Dynamics on Homogeneous Spaces and Counting Lattice Points

Thomas Zamojski

Fall 2011

Contents

Introduction

The goal of this course is to investigate certain elements of dynamics on homogeneous spaces and their applications to number theory.

Dynamical Systems

A dynamical system is given by the data (X, G, α) of

- a *phase space* X, the elements of which can be referred to as states. One can impose more structure, for example in topological dynamics, X should carry a topology.
- a *time space* G consisting of a locally compact second countable topological group (or semigroup). Classically G is the natural numbers $\mathbb{N}$, the integers $\mathbb{Z}$ or the real numbers $\mathbb{R}$ with the usual topologies, but more general groups will occur in these notes.
- a *law of evolution* (or rule of evolution) α consisting of a continuous (semi)-group action $\alpha : G \times X \to X$.

It follows from the definition of a group action that the rule of evolution has the following two characteristics: it is deterministic and time-independent. This means that the "future" is completely determined by the current state and this independently of the "present" time.

A nowadays obvious example in nature of a dynamical system is our solar system. The phase space consists of the position and momentum of the planets, time is our common notion of time , and the rule of evolution is governed by the laws of gravity.

Another example is given by an invertible map $T : X \to X$. It induces an action of $\mathbb{Z}$ by iteration of composition of functions: $n.x := T^n(x)$. Conversely, all $\mathbb{Z}$-actions arise this way.

Many fields of mathematics study groups of automorphisms preserving a certain structure, the main one coming to mind being geometry. What distinguishes dynamical systems from such fields however is its emphasis on asymptotic behaviour of the system. Limiting properties studied include qualitative ones such as periodicity, recurrence and stability, etc., but also some quantitative invariants such as entropy, growth of orbits, growth of the number of periodic orbits, etc.

Homogeneous Spaces

A homogeneous space X is one that looks the same from every point. More precisely, it admits a group G of automorphisms that acts transitively: for every $x, y \in X$, there exists $g \in G$ such that $gx = y$. Such spaces include euclidean space, hyperbolic space, spheres, groups themselves and many others. Because of the abundance of symmetries, one can usually prove more theorems about homogeneous spaces, hence their predominance in mathematics.

For us, at least G should be a locally compact second countable group acting continuously on X. The stabiliser of a point is then a closed subgroup. Upon choosing a group G acting transitively on X and a base point with stabiliser H in G, X identifies with the coset space G/H.

We can let subgroups of G act by left multiplication on the cosets. If the dynamics of G acting on X is trivial (one single orbit), closed subgroups of G on the contrary can have interesting dynamics. This is due to the possible interplay between subgroups and the stabiliser. In particular, if this interaction is arithmetic in nature, dynamical properties of the system can lead to powerful statements in number theory. In fact, the study of dynamics on homogeneous spaces was initiated from number theory.

We will early on specialise our study to real or p-adic Lie groups. However, we will avoid having to use structure theory as much as possible, and the reader not aquainted with it should be thinking of $SL_n(\mathbb{R})$. In fact some theorems will be proved only for $SL_2(\mathbb{R})$.

Applications to Number Theory

Conventions

All topological groups are assumed to be locally compact second countable. A space X with a G-action, on the left unless otherwise mentioned, can be termed a *dynamical system* as well as a *G-space*. If $x \in X$ and $g \in G$, we write gx for the element of X given by the action of g on x. The *orbit* of x is

$Gx = \{gx : g \in G\}$, and the *stabiliser* of x is $Stab_G(x) = \{g \in G : gx = x\}$. One can pull-back the action to functions on X: for a function f on X, we define $gf(x) := f(g^{-1}x)$. We also have an induced action on measures: if μ is a measure on X and A a subset, then $g\mu(A) := \mu(g^{-1}A)$.

The measured dynamical systems (X, G, μ) encountered in these notes will always be *nonsingular*: for any subset A and $g \in G$, if $\mu(A) = 0$ then $\mu(g^{-1}A) = 0$. In other words, G preserves the measure class of μ. It models situations where nothing is created out of nothing. Eventually, a main interest will be to study invariant measures. These satisfy the stronger condition $\mu(g^{-1}A) = \mu(A)$ for all measurable set A and all $g \in G$. Here, G preserves not only the measure class, but the measure itself.

A *σ-algebra* (or a *σ-field*) is a family of subsets of X closed under complementation and countable union (hence under countable intersection). The Borel σ-algebra $\mathcal{B}$ of X is the smallest algebra containing all open sets. In general, the σ-algebra $\sigma(\mathcal{F})$ generated by a family $\mathcal{F}$ of subsets is the smallest σ-algebra containing $\mathcal{F}$. A subset $B \in \mathcal{B}$ is called a *Borel set*.

A measure μ is *Borel regular* if every Borel set is μ-measurable and for each subset A (not necessarily measurable), there exists a Borel set B such that $A \subset B$ and $\mu(A) = \mu(B)$. Measures are always assumed to be Borel regular.

A *Borel space* $(X, \mathcal{B})$ is one that is isomorphic as a pair to a dense Borel subset of a compact metric space together with the restricted Borel σ-algebra. A *Borel G-space* is a Borel space with a Borel G-action.

A set of μ-measure zero is called *μ-null*. The complement is then *μ-conull* or of μ-full measure.

One

Elements of Ergodic Theory

From a dynamical point of view, ergodic theory comes from trying to quantify various recurrence properties. The statistics that will be of interest is the asymptotic frequency at which recurrence occurs. In order to study such statistical behaviour of orbits, it will be necessary to appeal to measure theory. Ergodicity will then give knowledge of the asymptotic behaviour of typical points or almost all points. In general, it is not clear how much information this gives about the dynamical system. However, for certain systems and choices of a measure, it gives valuable information.

1.1 Conditional Expectations and Measures

We recall that a *σ-algebra* (or a *σ-field*) is a family of subsets of X closed under complementation and countable union (hence under countable intersection). It represents the collection of observable events. The Borel σ-algebra $\mathcal{B}$ of X is the smallest algebra containing all open sets. In general, the σ-algebra $\sigma(\mathcal{F})$ generated by a family $\mathcal{F}$ of subsets is the smallest σ-algebra containing $\mathcal{F}$. Let X be equipped with a Borel measure μ.

Conditional measures come from assigning a measure to an event having some prior knowledge of it. It is introduced via the related conditional expectations, which are averaging operators. We begin with two fundamental examples.

Example 1.1.1. Suppose we have a partition $\{A_1, A_2, ...\}$ of X by sets of positive measure. Let $\mathcal{A}$ be $\sigma(A_1, A_2, ...)$ and f a function, we define the conditional expectation of f with respect to $\mathcal{A}$ as the function

$$\mathbb{E}^{\mathcal{A}} f\,(x) := \frac{1}{\mu(A_k)} \int_{A_k} f\, d\mu, \text{ for } x \in A_k.$$

This function is $\mathcal{A}$-measurable, meaning in this case that it is constant on the partition elements. For each A_k, we have the conditional measures defined by

$$\mu_{|A_k}(B) := \frac{\mu(A_k \cap B)}{\mu(A_k)}, \text{ for } B \in \mathcal{B}.$$

Succinctly, we can define a family of measures by $\mu_x^{\mathcal{A}} := \mu_{|A_k}$ for $x \in A_k$. Then conditional expectation and measure are related by the following:

$$\mu_x^{\mathcal{A}}(B) = \mathbb{E}^{\mathcal{A}}\mathbb{1}_B(x) \text{ and conversely } \mathbb{E}^{\mathcal{A}}f(x) = \int_X f\, d\mu_x^{\mathcal{A}}.$$

Example 1.1.2. Let $(Y, \mathcal{B}_Y, \mu)$ and $(Z, \mathcal{B}_Z, \nu)$ be finite Borel spaces (e.g. $[0,1] \subset \mathbb{R}$), and let $X = Y \times Z$ equipped with the product measure. Let $\mathcal{A}$ be the σ-algebra $\mathcal{B}_Y \times \{\emptyset, Z\}$. For an integrable function $f(y,z)$ and for almost all $y \in Y$, $f(y,z)$ is integrable as a function of z with respect to ν. We define

$$\mathbb{E}^{\mathcal{A}}f\,(y,z) = \frac{1}{\nu(Z)} \int_Z f(y,z)\, d\nu.$$

This function is $\mathcal{A}$-measurable, meaning independent of z. Note that f is being averaged over sets $X_y := \{y\} \times Z$ of $\mu \times \nu$-measure zero. We can define conditional measures by $\mu_{y,z}^{\mathcal{A}} = \delta_y \times \nu/\nu(Z)$, that is

$$\mu_{y,z}^{\mathcal{A}}(B) := \frac{\nu(B \cap X_y)}{\nu(X_y)} = \frac{\nu(B_y)}{\nu(Z)}.$$

The conditional expectation and measure are again related by

$$\mu_{y,z}^{\mathcal{A}}(B) = \mathbb{E}^{\mathcal{A}}\mathbb{1}_B(y,z) \text{ and conversely } \mathbb{E}^{\mathcal{A}}f\,(y,z) = \int_X f(y,z)\, d\mu_{y,z}^{\mathcal{A}}.$$

Recall that an *atom* of a measure is a smallest set of positive measure. In the first example, the measure $\mu_{|\mathcal{A}}$ is completely atomic, with atoms the sets A_k. In the second example however, $\mu_{|\mathcal{A}}$ has no atoms. In general, conditional expectations always exist:

Theorem 1.1.3. *Let $(X, \mathcal{B}, \mu)$ be a finite measure space, and $\mathcal{A} \subset \mathcal{B}$ a sub-σ-algebra. There exists an averaging map $\mathbb{E}^{\mathcal{A}} : L^1(X, \mathcal{B}, \mu) \to L^1(X, \mathcal{A}, \mu_{|\mathcal{A}})$ characterised by $\mathbb{E}^{\mathcal{A}}f$ being $\mathcal{A}$-measurable and for any $A \in \mathcal{A}$,*

$$\int_A \mathbb{E}^{\mathcal{A}}f\, d\mu = \int_A f\, d\mu.$$

Proof. It is enough to prove the theorem for $f \geq 0$. The measure $\nu := f\mu$ is absolutely continuous with respect to μ, and the same holds for the restrictions to $\mathcal{A}$. Then the image is given by the Radon-Nikodym derivative:

$$\frac{D\nu_{|\mathcal{A}}}{D\mu_{|\mathcal{A}}} = \mathbb{E}^{\mathcal{A}} f.$$

Now suppose we have two $\mathcal{A}$-measurable functions g_1, g_2 satisfying the theorem, then $A = \{x \in X \,:\, g_1(x) < g_2(x)\}$ is in $\mathcal{A}$, and thus

$$\int_A g_1 \, d\mu = \int_A f \, d\mu = \int_A g_2 \, d\mu.$$

Hence $\mu(A) = 0$. Similarly for the set where $g_1 > g_2$, and therefore g_1 and g_2 agree μ-a.e. □

Another proof is by showing $\mathbb{E}^{\mathcal{A}} : L^2(X, \mathcal{B}, \mu) \to L^2(X, \mathcal{A}, \mu_{|\mathcal{A}})$ is the orthogonal projection, and extending to L^1 by continuity. This can be done using the properties below and is left as an exercise.

Recall that equality of L^1 functions means that any representatives agree almost everywhere. We can state the fundamental properties of conditional expectations as follows.

Proposition 1.1.4. *The map $\mathbb{E}^{\mathcal{A}}$ of Theorem 1.1.3 satisfies the following:*

i) The operator $\mathbb{E}^{\mathcal{A}}$ is linear, positive and of norm one.

ii) For $f \in L^1(X, \mathcal{B}, \mu)$ and $g \in L^\infty(X, \mathcal{A}, \mu_{|\mathcal{A}})$,

$$\mathbb{E}^{\mathcal{A}}(gf) = g \, \mathbb{E}^{\mathcal{A}} f.$$

iii) If $\mathcal{A}' \subset \mathcal{A}$ is a sub-σ-algebra, then $\mathbb{E}^{\mathcal{A}'} f = \mathbb{E}^{\mathcal{A}'}(\mathbb{E}^{\mathcal{A}} f)$.

iv) If $f \in L^1(X, \mathcal{A}, \mu_{|\mathcal{A}})$, then $\mathbb{E}^{\mathcal{A}} f = f$.

v) $\left|\mathbb{E}^{\mathcal{A}} f\right| \leq \mathbb{E}^{\mathcal{A}} |f|$.

Proposition 1.1.5 (Jensen's Inequality)**.** *On a probability space X, if ϕ is a convex function, then*

$$\phi \circ (\mathbb{E}^{\mathcal{A}} f) \leq \mathbb{E}^{\mathcal{A}}(\phi \circ f).$$

The proofs of these propositions are left as an exercise.

To define conditional measures, one has to be careful with null sets. The issue is that $\mu_x^{\mathcal{A}}(B) := \mathbb{E}^{\mathcal{A}} \mathbb{1}_B(x)$ is defined for μ-a.e. x. However, the implied null set depends on B, whereas one needs $\mu_x^{\mathcal{A}}$ to be a measure on all of $\mathcal{B}$. Fortunately, for most spaces being considered in these notes, such a definition is possible.

Definition 1.1.6. A finite measure Borel space $(X, \mathcal{B}, \mu)$ is a dense Borel subset X of a compact (separable) metric space $\bar{X}$, together with the restricted Borel structure $\mathcal{B}$, and with a finite Borel measure μ.

The measure μ identifies with a measure μ on $\bar{X}$ such that $\mu(\bar{X} \setminus X) = 0$ and vice versa.

Theorem 1.1.7. *Let $(X, \mathcal{B}, \mu)$ be a finite measure Borel space, $\mathcal{A} \subset \mathcal{B}$ a sub-σ-algebra. Then there exists an essentially unique family of probability measures $\{\mu_x^{\mathcal{A}} : \mu$-a.e. $x \in X\}$ on X such that for μ-a.e. $x \in X$, for all μ-integrable function f,*

$$\mathbb{E}^{\mathcal{A}} f\,(x) = \int_X f\, d\mu_x^{\mathcal{A}}.$$

The proof follows a diagonal argument for compact metrisable spaces that will occur at least two other times. We begin by recording a useful lemma.

