1 + nv_2 \mid m, n \in \mathbb{Z}\}$. Then $\forall x, y \in \mathbb{R}^2, \exists\ m, n \in \mathbb{R}$ such that $(x, y) + (mv_1 + nv_2) \in \{av_1 + bv_2 \mid 0 \leq a, b < 1\}$*

Proof. Given that $(x, y) \in \mathbb{R}^2$, let $(x, y) = Mv_1 + Nv_2$, for some $M, N \in \mathbb{R}$

Choose $m, n \in \mathbb{R}$ such that the integral part of M, m and N, n is the same, then,

$$(Mv_1 + Nv_2) - (mv_1 + nv_2) \in \{av_1 + bv_2 \mid 0 \leq a, b < 1\}$$

□

Previously, we said that the same lattice can be generated by two different bases. However, we can generate one basis from the other using a uni-modular matrix.

Lemma 6.7. [5] *Two bases A, B in $\mathbb{R}^2$ are equivalent iff $A = BU$ or $B = AU$ for some uni-modular matrix U.*

Proof. '$\Rightarrow$':

Two bases A, B are equivalent if the lattices generated by them are equivalent,

$$\Lambda(A) = \Lambda(B)$$

Then there exists $U \in \mathcal{M}(\mathbb{Z})$ s.t. $A = BU$ and $U \in \mathcal{M}(\mathbb{Z})$ s.t. $B = AV$

Then $B = (BU)V$

$$B^T B = (BUV)^T BUV = (V^T U^T B^T)B(UV) = (UV)^T (B^T B)(UV)$$

Taking the determinant of the $L.H.S$ and $R.H.S$ of the above equality,

$$|B^T B| = |VU|^2 |B^T B|$$
$$|VU|^2 = 1 \Rightarrow |VU| = \pm 1$$
$$|V| = \pm 1, |U| = \pm 1$$

'$\Leftarrow$':

If $A = BU$ then $\Lambda(A) \subseteq \Lambda(B)$. Then $B = AU^{-1}$ means that $\Lambda(B) \subseteq \Lambda(A)$ as U^{-1} is uni-modular.

which implies that $\Lambda(A) = \Lambda(B)$ □

6.2. **Coset space of Lattices X_2.**

Now we are ready to define the coset space $X_2 := \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$ which parametrizes the space of uni-modular lattices.

We define,

$$X_2 = \{\Lambda < \mathbb{R}^2 | \Lambda \text{ discrete, } \frac{\mathbb{R}^2}{\Lambda} \text{ has co-volume } 1\}$$

Note that the following map given by $f(g) = g * \mathbb{Z}^2$ is not one-one:

$$f\colon \mathrm{SL}_2(\mathbb{R}) \to X_2, \text{ where } \mathbb{Z}^2 \in X_2.$$

Because for different matrices in $\mathrm{SL}_2(\mathbb{R})$, we might end up with the same lattice. For example:

Let,

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \to \left\{ m \begin{pmatrix} 1 \\ 0 \end{pmatrix} + n \begin{pmatrix} 0 \\ 1 \end{pmatrix} | m, n \in \mathbb{Z} \right\}$$

and

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \to \left\{ m \begin{pmatrix} 1 \\ 0 \end{pmatrix} + n \begin{pmatrix} 1 \\ 1 \end{pmatrix} | m, n \in \mathbb{Z} \right\}$$

Then the lattices generated by the two sets are the same as one can be reduced to the other via row operations. However, the mapping is onto because every lattice is an integral combination of basis vectors that come from $\mathbb{R}$.

The action of $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$ on a group of uni-modular matrices is well-defined because if we take a lattice $\Lambda \in X_2$ then $\Lambda = \{mv_1 + nv_2 \mid m, n \in \mathbb{Z}\}$.

The map is one-one because by Lemma 6.7, $\Lambda(A) = \Lambda(B)$ then $A = BU$ for some uni-modular matrix U.

It is a natural question to think about how to measure distance between two lattices?