Lemma 1.1.8. *Retaining the notation of Theorem 1.1.7, if the family of measures $\mu_x^{\mathcal{A}}$ exists a.e., then for any μ-null set N, $\mu_x^{\mathcal{A}}(N) = 0$ μ-a.e.*

Proof. Let $f = \mathbb{1}_N$, the definition of conditional expectation leads to

$$\int_X \mu_x^{\mathcal{A}}(N)\, d\mu = \mu(N) = 0.$$

□

Proof of Theorem 1.1.8. By Lemma 1.1.8, we can assume $X = \bar{X}$ is a compact metric space, therefore separable. Choose $\{f_0 = 1, f_1, ...\}$ a countable dense $\mathbb{Q}$-vector space of continuous functions. For these, one can choose a μ-full measure set on which $\mathbb{E}^{\mathcal{A}} f_n$ is defined for all n. Possibly throwing out another μ-null set, $\mathbb{E}^{\mathcal{A}} \cdot (x)$ is furthermore $\mathbb{Q}$-linear, bounded by the sup-norm and positive on $\{f_0, f_1, ...\}$. It therefore extends by continuity to all continuous functions. From Riesz Representation Theorem, there exists a family of measures $\{\mu_x^{\mathcal{A}}\}$ for μ-a.e. x representing these functionals. The fact that $\mu_x^{\mathcal{A}}$ is a probability measure is seen by considering $f = 1$.

Remains to show that for any μ-integrable function f,

i) $x \mapsto \displaystyle\int_X f\, d\mu_x^{\mathcal{A}}$ is $\mathcal{A}$-measurable

ii) $\displaystyle\int_A \int_X f\, d\mu_x^{\mathcal{A}}\, d\mu = \int_A f\, d\mu$ for any $A \in \mathcal{A}$.

By definition, these are already true for the countable collection $\{f_k\}$. Let $A \in \mathcal{A}$, $\varepsilon > 0$ and f_k be such that $|f - f_k| < \varepsilon$. Then

$$\begin{aligned} \left|\int_A \int_X f \, d\mu_x^{\mathcal{A}} \; - f \, d\mu\right| &\leq \left|\int_A \int_X f - f_k \, d\mu_x^{\mathcal{A}} \; d\mu\right| + \left|\int_A \int_X f_k \, d\mu_x^{\mathcal{A}} \; - f \, d\mu\right| \\ &\leq \int_A \int_X |f - f_k| \, d\mu_x^{\mathcal{A}} \; d\mu + \left|\int_A f_k - f \, d\mu\right| \\ &\leq 2\varepsilon\mu(A). \end{aligned}$$

Since μ is of finite mass and ε is arbitrary, this shows i) and ii) for continuous functions.

To prove it for integrable functions, we will freely use the dominated convergence theorem and the monotone convergence theorem. Let $\mathcal{M}$ be the monotone class of subsets B such that $\mathbb{1}_B$ satisfies the two conditions above. Since the conditions hold for continuous functions, and that for an open set U, there exists a sequence of continuous functions f_n increasing to $\mathbb{1}_U$, $\mathcal{M}$ contains open sets, and thus closed sets also. Similarly, it contains all G_δ-sets and F_σ-sets.

Let $\mathcal{R} = \{\bigsqcup_{k=1}^m U_k \cap F_k \; : U_k \text{ is open and } F_k \text{ is closed.}\}$. Since each $U_k \cap F_k$ is a G_δ-set, and by linearity of the conditions, $\mathcal{R} \subset \mathcal{M}$. By the monotone class theorem, $\sigma(\mathcal{R}) = \mathcal{B} \subset \mathcal{M}$.

We have shown that $\mathcal{M}$ contains $\mathbb{1}_B$ for any Borel set B. By linearity, it contains all Borel simple functions. By approximation, it therefore contains all positive μ-integrable function. Finally, using the positive part and negative part of a function, we conclude that the conditions are verified for all μ-integrable functions. $\square$

Corollary 1.1.9. *For a Borel map $T : (X, \mathcal{B}_X, \mu) \to (Y, \mathcal{B}_Y, \nu)$ and a sub-σ-algebra $\mathcal{A}$ of $\mathcal{B}_Y$, for μ-a.e. x,*

$$T_* \mu_x^{T^{-1}\mathcal{A}} = \nu_{Tx}^{\mathcal{A}}.$$

A σ-algebra $\mathcal{A}$ is *countably-generated* if $\mathcal{A} = \sigma(\{A_1, A_2, ...\})$. In this case it is useful to speak of the $\mathcal{A}$-*atom* of a point x:

$$[x]_{\mathcal{A}} := \bigcap_{x \in A \in \mathcal{A}} A = \bigcap_{x \in A_n} A_n \cap \bigcap_{x \notin A_n} X \setminus A_n.$$

By the last equality, it is always $\mathcal{A}$-measurable, which is not always the case if the algebra is not countably-generated.

For two σ-algebras $\mathcal{A}, \mathcal{A}'$, the relation $\mathcal{A} \subset_\mu \mathcal{A}'$ will mean that for any $A \in \mathcal{A}$, there exists $A' \in \mathcal{A}'$ such that $\mu(A \triangle A') = 0$. If $\mathcal{A} \subset_\mu \mathcal{A}'$ and $\mathcal{A}' \subset_\mu \mathcal{A}$, then we write $\mathcal{A} =_\mu \mathcal{A}'$.

Lemma 1.1.10. *Let $(X, \mathcal{B}, \mu)$ be a finite measure Borel space and $\mathcal{A} \subset \mathcal{B}$ any σ-algebra. Then there exists a countably-generated σ-algebra $\mathcal{C}$ such that $\mathcal{A} =_\mu \mathcal{C}$.*

Proof. Since $C(\bar{X})$ is separable and dense in $L^1(X, \mathcal{B}, \mu)$, the latter is separable. Since subsets of separable spaces are themselves separable, the set $\{\mathbb{1}_A : A \in \mathcal{A}\}$ is separable. Thus, there exist sets $A_1, A_2, ...$ such that for any $A \in \mathcal{A}$ and for $\varepsilon_n \to 0$, there exist a sequence $k(n)$ with $\mu(A \triangle A_{k(n)}) < \varepsilon_n$. The sequence $\mathbb{1}_{A_{k(n)}}$ is then a Cauchy sequence, converging to a function $f \in L^1(X, \mathcal{C}, \mu)$, where $\mathcal{C} = \sigma(A_1, A_2, ...)$. To conclude the proof, we note that $supp(f) \in \mathcal{C}$ satisfies $\mu(A \triangle supp(f)) = 0$. □

Then the following proposition states that we can consider only the much nicer case of countably generated algebras.

Proposition 1.1.11. *With notation as in Theorem 1.1.7, and whenever the conditional measure is defined at the various points,*

i) For $\mathcal{A}$ countably-generated, $\mu_x^{\mathcal{A}}([x]_{\mathcal{A}}) = 1$. Moreover, if $[x]_{\mathcal{A}} = [y]_{\mathcal{A}}$, then $\mu_x^{\mathcal{A}} = \mu_y^{\mathcal{A}}$.

ii) If $\mathcal{A} =_\mu \mathcal{A}'$, then $\mu_x^{\mathcal{A}} = \mu_x^{\mathcal{A}'}$ μ-a.e.

In general, it is true that conditional measures exist if and only if the σ-algebra is μ-equivalent to a countably-generated one.

1.2 Recurrence Properties

Recurrence properties as the name suggests are characterised by an event occuring over and over again. They can be of a topological nature as well as measure theoretic.

A basic theorem due to Poincare is that a measure-preserving transformation on a *finite* measure space must exhibit recurrence. This innocent looking fact will have interesting applications later.

Definition 1.2.1. A point $x \in X$ is called *recurrent* if for every neighbourhood U of x, there exists a sequence $g_k \to \infty$ such that $g_k.x \to x$. It is called *recurrent with respect to a set A* if there exists a sequence $g_k \to \infty$ such that $g_k.x \in A$.

Theorem 1.2.2 (Poincare's Recurrence Theorem)**.** *Let (X, μ) be a finite measure space, and let $T : X \to X$ be measure preserving. Then for any measurable set A, almost every point $x \in A$ is recurrent with respect to A.*

In the case that X has finite cardinality, the theorem is nothing else than the pigeonhole principle. For this reason, the theorem should be thought as its ergodic counterpart.

Proof. Consider B_n the set of points in A whose last return to A is at time n. It is measurable since

$$B_n = A \cap T^{-n}A \cap \bigcap_{k>n} T^{-k}(X \setminus A).$$

Then $T^{-m}B_n$ is a collection of points whose last return to A is at time $m+n$. Therefore $\{T^{-m}B_n : m \in \mathbb{N}\}$ is an infinite collection of disjoint sets, all of equal measure by invariance. Since the measure is finite, this is only possible if $\mu(B_n) = 0$. But then the set $\bigcup_{n=0}^{\infty} B_n$ of points of A that do not recur to A has measure zero. $\square$

This theorem is purely measure theoretic, but when X carries a topology, one can state the following convenient form of the theorem.

Corollary 1.2.3. *Let X be a separable metrisable space, μ a finite Borel measure on X and $T : X \to X$ be measure preserving. Then almost every point $x \in X$ is recurrent.*

Proof. Let $\{V_k : k \in \mathbb{N}\}$ be a countable basis for the topology of X. Let Y_k be the set of points of V_k that do not recur to V_k. The set $\bigcup_{k\in\mathbb{N}} Y_k$ of points that are not recurrent has measure zero by Poincare's Recurrence Theorem. $\square$

This means in particular that under the conditions of the corollary, the support of any invariant measure is contained in the set of all recurrent points. These are in turn contained in the closed invariant set of nonwandering points. In fact, the restriction of the dynamical system to the latter contains all information about various recurrence properties.

1.3 Asymptotic Distribution of Orbits

We consider a transformation $T : X \to X$ and its induced dynamical system. The frequency at which the orbit of $x \in X$ visits a set B prior to time N is

$$A_N(\mathbb{1}_B; x) := \frac{1}{N} \sum_{i=0}^{N-1} \mathbb{1}_B(T^i x).$$

The asymptotic behaviour of $A_N(\mathbb{1}_B;\cdot)$ as $N\to\infty$ reflects statistically recurrence to the set B. Instead of indicator functions, let's consider continuous functions and assume that an asymptotic frequency exists for all such functions. We define the *time average* of f along the orbit of x as

$$A(f;x) := \lim_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N-1} f(T^i x).$$

Then $A(\cdot;x)$ is a positive, invariant, bounded linear functional on the space of continuous functions. Therefore by the Riesz Representation Theorem, there exists an invariant measure μ_x such that $A(f;x) = \int_X f\, d\mu_x$.

Therefore the existence of a limiting frequency implies the existence of an invariant measure. One might ask

Question 1.3.1. Are there invariant measures and do they all arise this way?

In general the situation can be complicated. For finite measure Borel spaces however, the situation is more satisfying.

Changing the focus and fixing a function f instead, we have the following important theorem with a long history of its many proofs. We will give several ones in the next section.

Theorem 1.3.2 (Birkhoff's Ergodic Theorem)**.** *Let (X,μ) be a space of finite measure and $T : X\to X$ a measure preserving transformation. Let $\mathcal{I}$ be the σ-algebra of invariant subsets. For any integrable function f,*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=0}^{N-1} f(T^i x) = \mathbb{E}^{\mathcal{I}} f(x)$$

converges both in L^1 and for μ-almost every point.

The theorem can be interpreted in the following way: the limit of time averages must be invariant, and in fact converges to a degenerate average over invariant atoms. Note that most of the time $\mathcal{I}$ is not countably-generated, therefore the word degenerate.

The σ-algebra $\mathcal{I}$ could have been defined to be the σ-algebra of μ-almost invariant sets since they are μ-equivalent. The latter might be preferable since it is maximal with respect to $\subset_\mu$.

We have the following interesting answer to Question 1.3.1 for compact spaces:

Corollary 1.3.3. *For a continuous transformation $T : X\to X$ of a compact metrisable space X, the set of points x such that $A(f;x)$ exists for all continuous functions f has full measure with respect to any invariant measure.*

Proof. Bogolioubov-Krylov Theorem (Theorem **??**) states that there exists an invariant measure. Applying Birkhoff's Ergodic Theorem to a countable dense set of functions and then using approximation to extend to all continuous functions, we obtain that for any invariant measure μ and for μ-a.e. x,

$$A(f;x) = \int_X f \, d\mu_x^{\mathcal{I}}.$$

□

Convergence of time averages happens also in quadratic mean and is easier to prove. Recall that $\mathbb{E}^{\mathcal{I}} : L^2(X, \mathcal{B}, \mu) \to L^2(X, \mathcal{I}, \mu)$ is the orthogonal projection to the space of invariant functions. A measure-preserving transformation T on X induces a unitary operator U_T on $L^2(X, \mathcal{B}, \mu)$ via $U_T f = f \circ T$. Convergence of time averages is then a consequence of the following theorem.

Theorem 1.3.4 (Mean Ergodic Theorem)**.** *Let $\mathcal{H}$ be a Hilbert space and let $U : \mathcal{H} \to \mathcal{H}$ be a unitary operator. Let $\mathcal{H}_{\mathcal{I}} = \{v \in \mathcal{H} : Uv = v\}$ be the closed subspace of invariant vectors, and $P : \mathcal{H} \to \mathcal{H}_{\mathcal{I}}$ the orthogonal projection. Then for all $v \in \mathcal{H}$,*

$$\lim_{n\to\infty} \Big\| \frac{1}{n} \sum_{k=0}^{n-1} U^k v - Pv \Big\| = 0.$$

Proof. The theorem is clearly true for $v \in \mathcal{H}_{\mathcal{I}}$. By linearity, it suffices to show the theorem for ${\mathcal{H}_{\mathcal{I}}}^{\perp}$. Unitarity implies that $\mathcal{H}_{\mathcal{I}} = \{Uv - v : v \in \mathcal{H}\}^{\perp}$. Therefore one needs to show the theorem for vectors of the form $Uv - v$: we have $P(Uv - v) = 0$ and

$$\Big\| \frac{1}{n} \sum_{k=0}^{n-1} U^k (Uv - v) \Big\| = \Big\| \frac{1}{n} (U^n v - v) \Big\| \le \frac{2\|v\|}{n} \to 0.$$

□

1.4 Proofs of Birkhoff's Ergodic Theorem

We fix f an integrable function. We fix further the following notation:

$$\overline{A} := \limsup_{n\to\infty} A_n f, \qquad M_n f := \sup_{0\le k\le n} A_n f$$

$$\underline{A} := \liminf_{n\to\infty} A_n f, \qquad M f := \sup_n M_n f.$$

Proof 1. It is enough to assume that $0 \le f \le 1$ (why?). Let $\varepsilon > 0$. Both $\bar{A}$ and $\underline{\mathrm{A}}$ are T-invariant. We consider them to be fixed with respect to f. We define a measurable function $\tau_f : X \to \mathbb{N}$ by

$$\tau_f(x) := \min\{n \ : \ A_n f(x) > \bar{A}(x) - \varepsilon\}.$$

A little complication occurs because τ_f might not be bounded, but it is almost so. Indeed, since X has finite measure, there exists $M \gg 1$ such that the set $B := \{x \in X \ : \ \tau_f(x) > M\}$ has measure less than ε. Now for the function $g := f + \mathbb{1}_B$, τ_g is bounded by M (an exercise to verify?).

Now we pick inductively a sequence n_k in the following way: $n_0 = 0$, $n_1 = \tau_g(x)$ and $n_{k+1} - n_k = \tau_g(T^{n_k}x)$. Let

$$S_N(h;x) := N\, A_N h(x) = \sum_{i=0}^{N-1} h(T^i x).$$

For any N, let r be such that $n_r \le N - 1 < n_{r+1}$. Then

$$\begin{aligned} S_N(g;x) &= S_{n_1}(g;x) + S_{n_2-n_1}(g;T^{n_1}x) + \dots + S_{N-1-n_r}(g;T^{n_r}x) \\ &\ge (n_1 - n_0)(\bar{A}(T^{n_0}x) - \varepsilon) + \dots + (n_r - n_{r-1})(\bar{A}(T^{n_{r-1}}x) - \varepsilon) \\ &\ge (N-1-M)(\bar{A}(x) - \varepsilon), \end{aligned}$$

the last inequality following by invariance of $\overline{A}$ and the boundedness of τ_g. Remains to divide by N and integrate over X to obtain

$$\begin{aligned} \int_X g\, d\mu = \frac{1}{N}\int_X S_N(g;\cdot)\, d\mu \ge \frac{N-1-M}{N}\int_X \bar{A}\, d\mu - \varepsilon \\ \to \int_X \bar{A}\, d\mu - \varepsilon \quad \text{as } N \to \infty. \end{aligned}$$

By the definition of g, $\int_X g\, d\mu \le \int_X f\, d\mu + \varepsilon$, thus

$$\text{for any } \varepsilon > 0, \qquad \int_X f\, d\mu \ge \int_X \bar{A}\, d\mu - 2\varepsilon.$$

This being true for any $\varepsilon > 0$, and by making the same argument for $\underline{\mathrm{A}}$ or considering $1 - f$, we obtain:

$$\begin{aligned} \int_X \bar{A}\, d\mu \le \int_X f\, d\mu \le \int_X \underline{\mathrm{A}}\, d\mu \\ \implies \int_X \bar{A} - \underline{\mathrm{A}}\, d\mu = 0. \end{aligned}$$

Since $\bar{A} \ge \underline{\mathrm{A}}$, this implies both convergence in L^1 and μ-a.e. $\square$

Proof 2. Let $\varepsilon > 0$, $g = f - \mathbb{E}^{\mathcal{I}} f - \varepsilon$ and

$$G_n = \sup_{0 \le m \le n} \sum_{k=0}^{m-1} g \circ T^k.$$

For x not in the invariant set $A = \{x : G_n(x) \to \infty\}$,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{k-1} g(T^k x) \le \limsup_{n\to\infty} \frac{G_n(x)}{n} \le 0.$$

Note that $G_{n+1} - G_n \circ T = g - \min(0, G_n \circ T)$ decreases to g on A. By Dominated Convergence,

$$0 \le \int_A G_{n+1} - G_n \, d\mu = \int_A G_{n+1} - G_n \circ T \, d\mu \to \int_A g \, d\mu = \int_A \mathbb{E}^{\mathcal{I}} g \, d\mu.$$

However, $\mathbb{E}^{\mathcal{I}} g = -\varepsilon < 0$, therefore $\mu(A) = 0$. Therefore, for μ-a.e. x,

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{k-1} g(T^k x) = \overline{A}(x) - \mathbb{E}^{\mathcal{I}} f - \varepsilon \le 0.$$

Replacing f by $-f$, we obtain the reverse inequality:

$$\underline{A}(x) - \mathbb{E}^{\mathcal{I}} f + \varepsilon \ge 0,$$

which concludes the proof. □

The last proof presented here is one that generalises the best. The proof is via the important Maximal Ergodic Theorem, and in fact proves both theorems almost simultaneously. The proof is presented in (...).

If the amount of time a typical orbit spends in a set B is $\mu(B)/\mu(X)$. We might ask how likely is it to spend an unexpectedly large amount of time in B. At a fixed time n, Chebyshev's Inequality yields:

$$\mu(\{x : A_n \mathbb{1}_B(x) > \delta\} \le \frac{1}{\delta} \int_X A_n \mathbb{1}_B(x) \, d\mu = \frac{\mu(B)}{\delta}.$$

The Maximal Ergodic Theorem gives a similar statement for all n. We formulate a more general version first.

Theorem 1.4.1. *Let (X, μ) be a finite measure space, $T : X \to X$ be measure preserving and f an integrable function. Let λ be an invariant function such that its positive part λ^+ is integrable. Then the following is true: for the set $B_\lambda = \{x : Mf(x) > \lambda(x)\}$,*

$$\int_{B_\lambda} f - \lambda \, d\mu \ge 0.$$

Setting $\lambda = \delta$ a constant or $\lambda = \delta \mathbb{1}_A$ for an invariant set A, we obtain the usual formulation of the Maximal Ergodic Theorem:

Corollary 1.4.2 (Maximal Ergodic Theorem)**.** *For a constant δ and the set $B_\delta = \{x : Mf(x) \geq \delta\}$,*

$$\mu(B_\delta) \leq \frac{1}{\delta} \int_{B_\delta} f \, d\mu \leq \frac{1}{\delta} \int_X f \, d\mu.$$

Moreover, if A is invariant,

$$\mu(B_\delta \cap A) \leq \frac{1}{\delta} \int_{B_\delta \cap A} f \, d\mu \leq \frac{1}{\delta} \int_X f \, d\mu.$$

As a second corollary, setting $\lambda = \overline{A} - \varepsilon$ is most of the third proof of Birkhoff Ergodic Theorem. The only complication is that one must show that $\overline{A}$ is integrable.

Proof 3 of Birkhoff Ergodic Theorem. Consider $f^+ = \max(0, f)$ and its associated $\overline{A}f^+$. For a large bound M, $\lambda_M = \min(\overline{A}f^+, M) - 1/M \leq \overline{A}f^+$ is integrable. Then $B_{\lambda_M} = X$ and by Theorem 1.4.1,

$$\int_X f^+ \, d\mu \geq \int_X \lambda_M \, d\mu \to \int_X \overline{A}f^+ \, d\mu \quad \text{as } M \to \infty.$$

Therefore, $\overline{A}^+ \leq \overline{A}f^+$ must be integrable. Similarly, $\overline{A}^-$ is also integrable.

We apply Theorem 1.4.1 to $\lambda = \overline{A} - \varepsilon$ to conclude that

$$\int_X f \, d\mu \geq \int_X \overline{A} \, d\mu - \varepsilon.$$

As in the first proof, considering $-f$ instead finishes the proof. □

Proof of Theorem 1.4.1. We can assume that λ to be integrable, otherwise $\int_{B_\lambda} f - \lambda \, d\mu = \infty > 0$. Next, we can assume f is essentially bounded: if not, apply the theorem to the truncation $f_s = f \mathbb{1}_{\{x : |f(x)| \leq s\}}$ and let $s \to \infty$ using the Dominated Convergence Theorem.

Define $E_n := \{x : M_n f(x) \geq \lambda(x)\}$. If $x \notin E_n$, then $f(x) \leq M_n f(x) \leq \lambda(x)$, therefore $(f - \lambda)\mathbb{1}_{E_n} \geq (f - \lambda)$. Consider a long segment of the orbit of length $L \gg n$. By definition of M_n, if at some time $T^{k_0}x \in E_n$, then there exists $m \leq n$ such that $A_m f(T^{k_0}x) \geq \lambda(T^{k_0}x) = \lambda(x)$. Therefore,

$$\sum_{k=0}^{m-1} f(T^{k+k_0}x) - \lambda(T^{k+k_0}x) \geq 0.$$

In other words, each time the orbit enters E_n, there is a segment of length most n starting at that point on which the sum of $(f-\lambda)\mathbb{1}_{E_n}$ is positive. This can be done on all the segment, except perhaps at the end. Hence, there exists a j such that $j \leq L-1 \leq j+n$, and such that

$$\begin{aligned}\sum_{k=0}^{L-1}(f(T^k x)-\lambda(T^k x))\mathbb{1}_{E_n}(T^k x) &\geq \sum_{k=j}^{L-1}(f(T^k x)-\lambda(T^k x))\mathbb{1}_{E_n}(T^k x)\\ &\geq -n(\|f\|_\infty+\lambda^+(x)).\end{aligned}$$

Integrating, we obtain that

$$\int_{E_n} f-\lambda\, d\mu \geq \frac{-n}{L}(\|f\|_\infty+\|\lambda^+\|_1)\to 0 \quad \text{as } L\to\infty.$$

To conclude the proof, we apply Dominated Convergence as $n\to\infty$, and note that $E_n \to B_\lambda$. □

1.5 Ergodic Measures

Among the non-singular dynamical systems, ergodicity is synonymous to indecomposable: one cannot restrict non-trivially the system to an invariant subset. Moreover, any invariant measure can be decomposed into ergodic components.

Definition 1.5.1. A non-singular system (X,G,μ) is called *ergodic* if for every invariant set A, either $\mu(A)=0$ or $\mu(X\setminus A)=0$. It is called *properly ergodic* if there is no μ-conull orbit.

There is no possible confusion in the meaning of an invariant set due to the following theorem:

Theorem 1.5.2. *For a nonsingular action of a locally compact group G on a standard Borel space with measure μ, a set is μ-almost invariant under the full group if and only if it is μ-almost invariant under each element of the group.*

Depending on the emphasis, we can say that μ is G-ergodic (or ergodic if G is implied), or that the G-action is μ-ergodic (or ergodic if μ is implied). We give several different characterisation of ergodicity.

Definition 1.5.3. A space Y with a σ-algebra $\mathcal{B}$ is countably separated if there exists a countable collection of sets $B_1, B_2, ...$ in $\mathcal{B}$ such that for any two points $x,y\in Y$, there exists a n such that B_n contains exactly one of x or y.

Proposition 1.5.4. *For a G-space $(X, \mathcal{B}_X)$, μ being ergodic is equivalent to every G-invariant function $f : (X, \mathcal{B}_X) \to (Y, \mathcal{B}_Y)$ into a countably separated space being essentially constant.*

Proof. Suppose μ is ergodic. For any $B \in \mathcal{B}_Y$, $f^{-1}(B) \in \mathcal{B}_X$ is invariant. Let $\{B_1, B_2, ...\}$ be a countable separating collection of sets in $\mathcal{B}_Y$. Assume that the collection is closed under complements. Consider the set

$$B := \bigcap_{\mu(f^{-1}B_n)=0} Y \setminus B_n.$$

The set B is non-empty for otherwise μ would be the zero measure. Consider two points and a set B_n separating them. By ergodicity, either $\mu(f^{-1}B_n) = 0$ or $\mu(f^{-1}(Y \setminus B_n)) = 0$. Then B is contained in one of B_n or its complement, and therefore misses at least one the two points. We conclude that B is a single point. Since $\mu(X \setminus f^{-1}B) = 0$, this shows the first direction.

For the other direction, the set of real numbers is countably separated. For an invariant set $A \in \mathcal{B}_X$, $\mathbb{1}_A$ is an invariant measurable function, therefore essentially constant. Therefore either A or $X \setminus A$ has measure zero. □

The proposition can be phrased in the context of unitary operators as follows: for a finite measure space $(X, \mathcal{B}, \mu)$ with a measure-preserving G-action, we have a unitary representation $\pi : G \to \mathcal{U}(L^2(X))$, where $\pi(g)f = f \circ g$. Then the proposition states that the trivial representation is contained in π only once and is given as its restriction to constant functions.

Proposition 1.5.5. *In the case when Birkhoff's Ergodic Theorem holds, ergodicity of μ is equivalent to the assertion that for every integrable function f, for μ-a.e. x,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) = \frac{1}{\mu(X)} \int_X f \, d\mu.$$

Proof. The function $\mathbb{E}^{\mathcal{I}} f$ being invariant, ergodicity implies it is essentially constant. By properties of conditional expectations, it integrates to $\int_X f \, d\mu$, which completes the first direction.

Now consider an invariant set A and apply Birkhoff's Ergodic Theorem to $\mathbb{1}_A$: for μ-a.e. x,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{1}_A(T^k x) = \frac{\mu(A)}{\mu(X)}.$$

The left hand side being 0 or 1, we conclude that $\mu(A)$ is either zero or $\mu(X)$. $\square$

For ergodic measures, Birkhoff Ergodic Theorem can therefore be stated as typical orbits satisfy that time average equals space average.

A straightforward consequence is that two different ergodic measures must be singular.