Let

$$\Lambda_1 = \{mv_1 + nv_2 \mid m, n, \in \mathbb{Z}\}$$

and

$$\Lambda_2 = \{mw_1 + nw_2 \mid m, n, \in \mathbb{Z}\}$$

then

$$d(\Lambda_1, \Lambda_2) = d((v_1, v_2), (w_1, w_2)) = \|v_1 - w_1\| + \|v_2 - w_2\|$$

gives us the way to define distance between two lattices.

6.3. **Convergence of lattices.**

Having seen the distance between two lattices, let's dive further to interpret the notion of converging sequence of lattices. What do we mean when we say a sequence of lattices converges to a lattice and when can we say that?

We know that a sequence of lattices converges to a lattice when:

$a)\forall R,\ \exists\ k_0\ s.t.$ for $k > k_0, f(R, \Lambda_k) = f(R, \Lambda_{k_0}) \Rightarrow$ After a certain point, number of lattice points in latttices Λ_k must become equal to number of lattice points in Λ_{k_0}(R is the radius of a fixed ball in the space of lattices).

$b)$ Define $B(R, \Lambda_k) = \Lambda_k \cap B(0, R) = \{x_1, x_2, ...x_{m_k}\}$(lattice points in a ball of radius R on a lattice space). Then,

$$\exists\ \{y_1, y_2, ..., y_{m_k}\} = B(R)\ s.t.\ \forall R,\ B(R, \Lambda_k) \to B(R)$$

meaning that

$$\{x_1, x_2, ..., x_k\} \to \{y_1, y_2, ..., y_{m_k}\}$$

Let's look at the following counter-example of a non-converging sequence of lattices:

Suppose we have a sequence of the lattices of the form,

$$\Lambda_k = \left\{ m \begin{pmatrix} k \\ 0 \end{pmatrix} + n \begin{pmatrix} 0 \\ \frac{1}{k} \end{pmatrix} | m, n \in \mathbb{Z} \right\}$$

Fix a ball of radius $R = 1$. Then $\Lambda_k \cap B(0,1)$ contains all points on only y axis and as k increases, points on the y axis become closer and closer to each other which implies that there is no shortest non-zero vector, meaning that Λ_k is not discrete. This type of lattice configuration clearly does not satisfy the two afore-mentioned conditions required for convergence of lattices. Hence, the above sequence of lattices does not converge to any lattice which implies that the space of Λ_k is not compact.

6.4. **Reduced basis.**

In $\mathbb{R}^2$, a basis v_1, v_2 of a lattice Λ is called *reduced* if v_1 and v_2 are shortest length vectors in the lattice. Note that every lattice has a reduced basis because every lattice has a shortest vector. However, a sequence of lattices having their respective reduced basis may not have a basis to which those reduced bases converge to (refer to the counter-example above (in §6.3)). Existence of reduced basis for a sequence is therefore guaranteed if the basis vectors in the sequence are bounded. Let's look at the following lemma that determines that we have found the shortest length vectors for the basis of a lattice.

Lemma 6.8. *(Definition of reduced basis) Let $v_1 = (x, y)$ be the shortest length non-zero vector such that either $x \geq 0$ or if $x = 0$, then $y > 0$. Then $\{v_1, v_2\}$ forms the reduced basis of a lattice Λ if v_2 satisfies the following:*

a) $det(v_1\ v_2) = 1$

b) If we replace v_2 by $v_2 + nv_1$ for some $n \in \mathbb{Z}$, then $v_2^{\perp} = v_2 + \lambda v_1$ (via Gram-Schmidt orthogonalization), where $\|\lambda\| \leq \frac{1}{2}$.