Proposition 1.5.6. *For a G-space X, any two nonproportional invariant ergodic measures μ and ν are singular.*

Proof. By ergodicity, either the two measures have same support, or they have disjoint support. That is, either they are absolutely continuous with respect to each other, or they are singular. In the former case, the Radon-Nikodym derivative is an invariant function, hence constant almost everywhere. $\square$

From a functional analysis point of view, the Banach-Alaoglu-Bourbaki theorem implies that the space of invariant probability measures on X with its weak-*-topology is compact and convex. Ergodic measures are characterised as extreme points:

Proposition 1.5.7. *A probability measure μ is ergodic if and only if it is an extreme point in the space of invariant probability measures: if one has $\mu = \lambda\nu_1 + (1-\lambda)\nu_2$ for some $0 \leq \lambda \leq 1$ and for two different invariant probability measures ν_1 and ν_2, then λ must equal 0 or 1.*

Proof. Suppose μ is invariant, but not ergodic. Then there exists an invariant set A such that both A and $X \setminus A$ have positive measure. Let $\nu_1 = \mu_{|A}$ and $\nu_2 = \mu_{|X\setminus A}$, then

$$\mu = \mu(A)\nu_1 + (1-\mu(A))\nu_2.$$

Thus μ is not an extreme point.

Suppose μ is invariant and ergodic. Let $\mu = \lambda\nu_1 + (1-\lambda)\nu_2$ for some $\lambda > 0$. Then ν_1 is absolutely continuous with respect to μ. Let f be the Radon-Nikodym derivative $\frac{D\nu_1}{D\mu}$. The invariance of the measures implies invariance of f, and by ergodicity, it is constant μ-a.e. Since the measures have mass one, $f = 1$ μ-a.e. So $\nu_1 = \mu$, and therefore $\nu_2 = \mu$ as well. $\square$

One can then invoke Choquet's Theorem to obtain an ergodic decomposition: every invariant probability measure is an integral of ergodic invariant measures. For finite measure Borel spaces however, we have the following somewhat more explicit version:

Theorem 1.5.8 (Ergodic Decomposition). *Let $(X, \mathcal{B}, \mu)$ be a Borel G-space with a finite invariant measure μ. Let $\mathcal{I}$ be the σ-algebra of invariant sets in $\mathcal{B}$. Then $\mu_x^{\mathcal{I}}$ is ergodic for μ-a.e. x and*

$$\mu = \int_X \mu_x^{\mathcal{I}} \, d\mu.$$

Proof. Let $\mathcal{A}$ be a countably-generated σ-algebra μ-equivalent to $\mathcal{I}$, which exists by Lemma 1.1.10. By proposition 1.1.11, $\mu_x^{\mathcal{A}} = \mu_x^{\mathcal{I}}$ for μ-a.e. x are probability measures supported on the atoms $[x]_{\mathcal{A}}$. Since the acting group is countable, we can enlarge the implied null set N so that outside of it $T_* \mu_x^{\mathcal{A}} = T_* \mu_x^{T^{-1}\mathcal{A}} = \mu_{Tx}^{\mathcal{A}}$. Enlarging N again, we can make it T-invariant. Thus for $x \notin N$, $\mu_x^{\mathcal{A}} = \mu_{Tx}^{\mathcal{A}}$ is T-invariant.

To proceed with ergodicity, we can enlarge N to make sure that Birkhoff Ergodic Theorem holds for $x \notin N$ for a dense countable set $\{f_1, f_2, ...\}$ of continuous functions on $\overline{X}$. Let $N' = N \cup \{x \in X \, : \mu_x^{\mathcal{A}}(N) > 0\}$, which has measure zero by Lemma 1.1.8. For $x \notin N'$, for any $y \in [x]_{\mathcal{A}} \setminus N$, $\mu_x^{\mathcal{A}} = \mu_y^{\mathcal{A}}$, and thus

$$\frac{1}{n} \sum_{k=0}^{n-1} f_i(T^k y) \to \int_X f_i \, d\mu_y^{\mathcal{A}} = \int_X f_i \, d\mu_x^{\mathcal{A}},$$

holds for $\mu_x^{\mathcal{A}}$-a.e. y. Therefore, the operator $\mathbb{E}^{\mathcal{I}}$ corresponding to $\mu_x^{\mathcal{A}}$ is projection to constant functions on a dense set of functions in L^2, thus on all L^2 functions. This is equivalent to ergodicity. $\square$

In most cases, the space of invariant measures is infinite-dimensional. However, in the special case that the system admits a unique invariant measure (automatically ergodic), we have the following stronger result:

Proposition 1.5.9. *Let $(X, \mathcal{B}, \mu)$ be a finite measure Borel space, and let $T : X \to X$ be a measure-preserving transformation. Suppose μ is the unique invariant measure. Then for any continuous function f, time averages converge uniformly for every point $x \in X$.*

Proof. Suppose that for some continuous function f, convergence is not uniform. Then there exist numbers $a < b$, sequences x_k and y_k in X, and a sequence n_k such that

$$A_{n_k}(f; x_k) < a < b < A_{n_k}(f; y_k).$$

Taking a converging subsequence, we define

$$I(f) := \lim_{k\to\infty} A_{n_k}(f; x_k), \quad J(f) := \lim_{k\to\infty} A_{n_k}(f; y_k).$$

By a diagonal argument as in Proposition 1.1.7, I and J can be defined for any continuous function f. Being limits of time averages, they are invariant functionals. By the Riesz Representation Theorem, they are represented by invariant measures. Since they must be different, this contradicts unique ergodicity. $\square$

1.6 Some examples

Example 1.6.1. Let $Y = \{1, ..., m\}$ be a finite set, j having probability p_j. Let $X = Y^{\mathbb{Z}}$ with its product measure μ and let $T : X \to X$ be the Bernoulli shift, that is the transformation defined coordinate-wise by $T(x_k) = x_{k+1}$. Kolmogorov 0-1 law implies ergodicity of the system. The strong law of large numbers is then a consequence of Birkhoff's ergodic theorem for the set A of elements whose 0^{th}-coordinate is j:

$$\text{frequency of occurence of } j \text{ in } [0, n] = \frac{1}{n+1} \sum_{k=0}^{n} \mathbb{1}_A(T^k x) \to$$

$$\to \int_X \mathbb{1}_A \, d\mu = p_j \quad w.p.1 \text{ and as } n \to \infty.$$

Example 1.6.2. Identify the circle S^1 with $\mathbb{R}/\mathbb{Z}$ and let $T : S^1 \to S^1$ be the expanding map $T(x) = 10x$. It preserves the Lebesgue measure. If x is rational, the periodicity of the decimal expansion means that x has a periodic orbit. If x is irrational however, we wish to say that it is uniformly distributed. Let μ be the Lebesgue measure and let A be an invariant set of measure less than one. Pick a small interval I of length 10^{-n} around a point of no density of A, so that $\mu(I \setminus A) > (1-\varepsilon)\mu(I)$. Since A is invariant, $T(S^1 \setminus A) = S^1 \setminus A$. The map T expands the interval I by a factor 10 as long as it is injective. Therefore $\mu(T^n(I) \setminus A) > 1 - \varepsilon$, and the system is ergodic.

Therefore, for almost every irrational number, the frequency of occurance of any block of k digits is asymptotically 10^{-k}. The decimals of an irrational number are thus random, and the previous example with uniform probability $1/10$ on the digits is a good model for picking an arbitrary number.

Example 1.6.3. Identify the torus $\mathbb{T}^n$ with $\mathbb{R}^n/\mathbb{Z}^n$ and let $T : \mathbb{T}^n \to \mathbb{T}^n$ be defined as $T(x_1, ..., x_n) = (x_1 + \alpha_1, ..., x_n + \alpha_n)$. It preserves the Lebesgue measure. If $\{1, \alpha_1, ..., \alpha_n\}$ is linearly independent over $\mathbb{Q}$, then the system is uniquely ergodic. First we prove ergodicity using harmonic analysis. Let f be an invariant function and let $f(x) = \sum_{\gamma \in \mathbb{Z}^n} c_\gamma exp(2\pi i \gamma . x)$ be its Fourier expansion. Invariance of f means that $c_\gamma = c_\gamma exp(k.x)$. By the linear independence condition, this implies that $c_\gamma = 0$ unless $\gamma = 0$. That is f is essentially constant, proving ergodicity. Thus T has at least one dense orbit (see Proposition 1.7.1). But $\mathbb{T}^n$ being abelian, any orbit is a translate of that dense orbit. Therefore all orbits are dense, and the system is uniquely ergodic.

The case $n = 1$ says that an irrational rotation of the circle is uniquely ergodic and has the following interesting consequence: we can determine the limiting distribution of the first digit of the sequence $\{2^n\}$. Indeed a number x has first digit k if $\log_{10}(k) \leq \{\log_{10}(x)\} < \log_{10}(k+1)$ ($\{y\}$ is the fractional part of y). Since $\log_{10}(2)$ is irrational, we obtain that the frequency of occurance of the digit k is

$$\frac{1}{n+1}\sum_{k=0}^{n} \mathbb{1}_{[\log_{10}(k),\log_{10}(k+1))}(n\log_{10}(2)) \to \log_{10}(\frac{k+1}{k}).$$

Example 1.6.4. Irrational rotation on the circle has the following vast generalisation: Weyl's Polynomial Equidistribution Theorem. It states that for a polynomial $p(x) = a_k x^k + ... + a_0$ with at least one irrational coefficient other than a_0, the sequence $\{p(n) \mod 1 \,:\, n \in \mathbb{N}\}$ gets equidistributed with respect to the Lebesgue measure. Furstenberg gives an ergodic proof of the theorem.

1.7 Ergodicity and Orbit Structures

Typical orbits with respect to an invariant measure were seen to satisfy statistical recurrence. Ergodicity can be viewed as a statistical counterpart to topological transitivity:

Proposition 1.7.1. *Let G act continuously on a complete separable metrisable space X. Suppose Birkhoff Ergodic Theorem applies. Then a measure μ being ergodic implies that μ-a.e. point has an orbit dense in $supp(\mu)$.*

Proof. Since $supp(\mu)$ is closed, we can assume $supp(\mu) = X$. Take a countable basis $U_1, U_2, ...$ of open sets in X. By definition of the support of a measure, $\mu(U_k) > 0$. Applying Birkhoff Ergodic Theorem to each of $\mathbb{1}_{U_k}$, we see that there is a set of full measure of points $x \in X$ having their orbit $G.x$ visiting every U_k, and therefore $G.x$ is dense. $\square$

An essentially transitive system is one having a conull orbit. As far as measure theory is concerned, such systems are trivial. On the other hand, proper ergodicity reflects a complicated orbit structure. An instance of this is the following theorem. Recall that a set is locally closed if it is the intersection of an open set with a closed set, or equivalenty, if it is open in its closure.

Theorem 1.7.2. *Let G act continuously on a complete separable metrisable space X. Then the following statements are equivalent.*

i) Every orbit is locally closed.

ii) The orbit space X/G is T_0 in the quotient topology.

iii) The orbit space X/G is countably separated in the quotient Borel structure.

iv) Every quasi-invariant ergodic Borel measure μ on X is supported by an orbit.

Proof. Suppose every orbit is locally closed. For two points $x, y \in X$, either Gy is contained in the closure $\overline{Gx}$ of Gx, or they are disjoint. In the latter case, the complement of $\overline{Gx}$ is an open set separating the two orbits. In the former case, let U be an open set such that $Gx = U \cap \overline{Gx}$. Then U separates the two orbits, assuming they are different. So (i) implies (ii).

Take a countable basis for the topology $U_1, U_2, ...$ Suppose they separate orbits. Then $GU_1, GU_2, ...$ are Borel invariant sets separating orbits. So (ii) implies (iii).

For (iii) implies (iv), the argument is almost identical to the one in Proposition 1.5.4.

The most difficult part is showing (iv) implies (i). It is a theorem of Glimm. See Zimmer for a proof of this fact. □

The next proposition leads to yet another characterisation of the above, but is important in its own right.

Proposition 1.7.3. *Let G act continuously on a complete separable metrisable space X. The orbit Gx through x is locally closed if and only if the map $\phi : G/Stab_G(x) \to Gx$ is a homeomorphism, where Gx has the relative topology as a subset of X.*

Proof. By passing to the closure of $G.x$, we can assume that it is dense in X. The map ϕ being a continuous bijection, it is a homeomorphism if and only if it is an open map.

Suppose first that Gx is open in X. Let U be any neighbourhood of e in G, and choose a compact neighbourhood K of e such that $K = K^{-1}$ and $K^2 \subset U$. For a countable dense set $\{g_k\}$ in G, $Gx = \bigcup g_k Kx$. Since Gx

is an open dense set of a Baire space, it is itself a Baire space, and thus at least one of $g_k Kx$ contains an open set. By homogeneity, they all contain an open set. There exists a $k \in K$ for which Kx is a neighbourhood of kx. Then $k^{-1}Kx$ is a neighbourhood of x. In particular $Ux \supset K^2x$ is a neighbourhood of x, and ϕ is thus an open map.

Conversely, suppose ϕ is a homeomorphism. Then Gx is a Baire space, and since G is σ-compact, there exists some compact set $K \subset Gx$ containing an open set. That is, there exists an open set $U \subset X$ such that $K \supset U \cap Gx$. Then $Gx = GK \supset GU \cap Gx$. That is $Gx = GU$ is open. □

Certain natural classes of actions satisfy the conditions of Theorem 1.7.2, for instance, algebraic groups acting on varieties or actions of compact groups. As a corollary, these do not admit an interesting ergodic theory. For compact groups, one has the following stronger statement:

Corollary 1.7.4. *Any action of a compact group G on a countably separated Borel space X has a countably separated orbit space X/G.*

Proof. A compact group G acting continuously on a topological space X has compact orbits. It is a theorem of Varadarajan that for any countably separated Borel G-space X, there exist a compact metric space Y with a continuous G-action and a G-equivariant injective map $X \hookrightarrow Y$. The statement then follows from Theorem 1.7.2. □

1.8 Mixing

If one puts a proportion $\mu(A)$ of salt into water and waits for some time, then the mixture becomes homogeneous as the salt mixes within the water. Then the amount of salt found in a portion of size $\mu(B)$ of the mixture is $\mu(A)\mu(B)$. A dynamical system satisfying this property is said to be mixing:

Definition 1.8.1. A G-space X with a finite quasi-invariant measure μ is said to be (strongly) mixing if for any subsets $A, B \subset X$,

$$\lim_{g \to \infty} \mu(g^{-1}A \cap B) = \frac{\mu(A)\mu(B)}{\mu(X)}.$$

Statistically speaking, a system is mixing if any two sets become independent in the long-run. It relates to ergodicity via the following proposition:

Proposition 1.8.2. *For a transformation $T : X \to X$, a finite invariant measure μ is ergodic if and only if it is mixing on average: for any set A and B,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mu(T^{-k}A \cap B) = \frac{\mu(A)\mu(B)}{\mu(X)}.$$

Proof. Suppose that μ is ergodic and A, B two subsets in X. We apply the Mean Ergodic Theorem to the function $\mathbb{1}_A$, and taking inner product with $\mathbb{1}_B$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} < \mathbb{1}_A \circ T^k, \mathbb{1}_B > = < \frac{\mu(A)}{\mu(X)}, \mathbb{1}_B > = \frac{\mu(A)\mu(B)}{\mu(X)}.$$

So μ is mixing on average.