Proof. We want to achieve a bound on v_2 given v_1 is the shortest non-zero vector.

$$\begin{aligned} 1 &= \|v_1\| \ \|v_2^{\perp}\| = det(v_1 \ v_2^{\perp}) \\ &= det(v_1 \ v_2 + \lambda v_1) = det(v_1 \ v_2) + det(v_1 \ \lambda v_1) \\ &= det(v_1 \ v_2) + 0 = det(v_1 \ v_2) \end{aligned} \tag{9}$$

Now, given that $v_2^{\perp} = v_2 + \lambda v_1$, taking norms,

$$\|v_2^{\perp}\| \leq \|v_2\| + \|\lambda v_1\| \Rightarrow \|v_2^{\perp}\| \leq \frac{1}{\|v_1\|} + |\lambda|\|v_1\|$$

Since $\|v_1\| \geq \epsilon, \|v_2^{\perp}\| \leq \frac{1}{\epsilon} + |\lambda|\|v_1\|$

Since v_1 is the shortest non-zero vector,

$$\begin{aligned} &\|v_1\|^2 \leq \|v_2\|^2 \\ &\|v_1\|^2 \leq \|v_2^{\perp} - \lambda v_1\|^2 = \|v_2^{\perp}\|^2 + |\lambda|^2\|v_1\|^2 \\ &(1 - |\lambda|^2)\|v_1\|^2 \leq \|v_2^{\perp}\|^2 \end{aligned} \tag{10}$$

Since $|\lambda|$ is at most $\frac{1}{2}$, from (9) and (10),

$$\frac{3}{4}\|v_1\|^2 \leq \|v_2^{\perp}\|^2 \leq \frac{1}{\|v_1\|^2}$$

which implies that,

$$\|v_1\|^4 \leq \frac{4}{3} \Rightarrow \|v_1\| \leq \frac{\sqrt{2}}{3^{1/4}}$$

$$\|v_2\| = \frac{3^{1/4}}{\sqrt{2}} \text{ since } \|v_1\| \ \|v_2\| = 1$$

Hence, v_1 and v_2 are both bounded and short. Therefore, $B = \{v_1 \ v_2\}$ forms a reduced basis. □

Lemma 6.9. [2] *A sequence of lattices converges to a single lattice if and only if there exists a reduced basis of the sequence of lattices Λ_n that converges to the reduced basis of a single lattice Λ.*

Proof. Let $\{b_1^r, ..., b_n^r\}$ be the reduced basis of sequence of lattices Λ_r and $\{b_1, ..., b_n\}$ be the reduced basis of a lattice Λ. Then we want to show that

$$\Lambda_r \longrightarrow \Lambda \Leftrightarrow b_i^r \longrightarrow b_i \text{ for } 1 \leq i \leq n$$

Existence of such basis is proven in Lemma 6.8 . We shall show the convergence by using linear transformations of lattices.

'$\Rightarrow$':

Suppose, $\Lambda_r \longrightarrow \Lambda$ and let A be a linear transformation which takes basis vectors of Λ_r to Λ, $b_i^r = Ab_i$. Then,

$$|b_i^r - b_i| = |Ab_i - b_i| = |(A - I)b_i|$$

where I is the identity transformation.

By the following property of linear transformations,

$$|Ax| \leq \sqrt{n}\ \|A\||x|$$

we get,

$$|(A - I)b_i| \leq \sqrt{n}\ \|A - I\|\ |b_i|$$

By assumption, $\Lambda_r \longrightarrow \Lambda$ implies that $\Lambda_r = A\Lambda$ and $\|A - I\| \to 0$.

This immediately proves that $b_i^r \longrightarrow b_i$.

'$\Leftarrow$':

Now suppose $b_i^r \longrightarrow b_i$. Let $b_i^r = Ab_i$, then since,

$$\begin{aligned}
&|b_i^r - b_i| \to 0\\
&|Ab_i - b_i| \to 0\\
&|(A - I)b_i| \to 0 \Longrightarrow \|A - I\||b_i| \to 0
\end{aligned}$$

This implies that $\|A - I\| \to 0$ because the basis vectors are non-zero and $A - I$ is a linear transformation.

Therefore, $\Lambda_r \longrightarrow \Lambda$. □

A sequence of lattices may not converge when a lattice could go off to infinity. Since there was no shortest vector (discreteness condition was not fulfilled) as stated in the counter-example before, the sequence of lattices did not converge. Therefore,

$$\Lambda_n \to \infty \Leftrightarrow \exists\ v_n \in \Lambda_n\ s.t.\ \|v_n\| \to 0$$

Now we are ready to state and prove Mahler's compactness criterion:

6.5. Mahler's compactness criterion.

Theorem 6.10 (Mahler's criterion)**.** *For every $\epsilon > 0$, the set*

$$\{\Lambda \in X_2 \mid \forall v \in \Lambda, \|v\| \geq \epsilon\}$$

is compact.

Proof. To prove that the set is compact, it suffices to find a converging sequence of lattices in X_2 that converge to a single lattice Λ. Passing to a subsequence, this is equivalent to showing that there exists a reduced basis of the subsequence that converges to reduced basis of a single lattice Λ (Lemma 6.9). Since one can find shortest non-zero vectors $\{v_1, v_2\}$ because the vectors in such lattices are bounded (Lemma 6.8), they become the reduced basis of the subsequence. Therefore, the subsequence converges to a single lattice by Lemma 6.9, proving that the set is compact. $\square$

7. Non-divergence of unipotent flows in $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$

The final section of this paper aims to apply the results and theory of the previous sections in proving the non-divergence of *horocycle flows*. The non-divergence result coupled with Theorem 1.1 will prove the fact that minimal sets for horocycle flows are compact. We will start by introducing the reader to concept of a *horocycle flow*, and the quality of non-divergence that it portrays, thereby proving the non-divergence theorem. Combining the non-divergence result with the result from Mahler's criterion will help us prove the main result of this paper.

Definition 7.1. *The action of* $u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ *(the standard unipotent) gives the horocycle flow, where the one-parameter unipotent subgroup of* $\mathbb{R}$ *is given by*

$$U = \left\{ \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} \mid t \in \mathbb{R} \right\}$$

Theorem 7.2. [4] *Let* $\{u_t\}$ *be a one-parameter unipotent subgroup of* $\mathrm{SL}_2(\mathbb{R})$. *Then for any lattice* Λ *in* $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z}), u_t\Lambda$ *does not tend to* ∞ *as* $t \to \infty$.

Equivalently, there exists an $\epsilon > 0$, *such that* K_ϵ *is compact and* $K_\epsilon \subset X_2 := \{\Lambda \in X_2 |\ inf\ \|v\| > \epsilon\}$, *such that the set* $\{t \in \mathbb{R}_+ : u_t\Lambda \in K_\epsilon\}$ *is unbounded, where* $\Lambda \in X_2$. *This property is called non-divergence.*

Proof. We will show that the time spent by the points under the action of horocycle flow outside the compact set is very small. Equivalently,

$$m(\{t \mid 0 \leq t \leq T, u_t\Lambda \notin K_\epsilon\}) \leq \epsilon T$$

where m is the length of the interval for all $T > 0$ and all $\Lambda \in X_2$.

The action of the *horocycle flow* fixes the x-axis and shears the y-axis towards the direction of x-axis. Specifically, the image of vector (x, y) under the action of $\begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$ is $(x + ty, y)$. It suffices to prove the theorem with the use of primitive vectors (vectors v such that the equation $v = nw$ with $n \in \mathbb{Z}$ and $w \in \Lambda$ implies that $n = \pm 1$). Why?

If v is any vector in Λ, then for some time t, there may exist a short vector $u_t v$ which prevents $u_t\Lambda$ to lie in K_ϵ.

Then for each primitive vector v, define the following:

$$J_v = \{t \in [0, T] \mid \|u_t v\| \leq \epsilon'\};$$

$$I_v = \{t \in [0, T] \mid \|u_t v\| \leq \rho\}$$

where $\epsilon' < \rho$, which implies that $J_v \subset I_v$.

J_v is the set of bad times for the vector v and I_v is the set of protecting times for v. We will use the following lemma to show that for any two primitive vectors v and v' in the regions of J_v and I_v, v must equal v'.

Lemma 7.3. *For any $\Lambda \in X_2$, the cardinality of the set*

$$\{\Lambda \cap \{v \in \mathbb{R}^2 : \|v\| < \frac{1}{2}\}\}$$

is 2.