Suppose μ is mixing on average and let A be invariant. We apply mixing on average to $B = X \setminus A$ to get $\mu(A)\mu(X \setminus A) = 0$. Hence it is ergodic. $\square$

The second part of the proof in fact shows that mixing on average implies ergodicity, and this for a general group action. In particular, a mixing system is automatically ergodic.

Mixing is a rather special property. The following slightly weaker version seems to happen considerably more often.

Definition 1.8.3. For a system (X, T, μ) of a finite measure space with a measure preserving transformation $T : X \to X$ is said to be weakly-mixing if for any subsets $A, B \subset X$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \big|\mu(T^{-k}A \cap B) - \frac{\mu(A)\mu(B)}{\mu(X)}\big| = 0.$$

Mixing implies weak-mixing, which in turn implies ergodicity. However, since in the definition of weak-mixing, the summand can be large for a set of density zero of values of k, weak-mixing does not imply mixing.

Weak-mixing has many other characterisations that seem perhaps more natural. The following one leads to immediate generalisation.

Proposition 1.8.4. *The system (X, T, μ) is weakly-mixing if and only if for any ergodic system (Y, S, ν), the product system $(X \times Y, T \times S, \mu \times \nu)$ is ergodic.*

1.9 Entropy

Entropy is our first invariant of growth-type. It measures the exponential growth rate of statistically significant orbit segments distinguishable by finite partitions. It is based on the notion of the information function.

Suppose we have an event with n possible outcomes with known probabilities $p_1, ..., p_n$. How much information is gained by observing the event? In other words, how uncertain are we of the outcome? If we have a function $H(p_1, ..., p_n)$ measuring information, then we would like it to satisfy the following three intuitive properties:

i) *Continuity*: H is continuous in the variables p_i.

ii) *Monotonicity*: $H(\frac{1}{n}, ..., \frac{1}{n})$ is monotone increasing with respect to n: that is uncertainty increases with the number of possible equally probable outcomes.

iii) *Divisibility*: If an event A should be broken into two events B and C, then $H(A)$ should be the weighted sum of $H(A)$ and $H(B)$. That is if $A = \{p_1, ..., p_n\}$, $B = \{p_1 + ... + p_k, p_{k+1}, ..., p_n\}$ and $C = \{q_1, ..., q_k\}$ where $(p_1 + ... + p_k)q_m = p_m$, then

$$H(p_1, ..., p_n) = H(p_1 + ... + p_k, p_{k+1}, ..., p_n) + (p_1 + ... + p_k)H(q_1, ..., q_k).$$

Proposition 1.9.1. *The only function satisfying the above three properties up to a multiplicative constant is*

$$H(p_1, ..., p_n) = \sum_{k=1}^{n} p_k \log \frac{1}{p_k} = -\sum_{k=1}^{n} p_k \log p_k.$$

(Here we must agree that we do not consider summands where $p_k = 0$, or that $0 \log 0 = 0$.)

The function H is called the *entropy* of the partition. That it is the right function intuitively is reinforced by a list of further properties that we omit as they will be encountered in various forms later.

Suppose that μ is a probability measure on X. Given a finite partition $\mathcal{P} = \{P_1, ...P_n\}$ into measurable sets, we have $H_\mu(\mathcal{P}) = H(\mu(P_1), ..., \mu(P_n))$. The information function $I_\mathcal{P}(x)$ of the partition $\mathcal{P}$ is $-\log \mu(P_k)$, where $x \in P_k$. This is about the number of bits required to encode the k^{th} outcome in an optimal encoding, or in other words the amount of information carried by the outcome. From the definitions, we see that $H(\mathcal{P}) = \int_X I_\mathcal{P}(x)\, d\mu$, hence entropy is the expected number of bits necessary to record the event, or expected information gained by observing the event.

Under the assumption that we disregard sets of measure zero, given two finite partitions $\mathcal{P}$ and $\mathcal{Q}$, we can define a conditional information function in a natural way: $I_{\mathcal{P}|\mathcal{Q}}(x) = -\log \mu_x^{\mathcal{Q}}(\mathcal{P}(x))$, where $\mathcal{P}(x)$ is the partition element of $\mathcal{P}$ containing x. Then conditional entropy is given by $H(\mathcal{P}|\mathcal{Q}) = \int_X I_{\mathcal{P}|\mathcal{Q}}\, d\mu$.

Consider the space $\mathscr{P}$ of finite measurable partitions of finite entropy up to measure zero. It is partially ordered according to $\mathcal{P} \leq \mathcal{Q}$ if for all $Q \in \mathcal{Q}$, there exists $P \in \mathcal{P}$ such that $\mu(P \triangle Q) = 0$. In this case we call $\mathcal{P}$ a refinement of $\mathcal{Q}$. The next proposition will in particular show that the space $\mathscr{P}$ admits joints: $\mathcal{P} \vee \mathcal{Q} = \{P \cap Q : P \in \mathcal{P}, Q \in \mathcal{Q}\}$. Conditional entropy can then be written as a weighted sum of entropies:

$$H(\mathcal{P}|\mathcal{Q}) = \sum_{Q \in \mathcal{Q}} \mu(Q) H_{\mu_{|Q}}(\mathcal{P} \vee \mathcal{Q}) = -\sum_{Q \in \mathcal{Q}} \mu(Q) \sum_{P \in \mathcal{P}} \mu_{|Q}(P) \log \mu_{|Q}(P).$$

Two partitions $\mathcal{P}$ and $\mathcal{Q}$ are called *independent* if $\mu(P \cap Q) = \mu(P)\mu(Q)$ for any $P \in \mathcal{P}$ and for any $Q \in \mathcal{Q}$. Note then that $\mu_{|Q}(P) = \mu(P)$.

Properties of entropy and conditional entropy are recorded in the following proposition:

Proposition 1.9.2. *Let $\mathcal{P}, \mathcal{Q}$ and $\mathcal{R}$ be three measurable finite partitions of a probability space $(X, \mathcal{B}, \mu)$. Then*

i) $0 < -\log \sup_{P \in \mathcal{P}} \mu(P) \leq H(\mathcal{P}) \leq \log \#\mathcal{P}$, *the maximal entropy $\log \#\mathcal{P}$ being achieved if and only if all elements of $\mathcal{P}$ have equal measure.*

ii) $0 \leq H(\mathcal{P}|\mathcal{Q}) \leq H(\mathcal{P})$: *the first equality if and only if $\mathcal{Q} \leq \mathcal{P}$, the second equality if and only if the partitions are independent.*

iii) If $\mathcal{Q} \leq \mathcal{R}$, then $H(\mathcal{P}|\mathcal{Q}) \leq H(\mathcal{P}|\mathcal{R})$.

iv) $H(\mathcal{P} \vee \mathcal{Q}|\mathcal{R}) = H(\mathcal{P}|\mathcal{R}) + H(\mathcal{Q}|\mathcal{P} \vee \mathcal{R})$. *In the particular case that $\mathcal{R}$ is trivial, we obtain $H(\mathcal{P} \vee \mathcal{Q}) = H(\mathcal{P}) + H(\mathcal{Q}|\mathcal{P})$.*

v) $H(\mathcal{P} \vee \mathcal{Q}|\mathcal{R}) \leq H(\mathcal{P}|\mathcal{R}) + H(\mathcal{Q}|\mathcal{R})$. *In the particular case that $\mathcal{R}$ is trivial, we obtain $H(\mathcal{P} \vee \mathcal{Q}) \leq H(\mathcal{P}) + H(\mathcal{Q})$.*

vi) $H(\mathcal{P}|\mathcal{Q}) + H(\mathcal{Q}|\mathcal{R}) \geq H(\mathcal{P}|\mathcal{R})$.

vii) $H_\mu(\mathcal{P})$ *is concave as a function of μ: If μ and ν are probability measures on X, and $\lambda \in [0, 1]$, then*

$$\lambda H_\mu(\mathcal{P}) + (1 - \lambda) H_\nu(\mathcal{P}) \leq H_{\lambda\mu + (1-\lambda)\nu}(\mathcal{P}).$$

viii) For a nonsingular transformation $T : X \to X$, $H_\mu(T^{-1}\mathcal{P}) = H_{T_*\mu}(\mathcal{P})$.

Corollary 1.9.3. *The space $\mathscr{P}$ can be equipped with the* Rokhlin metric

$$d(\mathcal{P}, \mathcal{Q}) = H(\mathcal{P}|\mathcal{Q}) + H(\mathcal{Q}|\mathcal{P}).$$

We are now ready to define entropy of a measure preserving transformation. Let $(X, \mathcal{B}, \mu)$ be a finite measure Borel space and let $T : X \to X$ be measure-preserving. Define iterated partitions by

$$\mathcal{P}^T_{-n} := \bigvee_{k=0}^{n-1} T^{-k}\mathcal{P}.$$

Lemma 1.9.4. *For a partition $\mathcal{P}$ of finite entropy, the following limit exists:*

$$h_\mu(T, \mathcal{P}) := \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P}) = \lim_{n\to\infty} H(\mathcal{P}| \bigvee_{k=1}^{n-1} T^{-k}\mathcal{P}).$$

Proof. By inductively applying property iv) of entropy and property viii),

$$a_n := H(\bigvee_{k=0}^{n-1} T^{-k}\mathcal{P}) = \sum_{m=1}^{n} H(\mathcal{P}| \bigvee_{k=0}^{m-1} T^{-k}\mathcal{P}) =: \sum_{m=1}^{n} b_m.$$

Property ii) and iii) show that b_m is a nonincreasing sequence bounded between zero and $H(\mathcal{P})$. The lemma follows. □

The quantity $h_\mu(T, \mathcal{P})$ measures the exponential rate of decay of the average size of an element of iterated partitions. For ergodic transformations, this size is almost always close to the average:

Theorem 1.9.5 (Shannon-McMillan-Breiman Theorem)**.** *Let $(X, \mathcal{B}, \mu)$ be a finite measure Borel space, $T : X \to X$ be measure-preserving and ergodic, and let $\mathcal{P}$ be a partition of finite entropy. Then for μ-a.e. x,*

$$\lim_{n\to\infty} -\frac{1}{n} \log \mu(\mathcal{P}^T_{-n}(x)) = h_\mu(T, \mathcal{P}).$$

Relative entropy $h_\mu(T, \mathcal{P})$ satisfies a list of properties similar to the one for the entropy of a partition. One that has no counterpart and is worth mentioning is that for any $m \in \mathbb{N}$,

$$h_\mu(T, \mathcal{P}) = h_\mu(T, \bigvee_{k=0}^{m} T^{-k}\mathcal{P}) = h_\mu(T, \bigvee_{k=-m}^{m} T^{-k}\mathcal{P}).$$

Definition 1.9.6. The *measure-theoretic entropy* of T with respect to μ is

$$h_\mu(T) := \sup_{\mathcal{P} \in \mathscr{P}} h_\mu(T, \mathcal{P}).$$

The definition of entropy $h_\mu(T)$ would not change if we were to allow countable partitions as well since they can be approximated by finite ones.

From an information theoretic point of view, entropy of a transformation T is the expected information obtained at the present state knowing the past. For example, for a zero entropy transformation, one can know approximately the future of an orbit from the knowledge of its past.

As before, we list some properties of entropy.

Proposition 1.9.7 (Properties of Entropy)**.** *Let $(X, \mathcal{B}, \mu)$ be a finite measure Borel space and $T : X \to X$ be measure-preserving. Then*

i) Entropy respects the ergodic decomposition:

$$h_\mu(T) = \int_X h_{\mu_x^{\mathcal{I}}}(T)\, d\mu.$$

ii) For any $m \in \mathbb{Z}$, $h_\mu(T^m) = |m| h_\mu(T)$.

iii) If T is ergodic, A is a set of positive measure, and $T_{|A}$ is the induced transformation on A, then $h_{\mu_{|A}}(T_{|A}) = \frac{1}{\mu(A)} h_\mu(T)$.

We finish this section with some remarks about computations. The supremum in the definition of entropy is in many cases achieved by a partition. A partition $\mathcal{P}$ is said to be generating if the σ-algebra $\mathcal{P}_{-\infty}^T \vee \mathcal{P}_\infty^T$ separates points in the sense that μ-a.e. x has atom $\{x\}$. Generating partitions always achieve the supremum.

We give two examples of computations of entropy.

Example 1.9.8. Let $T : S^1 \to S^1$ be a rotation of the circle, μ the Lebesgue measure and $\mathcal{P}$ any partition into p intervals. The number of partition elements of $\mathcal{P}_{-n}^T$ is bounded by np. By Proposition 1.9.2 i), we obtained $h_\mu(T, \mathcal{P}) = 0$ for any such partition. But that family of partitions being sufficient, $h_\mu(T) = 0$.

Example 1.9.9. Let $T : \mathbb{T}^2 \to \mathbb{T}^2$ be a hyperbolic transformation, that is given by a real diagonalisable integral linear transformation with eigenvalues $\lambda > 1$ and $\lambda^{-1} < 1$. Let μ be the Lebesgue measure. Then we show that $h_\mu(T) = \log(\lambda)$.