Proof. We want to show that no two linearly independent vectors can have norm less than 1.

Suppose not, then $\exists t \in I_v \cap I_{v'}$ such that $\|u_t v\|\|u_t v'\| < 1$ for linearly independent v, v' in the above set. This gives us a contradiction as we are working in the space of uni-modular lattices whose co-volume is 1. Therefore, the vectors for which $\|u_t v\|\|u_t v'\| < 1$, are the primitive vectors. Hence, $v = \pm v'$ or in other words, the cardinality of the set is 2 $(v, -v)$. □

Since $v = \pm v', \|u_t v\| = \|u_t v'\| \Rightarrow I_v = I_{v'}$.

Now we claim that,

Claim 7.4. [4]

$$m(J_v) \leq \frac{8\epsilon'}{\rho} m(I_v)$$

for some primitive vector v.

Proof. Since, $J_v = \{t \in [0,T] \mid \|u_t v\| \leq \epsilon'\}$, let $v = \begin{pmatrix} x \\ y \end{pmatrix}$, such that $\|u_t v\| = \begin{pmatrix} x+ty \\ y \end{pmatrix} \leq \epsilon'$.

This implies that $|x+ty| < \epsilon'$ and $|y| < \epsilon'$, therefore,

$$\begin{aligned} &-\epsilon' < x+ty < \epsilon' \\ &-\epsilon - x < ty < \epsilon' - x \\ &\frac{-\epsilon'}{y} - \frac{x}{y} < t < \frac{\epsilon'}{y} - \frac{x}{y} \end{aligned}$$

which gives length of time spent by v inside the region of J_v,

$$m(J_v) \leq \frac{2\epsilon'}{y}. \tag{11}$$

Similarly, let $I_v = [t_1, t_2]$, then the interval must contain some t', such that $\|u_{t'} v\| < \rho \Rightarrow |x + t'y| \leq \frac{\rho}{4}$

$\|u_{t_1} v\| = \rho$ gives $|x + t_1 y| \geq \frac{\rho}{2}$ if $|y| \leq \frac{\rho}{2}$

Therefore, from previous equations,

$$m(I_v) = t_2 - t_1 \geq t' - t_1 \geq \frac{\rho}{4y} \tag{12}$$

(11) and (12) imply that

$$\frac{m(J_v)}{m(I_v)} \leq \frac{\frac{2\epsilon'}{y}}{\frac{\rho}{4y}} \leq \frac{8\epsilon'}{\rho}$$

$\square$

Therefore,

$$\begin{aligned} &m(\{t \mid u_t\Lambda \notin K_\epsilon\}) = m(\{t \in [0,T] \mid \|u_t v\| \leq \epsilon'\}) \\ &m(\{t \in [0,T] \mid \|u_t v\| \leq \epsilon'\}) = m(\cup\, J_v) \\ &m(\cup\, J_v) \leq \sum m(J_v) \leq \frac{8\epsilon'}{\rho} \sum m(I_v) \leq \frac{8\epsilon'}{\rho} m(\cup I_v) \leq \epsilon T \end{aligned}$$

when $\epsilon' \leq \frac{\epsilon\rho}{8}$

Hence,

$$m(\{t \mid u_t\Lambda \notin K_\epsilon\}) \leq \epsilon T$$

The proof shows that the time spent by orbits under action of u_t outside the compact set K_ϵ is very tiny. Therefore, most points return to the compact set K_ϵ showing that $\{t \in \mathbb{R}_+ : u_t\Lambda \in K_\epsilon\}$ is unbounded. $\square$

Now we can combine the result of Theorem 6.10, the non-divergence result and Theorem 1.1 to prove Theorem 1.2 given by:

Theorem 7.5. *Any minimal set for the action of U on X_2 is compact.*

Proof. Let U be the action given by left multiplication with the subgroup

$$U := \left\{ u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} : t \in \mathbb{R} \right\}$$

acting on the space X_2.

Let Z be any non-empty, closed u_t - invariant subset, then we want to show that Z is compact.

We will use Theorem 1.1 to derive the conclusion. In order to show that Z is compact, we must satisfy the following three conditions:

1) $\exists\ V \subset Z, V$ open subset such that it's closure is compact.

2) $\exists x \in Z$, s.t. $\forall u_{s_0} \in U, \exists K = K(x, s_0)$, s.t. K is compact and $\{ns_0 > 0 : u_{ns_0}x \in K\}$ and $\{ns_0 < 0 : u_{ns_0}x \in K\}$ are unbounded.

3) $UV = Z$.

Let's see the proof for *Condition 1*:

Proof. Let $x \in Z$, then we can choose V to be the compact neighborhood of x which lies in Z, given by $V = B(x, 1) \cap\ Z$. Closure of V is compact because Z is closed due to minimality and B is a compact neighborhood. $\square$

Let's see how *Condition 2* unfolds:

Proof. Let K' be compact set given by applying non-divergence to $x \in Z$, s.t. $\{t \in \mathbb{R}_+ : u_t x \in K'\}$ is unbounded (Theorem 7.2).

We want to find $K = K(K', s_0)$ and a sequence $n_i \to \infty$ such that $u_{n_i s_0} \in K$.

First, we will show that such a sequence of natural numbers n_i exists. Why?

Assume, $t_1 < t_2 < t_3 < ...$ so that

$$t_{i+1} - t_i > s_0 + 1$$

Therefore,

$$\forall i, \exists\, n_i \in \mathbb{N} \text{ so that } t_i < n_i s_0 < t_{i+1}$$

which implies that

$$n_i s_0 - t_i \leq s_0$$

Let

$$K = \{u_\tau x \mid |\tau| \leq s_0\}K'$$

where

$$\tau = n_i s_0 - t_i$$

then K is compact.

We claim that $u_{n_i s_0} x \in K$ because

$$u_{n_i s_0} x = u_{t_i + (n_i s_0 - t_i)} x = u_{(n_i s_0 - t_i)}(u_{t_i} x)$$

where $u_{t_i} x \in K'$.

Therefore, K is the compact subset for which $\{ns_0 > 0 : u_{ns_0} x \in K\}$ is unbounded, satisfying *condition 2* of Theorem 1.1.

A similar argument holds for $ns_0 < 0$. $\square$

Condition 3 can be proven as follows:

Proof. $UV \subseteq Z$ because $V \subset Z$ and Z is U-invariant.

To show that $Z \subseteq UV$, let $z \in Z$, then we want to find $u_t \in U$, s.t. $u_t z \in V$.

Since Z is minimal, the orbit of z under the action of u_t is dense in Z, which means $Uz \cap V \neq \phi$ since $V \subset Z$. Therefore, $Z \subseteq UV$.

Hence, $UV = Z$. $\square$

Hence, fulfillment of all conditions of Theorem 1.1 concludes that Z is compact, or *any minimal subset for the action of* U *on* X_2 *is compact.*

$\square$

Acknowledgements

I would like to extend my sincere gratitude towards my advisor for his relentless support and guidance throughout this project. Without his patience and valuable time that he spent on correcting me, this paper would not be a reality. I would also like to thank my friends and family for their support and for letting me explore my curiosity with belief in my work.

References

[1] Borris Hasselblatt and Anatole Katok. *A First Course in Dynamics With a Panorama of Recent Developments*. Cambridge Univ. Press, 2003.
[2] John W.S Cassels.*An Introduction to Geometry of Numbers*. Springer- Verlag Berlin Heidelberg. 1997
[3] G.A. Margulis. *Compactness of Minimal Closed Invariant Sets of Actions of Unipotent Groups*. 1989
[4] Manfred Einsiedler and Thomas Ward. *Ergodic Theory with a view towards Number Theory*. Springer-Verlag London Limited, 2011.
[5] Oded Regev. *Lattices in Computer Science: Introduction*. Lecture, Fall 2004.

Department of Mathematics, University of California, San Diego