1.10 Ergodic Theory for Amenable Groups

We have tried to make definitions as general as possible. For example, the definition of ergodicity is valid for any nonsingular dynamical system. However, the most salient features of ergodic theory requires to average over larger and larger parts of an orbit. This process should lead to invariance in the limit. For the case of weak-mixing, Proposition 1.8.4 allows one to avoid sampling the orbit. But for pointwise and mean theorems, such a reformulation doesn't exist.

Over $\mathbb{Z}$, the way classically one samples the orbit is using the sequence of sets $F_n = \{0, 1, ..., n-1\}$. The essential feature is that if we translate by a bounded amount, for large n the sets F_n don't move that much relative to themselves. So averaging over F_n or gF_n doesn't change anything in the limit.

It is perhaps at first surprising, but not all groups have a sequence of sets F_n as described. Those that do are termed amenable:

Definition 1.10.1. A group G with invariant measure μ is *amenable* if for every compact set $K \subset G$ and $\varepsilon > 0$, there exists a positive finite measure (K, ε)-invariant set F in the sense that for any $g \in K$, $\mu(F \triangle gF) < \varepsilon\mu(F)$.

A sequence F_n of sets that are eventually (K, ε)-invariant for every K and ε is called a *Folner sequence.*

There are many other possible definitions for amenability, but the one described is the most useful for our purposes. Briefly, the following groups are amenable: compact and abelian groups, closed subgroups of an amenable group, and amenable-by-amenable extensions, and thus solvable groups. For connected Lie groups, they are exactly the compact extensions of solvable groups.

We can sample orbits along any Folner sequence F_n. Let

$$A_{F_n} f(x) := \frac{1}{\mu(F_n)} \int_{F_n} f(gx)\, d\mu(g).$$

The Mean Ergodic Theorem is proven in almost the same way as for the classical version. However, pointwise theorems pose more difficulty. Even for $\mathbb{Z}$, one can exhibit functions for which $A_{F_n} f(x)$ does not converge, where $F_n = \{n^2, ..., n^2 + n\}$ is seen to be Folner. Nonetheless, Lindenstrauss established convergence for tempered Folner sequences. Any Folner sequence has a tempered Folner subsequence, thus one has a completely satisfying version of Birkhoff Ergodic Theorem for amenable groups.

Definition 1.10.2. A Folner sequence F_n is *tempered* if it satisfies the following growth condition: there exists a constant C such that for any n,

$$\mu(\bigcup_{k<n} F_k^{-1}F_n) < C\mu(F_n).$$

Theorem 1.10.3. *Let G be an amenable group with invariant measure μ acting by measure preserving transformations on a finite measure space (X,ν), let $\mathcal{I}$ denote the σ-algebra of invariant sets, and let F_n be a tempered Folner sequence. Then for any integrable function f and μ-a.e. x,*

$$\lim_{n\to\infty} \frac{1}{\mu(F_n)} \int_{F_n} f(gx)\,d\mu(g) = \mathbb{E}^{\mathcal{I}} f(x).$$

Sketch of a proof. To give an idea of the proof, we assume $F_n \subset F_{n+1}$. As before, it is enough to show that for any positive integrable function f,

$$\overline{A}f(x) := \limsup_{n\to\infty} A_{F_n} f(x) \leq \int_X f\,d\nu =: \nu(f).$$

Let $c > 1$, and $\varepsilon > 0$. Suppose $B = \{x \in X : \overline{A}f(x) > c\,\nu(f)\}$ has positive measure, and let $K \subset G$ be a compact set such that $\nu(KB) > 1-\varepsilon$. Let $B_n = \{x \in X : A_{F_n} f(x) > c\,\nu(f)\}$. From the definition of limsup, for any large M, one can find large N_i and a_i so that

$$A = \bigcap_{i=1}^{M} \bigcup_{j=0}^{N_i-1} B_{a_i+j}$$

satisfies $\nu(KA) > 1 - 2\varepsilon$, and so that a_{i+1} is so much bigger than a_i that the Folner sets are very tempered: $\mu(KF_{a_i+N_i}^{-1}F_{a_{i+1}}) < (1+\delta)\mu(F_{a_{i+1}})$ for a small δ.

For an extremely larger set F, for a large set of x, $A_F \mathbb{1}_{KA}(x) > 1 - 3\varepsilon$. If F is also invariant enough, the Mean Ergodic Theorem implies that for a large set of x, $A_F f(x) < (1+\varepsilon)\nu(f)$. Thus there exists an x_0 satisfying both conditions.

To finish the proof, one covers the set Fx_0 with translates $F_{a_i+j}gx_0 = F_{i,j}gx_0$ such that $A_{F_{i,j}} f(gx_0) > c\,\nu(f)$. The heart of the proof lies in finding a Vitali type subcover, that is by mostly disjoint sets that covers almost all of Fx_0. Then one finds that

$$\sum_{g\in F} f(gx_0) \gtrapprox \sum_{\text{subcover}} A_{F_{i,j}} f(gx_0)\mu(F_{i,j}) > c\,\nu(f) \sum_{\text{subcover}} \mu(F_{i,j}) \approx c\,\nu(f)\mu(F),$$

yet on the other hand

$$\sum_{g\in F} f(gx_0) = A_F f(x_0)\mu(F) < (1+\varepsilon)\nu(f)\mu(F).$$

It is interesting to note that to find a subcover, one uses a simple randomized algorithm. The desired Vitali type property needs the sequence to be tempered. □

1.11 Invariant Measures

We have thus far avoided the question of existence of invariant measures. In some cases, the answer is easy. For example

i) For a discrete group, the counting measure is invariant.

ii) For a Lie group (real or p-adic), one can push-forward by translations any top-dimensional form at the identity. The differential form obtained is invariant, and the measure it defines as well.

iii) For a compact group K with invariant probability measure μ acting on a space X, one can take any measure ν on X and average it over K: $\nu' = \displaystyle\int_K k\nu \, d\mu(k)$. The measure ν' is K-invariant.

For groups in fact, we have the following completely satisfying answer:

Theorem 1.11.1 (Haar)**.** *Any locally compact second countable group supports an invariant measure.*

1.12 Moore's Ergodicity Theorem

An important and natural question that one might ask is the following: given an ergodic G-space X and a subgroup H, when is H ergodic on X? For semisimple Lie groups, Moore's Ergodicity Theorem gives a complete answer to the above question. In a subsequent chapter will be presented an application of mixing to counting integral points of symmetric varieties.

Notions from Lie Theory

The reader not aquainted with the theory of Lie groups should be thinking of matrix groups, even $SL_n(\mathbb{R})$ for simplicity. Such groups are manifolds with an analytic group structure, but perhaps more importantly, the tangent space at the identity has a structure of a Lie algebra.

Thus, one speaks of a semisimple Lie group as one having a nontrivial semisimple Lie algebra (we also ask for a finite center, and finite dimensional). Concretely, a simple Lie group is a nonabelian Lie group whose sole normal subgroups are 0-dimensional and lie in the center. A semisimple Lie group is then up to isogeny a product of simple groups. That is, G is semisimple if it can be represented as

$$G = (S \times G_1 \times ... \times G_m)/F,$$

where S is compact, G_i are simple, and F is a finite central group.

A lattice Γ in a Lie group G is a discrete closed subgroup having finite invariant covolume. That is, the space G/Γ is locally isomorphic to G (except perhaps at finitely many points) and the Haar measure on G pushes-forward to a G-invariant measure on G/Γ for which the whole space has finite measure.

Statements and Examples

Moore's Ergodicity Theorem is truly a statement about simple groups, and to lift it from simple to semisimple, one needs exactly the notion of irreducibility. The issue is exemplified in the following: suppose $G = G_1 \times G_2$, $X = X_1 \times X_2$ where X_k is an ergodic G_k-space, and G acts componentwise on X. Then G_1 seen as a closed subgroup of G does not act ergodically on X. Irreducibility avoids products of the kind.

Definition 1.12.1. Let $G = G_1 \times ... \times G_m$ be a product of non-compact connected simple Lie groups with finite center. Let X be an ergodic G-space with invariant finite measure. Then the action is said to be *irreducible* if for any non-central normal subgroup $N \subset G$, N is ergodic on X.

A lattice Γ in G is termed *irreducible* if for any non-central normal subgroup N of G, ΓN is dense in G. The action of G by left translation on G/Γ is irreducible if and only if the lattice Γ is irreducible. A nontrivial example of such a lattice would be as follows: let id, σ be the embeddings of $\mathbb{Q}[\sqrt{2}]$ in $\mathbb{R}$. Then $SL_2(\mathbb{Z}[\sqrt{2}])$ embedded in $SL_2(\mathbb{R}^2)$ via $\gamma \mapsto (\gamma, \sigma(\gamma))$ is an irreducible lattice.

Theorem 1.12.2 (Moore's Ergodicity Theorem)**.** *Let $G = G_1 \times ... \times G_m$ be a product of non-compact connected simple Lie groups with finite center. Suppose X is an irreducible ergodic G-space with finite invariant measure.*

Then a subgroup $H \subset G$ is mixing on X (hence ergodic) if and only if H has non-compact closure.

To give some nontrivial examples, we first introduce Moore's duality:

Proposition 1.12.3.

i) Let $H \subset G$ be a closed subgroup of a locally compact group G and X a nonsingular G-space. Then H is ergodic on X if and only if G acting diagonally on $X \times G/H$ is ergodic with respect to the product measure class.

ii) For two closed subgroups H_1 and H_2 of G, H_1 is ergodic on G/H_2 if and only if H_2 is ergodic on G/H_1.

Proof. Let $A \subset X \times G/H$ be G-invariant and neither null nor conull. For $y \in G/H$, define the section $A_y = \{x \in X \ : \ (x, y) \in A\}$. Invariance of A implies that for any $g \in G$, $gA_y = A_{gy}$. Fubini then implies that A_e is neither null nor conull. But A_e is H-invariant.

Conversely, let $B \subset X$ be H-invariant subset, and choose a Borel section $\sigma : G/H \to G$ of the projection $G \to G/H$. Define

$$A = \{(x, y) \in X \times G/H \ : x \in \sigma(y)B\}.$$

Since B is H-invariant and for $g \in G$, there exists an $h \in H$ such that $\sigma(gy) = g\sigma(y)h$, A is a G-invariant Borel set. If B is neither null nor conull, then the same holds for A.

For the second part, by the first part, H_1 is ergodic on G/H_2 if and only if G is ergodic on $G/H_1 \times G/H_2$, which is symmetric in H_1 and H_2. □

Example 1.12.4. Let G be as in Moore's Ergodicity Theorem, H a closed subgroup and Γ an irreducible lattice. Then Γ acting on G/H by translation is ergodic if and only if H is non-compact. Two interesting instances of this are:

i) $SL_n(\mathbb{Z})$ acting on $\mathbb{R}^n$ is ergodic, since this is equivalent to ergodicity of $\mathbb{R}^n \setminus \{0\}$, which is the homogeneous space $SL_n(\mathbb{R})/P$, where P is a maximal parabolic subgroup, hence non-compact.

ii) $SL_2(\mathbb{Z})$ acting on the real projective line $\mathbb{RP}^1$ is ergodic since the latter identifies again with $SL_2(\mathbb{R})/P$. Note here that there is no nontrivial $SL_2(\mathbb{Z})$-invariant measure, hence the necessity to include nonsingular systems in the theory.

Example 1.12.5. The geodesic flow and the horocyclic flow of a Riemannian manifold of constant negative curvature are mixing.

Definition 1.12.6. Given a strongly continuous unitary representation of G on the Hilbert space $\mathcal{H}$, for two vectors $v, w \in \mathcal{H}$, we define a matrix coefficient as the function $\phi_{v,w} : G \to \mathbb{C}$ by $g \mapsto \langle gv, w\rangle$.

The action of G on a space X with finite invariant measure is mixing if and only if all matrix coefficients vanish at infinity on $L^2_0(X)$, the orthogonal to constant functions in $L^2(X)$. Hence, Moore's Ergodicity Theorem is a consequence of the following:

Theorem 1.12.7 (Howe-Moore's Vanishing Theorem)**.** *Let G be a connected semisimple Lie group with finite center and $\mathcal{H}$ a strongly continuous unitary representation of G whose restriction to any noncompact simple factor has no nontrivial invariant vector. Then all matrix coefficients vanish at infinity.*

Proof of the Vanishing Theorem

Fix a sequence $\{a_n\}$ in G. Let

$$\begin{aligned} M^+ &= \{g \in G \ : \ e \text{ is an accumulation point of } \{a_n^{-1} g a_n\}\}. \\ N^+ &= \text{ closed subgroup generated by } M^+. \end{aligned}$$

Lemma 1.12.8 (Mautner Phenomenon)**.** *Let G be locally compact group and $\mathcal{H}$ a strongly continuous unitary representation of G. Let $\{a_n\}$ be a sequence in G and $v, v_0 \in \mathcal{H}$ be such that $a_n v \to v_0$. Then for all $g \in N^+$, $gv_0 = v_0$.*

Proof. Choose a subsequence a_k such that $a_k^{-1} g a_k \to e$. Then for any $w \in \mathcal{H}$,

$$\begin{aligned} |\langle gv_0 - v_0, w\rangle| &= \lim_{k\to\infty} |\langle g a_k v, w\rangle - \langle a_k v, w\rangle| \\ &= \lim_{k\to\infty} |\langle a_k^{-1} g a_k v, a_k^{-1} w\rangle - \langle v, a_k^{-1} w\rangle| \\ &\le \lim_{k\to\infty} \|(a_k^{-1} g a_k - e)v\| \|w\| = 0. \end{aligned}$$

□

We give a proof for G locally isomorphic to $SL_2(\mathbb{R})$ to avoid most of the structure theory of Lie groups. In that case the Lie algebra of G is $sl_2(\mathbb{R})$, and is generated by

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

We let then

$$A^+ = \exp \mathbb{R}_{\geq 0} H \quad U = \exp \mathbb{R} X \quad L = \exp \mathbb{R} Y.$$

The Cartan involution $\theta(\alpha) = -\alpha^T$ on the space of matrices leaves $sl_2(\mathbb{R})$ invariant. The $+1$ eigenspace is denoted $\mathfrak{k}$ and the -1 eigenspace by $\mathfrak{p}$. Let K be the connected Lie group having Lie algebra $\mathfrak{k}$. Structure theory yields an extension of θ to an involution Θ on G having K as the group of fixed points. The group K is a maximal compact subgroup of G, and $K \times \mathfrak{p} \to G$ given by $(k, X) \to k \exp(X)$ is a diffeomorphism onto G. Furthermore, $\mathfrak{p} = \cup_{k \in K} Ad(k)\mathbb{R}H$, from which we obtain the decomposition $G = K\bar{A}^+K$.

Lemma 1.12.9. *Suppose $\mathcal{H}$ is a strongly continuous unitary representation of G and that for all matrix coefficients ϕ, $\phi(a) \to 0$ as $a \in \bar{A}^+$ goes to infinity. Then all matrix coefficients vanish at infinity.*

Proof. Suppose that there exists a sequence $g_n \to \infty$ and $v, w \in \mathcal{H}$ such that $\phi_{v,w}(g_n)$ does not converge to zero. From the KAK-decomposition above, $g_n = k_n a_n k'_n$ for some $k_n, k'_n \in K$ and $a_n \in \bar{A}^+$. Taking converging subsequences, we assume $k_n \to k$ and $k'_n \to k'$. Then

$$\phi_{v,w}(g_n) = \langle a_n k'_n v, k_n^{-1} w \rangle \to \langle a_n k' v, k^{-1} w \rangle = \phi_{k'v, k^{-1}w}(a_n)$$

does not converge to zero, a contradiction. □

Choose $v \in \mathcal{H}$ and a sequence $\{a_n\}$ in $\bar{A}^+$. A small computation shows that $U \subset N^+$. Let v_0 be an accumulation point of $\{a_n v\}$. Mautner's lemma shows that v_0 is invariant under U. The next lemma shows that v_0 is invariant under the full group G. Since $\mathcal{H}$ has no nontrivial invariant vectors, this implies that $v_0 = 0$. Thus matrix coefficients vanish at infinity on $\bar{A}^+$, and by the previous lemma on G, finishing the proof.

Lemma 1.12.10. *Suppose v_0 is invariant under U, then it is invariant under G.*

Proof. We begin with $G = SL_2(\mathbb{R})$. The matrix coefficient ϕ_{v_0,v_0} is N-bi-invariant. We pick a sequence $t_n \to 0$, let $g_n = \begin{pmatrix} 0 & -t_n^{-1} \\ t_n & 0 \end{pmatrix}$. For

$a = \begin{pmatrix} \alpha & 0 \\ 0 & \alpha^{-1} \end{pmatrix}$,

$$\begin{pmatrix} 1 & \alpha t_n^{-1} \\ 0 & 1 \end{pmatrix} g_n \begin{pmatrix} 1 & \alpha^{-1} t_n^{-1} \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \alpha & 0 \\ t_n & \alpha^{-1} \end{pmatrix}.$$

Thus for any $a \in A$, $\phi_{v_0,v_0}(a) = \lim_{n\to\infty} \phi_{v_0,v_0}(g_n)$ is constant. Thus

$$\langle av_0, v_0\rangle = \langle ev_0, v_0\rangle = \|v_0\|^2.$$

Cauchy Schwarz in the equality case leads to av_0 is a constant multiple of v_0, which must be one.

For $G = PSL_2(\mathbb{R})$ follows as well since the representation lifts to $SL_2(\mathbb{R})$.

□

Two

Measure Classification Theorems

2.1 Statement of Ratner's Theorems

We consider the following elementary example first. Let $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ be a torus, and $v \in \mathbb{R}^n$. One obtains a flow $\phi_t(x) = x + tv \mod \mathbb{Z}^n$. Then there exists a vector subspace S of $\mathbb{R}^n$ such that

i) $v \in S$, so that $x + \mathbb{R}v \subset x + S$,

ii) $x + S \mod \mathbb{Z}^n$ is compact, hence diffeomorphic to a subtorus $\mathbb{T}^k$,

iii) $x + \mathbb{R}v$ is dense in $x + S \mod \mathbb{Z}^n$.

Furthermore, if μ is the Lebesgue measure on the subtorus $x + S \mod \mathbb{Z}^n$, then for any continuous function f on $\mathbb{T}^n$,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(\phi_t(x))dt = \int f\, d\mu.$$

Ratner's Theorems are a far reaching generalisation where $\mathbb{R}^n$ is replaced by a Lie group, $\mathbb{Z}^n$ by a lattice, and ϕ_t by a unipotent flow, or even a subgroup generated by unipotent one-parameter subgroups. The conclusion is that the orbit closures are nice geometric homogeneous subspaces, and for the case of a flow, the orbit is equidistributed on the closure.

Theorem 2.1.1 (Ratner's Orbit Closure)**.** *Let G be a Lie group, Γ a lattice in G, and U a subgroup of G generated by its unipotent one-parameter subgroups. Then for each $x \in G$, there exists a closed subgroup L of G such that*

i) L contains U,

ii) $Lx\Gamma/\Gamma$ *is closed and has finite* L*-invariant volume,*

iii) $Ux\Gamma/\Gamma$ *is dense in* $Lx\Gamma/\Gamma$*.*

Theorem 2.1.2 (Ratner's Equidistribution Theorem)**.** *Let* G *be a Lie group,* Γ *a lattice in* G*, and* u_t *a unipotent one-parameter subgroup. For* $x \in G$*, let* L *be the corresponding subgroup from the Orbit Closure Theorem, and* μ_L *the* L*-invariant measure on* $Lx\Gamma/\Gamma$*. Then for any compactly supported continuous function* f *on* G/Γ*,*

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(u_t x\Gamma/\Gamma)dt = \int f\, d\mu_L.$$

Theorem 2.1.3 (Ratner's Measure Classification)**.** *Let* G *be a Lie group,* Γ *a lattice in* G*, and* U *a subgroup of* G *generated by its unipotent one-parameter subgroups. Let* μ *be an* U*-invariant and ergodic probability measure on* G/Γ*. Then* μ *is homogeneous: there exist* $x \in G$ *and a subgroup* L *as in the Orbit Closure Theorem such that* μ *is the* L*-invariant probability measure on* $Lx\Gamma/\Gamma$*.*

Example 2.1.4. Let $G = PSL_2(\mathbb{R})$, $\Gamma = PSL_2(\mathbb{Z})$, so that G/Γ is the unit tangent bundle of the modular surface. Consider $u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$. The orbits $u_t x\Gamma/\Gamma$ are horocycles. They are either periodic or dense.

In contrast, if we consider $a_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}$, then the orbits are geodesics, and their closures can be fractal, of various Hausdorff dimension.

Example 2.1.5. For $G = SL_3(\mathbb{R})$, $\Gamma = SL_3(\mathbb{Z})$ and u_t as in the previous example, but embedded in the upper left corner of $SL_3(\mathbb{R})$, then some orbit closures are homogeneous under the smaller group $SL_2(\mathbb{R})$ also embedded in the upper left corner.

2.2 Margulis Theorem/Oppenheim Conjecture

That the Oppenheim Conjecture follows from the Orbit Closure Theorem was observed by Raghunatan (in the 60's, 70's?) who conjectured the above theorems proved by Ratner around 1990. The oppenheim conjecture was one of the main motivating application and was proved by Margulis in 1980, essentially proving a special case of the Orbit Closure Theorem. However, the proof being much lengthier, we present here how to derive it from Ratner's Theorem.

Recall that a quadratic form Q in n variables is *nondegenerate* if there does not exist $x \in \mathbb{R}^n$ such that for all $v \in \mathbb{R}^n$, $Q(x+v) = Q(x-v)$. In other words, it is not a quadratic form in fewer variables. It is *irrational* if it is not a scalar multiple of a quadratic form with rational coefficients. It is *indefinite* if it achieves both negative and positive values.

Theorem 2.2.1. *Let Q be a nondegenerate, irrational and indefinite quadratic form in $n \geq 3$ variables. Then the values $Q(\mathbb{Z}^n)$ is dense in $\mathbb{R}$.*

Proof. Without lost of generality, we can assume that $n = 3$. Otherwise, pick v_1 and v_2 in $\mathbb{Z}^n$ such that $Q(v_1)/Q(v_2)$ is negative and irrational, and pick $v_3 \in \mathbb{Z}^n$ generically so that the quadratic form $Q'(x) = Q(x_1v_1 + x_2v_2 + x_3v_3)$ is nondegenerate.

Let Q_0 be the quadratic form $x_1^2 + x_2^2 - x_3^2$. The group $SO(Q_0)(\mathbb{R})$ consists of elements $g \in SL_3(\mathbb{R})$ such that $Q_0(gv) = Q_0(v)$, and is isomorphic to $SO_{2,1}(\mathbb{R})$. It is generated by its unipotent one-parameter subgroups as it is locally isomorphic to $SL_2(\mathbb{R})$. Since Q has signature $(2,1)$ or $(1,2)$, there exists a $g \in SL_3(\mathbb{R})$ and $\lambda \in \mathbb{R}^x$ such that $Q = \lambda Q_0 \circ g$.

Let $G = SL_3(\mathbb{R})$, $H = SO(Q_0)^0$ and $\Gamma = SL_3(\mathbb{Z})$. Applying the Orbit Closure theorem to $x = g$, we obtain a closed intermediary group L between H and G such that the closure of $Hg\Gamma/\Gamma$ is $Lg\Gamma/\Gamma$. However, one shows that either $L = H$ or $L = G$.

If $L = G$, then $Q(\mathbb{Z}^3) = \lambda Q_0(g\mathbb{Z}^3) = \lambda Q_0(Hg\mathbb{Z}^3)$ which is dense in $Q_0(G\mathbb{Z}^3) = \mathbb{R}$. If however $L = H$, then Γ intersects $SO(Q)^0 = g^{-1}Hg$ in a lattice, and therefore $g^{-1}Hg$ is defined over $\mathbb{Q}$. This implies that Q is a multiple of an integral form. $\square$

The assumptions of the theorem are optimal. If the form is definite, then $Q(\mathbb{Z}^n)$ coincide with a norm squared of integral vectors, hence is discrete in $\mathbb{R}$. If the form is rational, then the values $Q(\mathbb{Z}^n)$ are rational with bounded denominator, hence discrete in $\mathbb{R}$.

If $n = 2$, let a be a root of an integral polynomial $p(t)$ of degree two. Then the form $Q(x,y) = x^2 - a^2y^2$ is nondegenerate, irrational and indefinite, yet does not have dense values on $\mathbb{Z}^2$. Indeed, supposing $y \neq 0$,

$$Q(x,y) = y^2\left(\frac{x}{y} - a\right)\left(\frac{x}{y} + a\right).$$

Remains to show that there exists a constant C such that $|a - x/y| > C/y^2$ (a is badly approximable). It suffices to show this for x/y in a compact interval around a, on which $p'(t)$ is bounded by $1/C$. Then by the Mean Value Theorem,

$$|a - \frac{x}{y}| \geq C|p(a) - p(x/y)| = C|p(x/y)| \geq \frac{C}{y^2}.$$

From the case $n = 2$, we can construct forms in $n \geq 3$ that degenerate to bad forms in two variables.

2.3 Einsiedler's Proof for $H = SL_2(\mathbb{R})$

We now present a proof of Ratner's Measure Classification Theorem for the case of a Lie group G, a discrete subgroup Γ, $X = G/\Gamma$, a subgroup H isomorphic to $SL_2(\mathbb{R})$ and a H-invariant and ergodic probability measure μ on X. The conclusion is that μ is homogeneous. The proof is elementary modulo two facts from the representation theory of SL_2.

Logistic of the Proof

We first sketch the proof:

i) Let $u_t = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$, for $t \in \mathbb{R}$. We saw in the proof of Moore's Ergodicity that μ is U-ergodic, for $U = u_{\mathbb{R}}$. We pick $\mathcal{G} \subset X$ a set of full measure of generic points for the u_t-flow:

$$\forall x \in \mathcal{G}\ \forall f \in C_c(X) \quad \frac{1}{T}\int_0^T f(u_t x)\,dt \to \int f\,d\mu.$$

ii) Let $L = Stab(\mu)^o$, a closed subgroup of G. We shall prove below that if μ is concentrated on a single L-orbit, that is there exists $x \in X$ such that $\mu(Lx) = 1$, then Lx is a closed orbit, and therefore μ is homogeneous.

iii) If a subset of the H-invariant set Lx has positive measure, then by ergodicity $\mu(Lx) = 1$. Therefore we assume that any orbit Lx has measure zero. This means that we can find many generic points y, y' not on the same L-orbit and as close together as we want. Moreover, we can arrange them to be transverse to L-orbits in the following sense. Let $\mathfrak{l}$ be the Lie algebra of L contained in the Lie algebra $\mathfrak{g}$ of G. The subgroup H acts on $\mathfrak{g}$ by the adjoint representation, leaving $\mathfrak{l}$ stable. The first fact about SL_2 that we use is that there exists a H-stable complement $\mathfrak{l}'$ to $\mathfrak{l}$ in $\mathfrak{g}$. Then there are generic points y', y such that $y' = \exp(v)y$ for some $v \in \mathfrak{l}'$ as small as we desire.

iv) The H-principle is the heart of the proof. It allows to conclude that in fact we can arrange to have not only $y' = \exp(v)y$ for $v \in \mathfrak{l}'$, but also v to be U-invariant, or in other words, $\exp(v)$ is in the centraliser

$C_G(U)$ of U in G. To obtain this, we flow along u_t for some appropriate amount of time. Then

$$uy' = u\exp(v)u^{-1}uy = \exp(Ad(u)v)uy.$$

The second fact about SL_2 we need is that the fastest growth of $Ad(u)v$ is in a U-stable direction, so that $uy' \approx cuy$ for some element $c \in C_G(U)$. Repeating the argument with closer y, y' and flowing for a longer time along u_t, we get that the approximation gets even better. We take a limit to obtain x, x' generic such that $x' = \exp(w)x$ for some $w \in \mathfrak{l}'$ and $\exp(w) \in C_G(U)$. Finally, we can repeat the argument again but at a different scale to obtain w as small as we want.

v) Since x, x' differ by an element $\exp(w) \in C_G(U) \cap \exp(\mathfrak{l}')$ and are both generic, it is easy to see using the Pointwise Convergence above that μ is invariant under the new element $\exp(w)$.

vi) For the final step, we use local compactness of $\mathfrak{l}'$ to pass from invariance under $\exp(\mathbb{Z}w_n)$ for $w_n \in \mathfrak{l}'$ arbitrarily small, to invariance under a one-parameter subgroup $\exp(\mathbb{R}w)$ for some $w \in \mathfrak{l}'$. Indeed, take w in a limit direction of the w_n. Since $w_n \to 0$, one can find for any $t \in \mathbb{R}$ integers $a_n \in \mathbb{Z}$ such that $a_n w_n \to tw$.

Proof: steps 2,3,5

We shall denote by B_r^S the ball of radius r in the group S centered at the identity.

Lemma 2.3.1. *(step 2) If for some $x \in X$, $\mu(Lx) = 1$, then Lx is closed.*

Proof. Let y be a limit point of Lx, and take a sequence $x_n = l_n x \to y$ for $l_n \in L$. If for some n, $x_n \in B_1^L y$, then $y \in Lx$. Otherwise, after taking a subsequence, $x_n \notin B_1^L x_m$ for any $n \neq m$. For $r < 1$ an injectivity radius at y, for large n, $x_n \in B_{r/2}^G y$, thus for large n the sets $B_{r/2}^L x_n$ are all disjoints. Since L contains a lattice, it is unimodular, and therefore $B_{r/2}^L x_n$ all have same measure, which contradicts $\mu(Lx) = 1$. □

Lemma 2.3.2. *(step 5) If $x, x' \in \mathcal{G}$, that is they are generic for the u_t-flow, and $x' = cx$ for some $c \in C_G(U)$, then μ is invariant by c.*

Proof. For $f \in C_c(X)$, define $f_c \in C_c(X)$ by $f_c(x) = f(cx)$, so

$$\frac{1}{T}\int_0^T f_c(u_t x)\,dt = \frac{1}{T}\int_0^T f(cu_t x)\,dt = \frac{1}{T}\int_0^T f(u_t x')\,dt.$$

Since both x, x' are generic, the right hand side converges to $\int_X f\, d\mu$ and the left hand side to $\int_X f_c\, d\mu = \int_X f\, d(c\mu)$. $\square$

One difficulty in the H-principle is to make sure that the limits still belong to $\mathcal{G}$. To do so, we define the following sets of large measure:

$$\begin{aligned}
&K \subset \mathcal{G} \text{ is a compact set with } \mu(K) > 0.9,\\
&X_1 = \left\{ x \in \mathcal{G} \,:\, \forall T > T_0,\ \frac{1}{T}\int_0^T \mathbb{1}_K(u_t x)\, dt > 0.8 \right\},\\
&X_2 = \left\{ x \in X \,:\, \frac{1}{m_L(B_1^L)}\int_{B_1^L} \mathbb{1}_{X_1}(lx)\, dm_L(l) > 0.9 \right\},
\end{aligned}$$

where T_0 is chosen large enough so that $\mu(X_1) > 0.99$, and m_L is a Haar measure on L.

Lemma 2.3.3. *$\mu(X_2) > 0.9$.*

Proof. Consider $f(x, l) = \mathbb{1}_{X_1}(lx)$ as a function on $X \times L$ with the product probability measure $\mu \times \tilde{m}_L$, where $\tilde{m}_L$ is the restriction of m_L to B_1^L. As μ is L-invariant, for any $l \in L$, $\int_X f\, d\mu = \mu(X_1)$. Then integrating with respect to L first concludes the proof. $\square$

From Step 3, we can assume $\mu(Lx) = 0$ for any $x \in X$. Thus for $x \in X_2 \cap supp(\mu)$ and for any $\delta > 0$, the set $\mu(B^G)\delta x \cap X_2$ has positive measure and is not contained in Lx. So for any $\delta > 0$, there exists $x' \in X_2$ such that $dist(x, x') < \delta$ and $x' = gx$ for some $g \in B_\delta^G \setminus L$.

Lemma 2.3.4. *For all $\varepsilon > 0$, there exists $y, y' \in X_1$ with $dist(y, y') < \varepsilon$ and $y' = \exp(v)y$ for some nonzero $v \in B_\varepsilon^{\mathfrak{l}'}$.*

Proof. Any $g \in G$ close to e can be written uniquely as $\exp(w)l$ for $w \in \mathfrak{l}'$ close to 0 and $l \in L$ close to e. Define $\pi_L(g) = l$. Now fixing g, define $\psi(l) = \pi_L(lg)$. Finally, define $E = \{l \in B_1^L : lx \in X_1\}$ and similarly for E' and x'. By definition of X_2, E and E' are large in B_1^L, and since ψ is close to the identity, $\psi(E') \cap B_1^L$ is also large. Therefore E intersects $\psi(E')$: there exists $l_1 \in E$ and $l_2 \in E'$ such that $l_1 = \psi(l_2)$. Both $l_1 x$ and $l_2 x'$ belong to X_1 and differ by $l_2 g l_1^{-1} = l_2 g \pi_L(l_2 g)^{-1} = \exp(v)$ for some $v \in B_\varepsilon^{\mathfrak{l}'} \setminus \{0\}$. $\square$

Proof step 4: the H-Principle

The finite dimensional representations of $SL_2(\mathbb{R})$ are completely reducible, hence it is enough to understand the irreducible ones. In fact, we have used this fact to define $\mathfrak{l}'$. All finite dimensional irreducible representations are symmetric tensor products of the standard representation on $\mathbb{R}^2$. One way to construct them up to isomorphism is by considering the space of homogeneous polynomials of degree n. For a polynomial $p(X,Y) = c_0X^n + c_1X^{n-1}Y + ... + c_nY^n$, the action is given by $u_tp(X,Y) = p(X, Y + tX)$. One computes

$$u_tp(X,Y) = (c_0+c_1t+...+c_nt^n)X^n+(c_1+2c_2t+...+nc_nt^{n-1})X^{n-1}Y+...+c_nY^n.$$

The direction of fastest increase is then X^n, which is stabilised by U.

We can decompose the representation $\mathfrak{l}'$ with the adjoint action of U into irreducible components $\oplus V_j$, each of the V_j isomorphic to a space of homogeneous polynomials of some degree. Suppose the vector v corresponds to polynomials p_j. For each j and for each scale $\eta > 0$, define

$$\begin{aligned} T_{p_j} &= \frac{\eta}{\max\{|c_1|, |c_2|^{1/2}, ..., |c_n|^{1/n}\}}, \\ T_v &= \min T_{p_j}, \end{aligned}$$

so that for $t = T_{p_j}$, the X^n coefficient of u_tp_j is at scale η, and the other coefficients are at smaller scale. This last assertion is captured in the following lemma.

Lemma 2.3.5. *Let $\eta > 0$ be a fixed scale, and $\varepsilon > 0$ arbitrarily smaller. Then there exist constants $n, C > 0$ such that for any $v \in B_\varepsilon^{\mathfrak{l}'} \setminus \{0\}$ and any $t \in [0, T_v]$, there exists a $w(t) \in \mathfrak{l}'$ stabilised by U, $\|w\| < C\eta$ and satisfying*

$$Ad(u_t)v = w(t) + O(\varepsilon^{1/n}),$$

where the implied constant in $O(\varepsilon^{1/n})$ is C (independently of t).

Furthermore, assuming $T_v < \infty$, there exist a constant $c > 0$ and a set $E_v \subset [0, T_v]$ of measure at least $0.9T_v$ such that if $t \in E_v$, then $\|w(t)\| > c\eta$.

Proof. Let $v \in B_\varepsilon^{\mathfrak{l}'} \setminus \{0\}$ correspond to polynomials p_j. We consider one of them, say $p(X,Y) = c_0X^n + c_1X^{n-1}Y + ... + c_nY^n$, where the coefficients c_i are $O(\varepsilon)$. By definition of T_p, the X^n coefficient of u_tp is bounded by $(n+1)\eta$, and the other coefficients are sums of terms c_it^j for $i > j$. But $t < \eta|c_i|^{-1/i}$, thus $|c_it^j|$ is $O(\varepsilon^{1/n})$, the implied constant depending on η and ε, but not on t.

Now assume $T_p < \infty$ and consider the polynomial map on $[0,1]$ given by

$$q(s) = c_0 + \frac{c_1}{T_p}s + \frac{c_2}{T_p^2}s^2 + ... + \frac{c_n}{T_p^n}s^n.$$

Its supremum is of the order of η. In fact it is between η^n and η. But polynomials on intervals are small relative to their supremum only on a small set relative to the length of the interval (see section on (C,α)-goodness). Since both scales are fixed, we obtain a constant $c > 0$ and a set $E_p \subset [0, T_v]$ of measure $0.9T_v$ such that for $t \in E_p$, $\|u_t p\| > c\eta$.

The same estimates for each p_j can then be summed, completing the lemma. $\square$

We now complete the proof of the H-principle, thus completing the proof of the measure classification theorem. Let $\eta > 0$ be a fixed scale and $\varepsilon > 0$ a smaller scale. By Lemma **??**, there exist $y, y' \in X_1$ with $y' = \exp(v)y$ for some $v \in B_\varepsilon^{\mathfrak{l}'} \setminus \{0\}$. If $T_v = \infty$, then v is already stabilised by U, so $\exp(v) \in C_G(U)$. Otherwise, making sure ε is small enough so that $T_v > T_0$ in the definition of X_1, we get that

$$E_T = \{t \in [0, T_v] \, : u_t y \in K\} \quad \text{and} \quad E_{T'} = \{t \in [0, T_v] \, : u_t y' \in K\},$$

both have measure larger than $0.8T_v$. Applying Lemma **??**, there exists a $t \in E_T \cap E_{T'} \cap E_v$ such that

i) $x = u_t y$ and $x' = u_t y'$ are in K,

ii) $x' = \exp(w + O(\varepsilon^{1/n}))x$, with $w \in \mathfrak{l}'$ stabilised by U,

iii) $c\eta < \|w\| < C\eta$.

Letting $\varepsilon \to 0$, we obtain again two points x, x' generic differing by an element $\exp(v) \in C_G(u)$ at scale at most η. The latter being arbitrary, this completes the H-principle.

2.4 Statement of EKL Measure Classification

The measure classification theorem for torus orbits is more subtle than Ratner's classification theorem. The reason is that contrarily to unipotent orbits, the local behaviour of toral orbits can be quite chaotic. We encountered in the H-principle above an instance of a local behaviour: unipotent orbits behave like polynomials on linear spaces, and these have the property that they are relatively small only for a relatively small time. However, toral orbits behave like exponentials on linear spaces, and these do not have that good property (unless we bound the exponents, that is take a small piece of the orbit).

Another subtlety is that in low dimension, it is known that the dynamical system is chaotic. For example, on $SL_2(\mathbb{R})/SL_2(\mathbb{Z})$ and for A the group of diagonal matrices, the system is Bernoulli. In particular, there are many A-invariant and ergodic measures, of any fractal dimension. A meaningful classification is then impossible. However, in higher rank and for split maximal torus, some rigidity can occur. It comes from the many stable and unstable folations for one-dimensional subtorii, which is used to create invariance under some unipotent groups and therefore to apply Ratner's classification theorem.

There is yet another important subtlety. Even in high dimension, there are lattices Γ in G for which the system that remains at "bounded" distance from a one-dimensional orbit, reducing the classification to one-dimensional classification, and thus is impossible. This exemplify that some arithmetics of the lattice should also be playing a role.

Finally, we note that an orbit closure theorem and an equidistribution theorem as for unipotent orbits fail even in the presence of measure classification. There are examples with exceptional bad directions in a maximal torus that accumulates around some fractal sets, therefore destroying homogeneity of the orbit closure. But these directions are few, and they are supported on a set of measure zero, thus they are not seen by the invariant measure.

Theorem 2.4.1 (EKL Measure Classification)**.** *Let $n \geq 3$, let A the subgroup of diagonal matrices in $G = SL_n(\mathbb{R})$ and let $\Gamma = SL_n(\mathbb{Z})$. Suppose μ is an A-invariant and ergodic measure on $X = G/\Gamma$ such that its entropy $h_\mu(a)$ is positive for some $a \in A$. Then μ is homogeneous.*

Conjecture 2.4.2 (Margulis)**.** *The same holds removing the positive entropy assumption.*

2.5 Littlewood's Conjecture

Margulis' conjecture implies the following long standing conjecture.

Conjecture 2.5.1 (Littlewood)**.** *For any $u, v \in \mathbb{R}$,*

$$\liminf\{n\langle nu\rangle\langle nv\rangle\ : n \in \mathbb{N}\} = 0,$$

where $\langle x\rangle$ denotes the distance to the nearest integer.

Proposition 2.5.2. *Let $g(u,v)$ and $a(r,s)$ be the following elements of $G = SL_3(\mathbb{R})$:*

$$g = \begin{pmatrix} 1 & 0 & 0 \\ u & 1 & 0 \\ v & 0 & 1 \end{pmatrix} \quad a = \begin{pmatrix} e^{-r-s} & 0 & 0 \\ 0 & e^r & 0 \\ 0 & 0 & e^s \end{pmatrix}.$$

Let $x = g\Gamma/\Gamma$, where $\Gamma = SL_3(\mathbb{Z})$, and A^+ the semigroup of elements $a(r,s)$ for $r, s > 0$. Then u, v satisfy Littlewood's Conjecture if and only if orbit A^+x is unbounded in G/Γ.

Proof. Suppose A^+x is unbounded. Mahler's Compactness Criterion (see later) implies that for all $\varepsilon > 0$, there exist an $a \in A^+$ and $z = (n, m_1, m_2) \in \mathbb{Z}^3$ such that $\|agz\| < \varepsilon$. The product of the entries of the vector agz is then less than $c\varepsilon^3$. But it is also A-invariant, thus is the product of the entries of gz:

$$|n(nu + m_1)(nv + m_2)| < c\varepsilon^3.$$

Thus the Littlewood's Conjecture is valid for u, v.

Suppose the Littlewood's Conjecture is true for u, v. Then for any $\varepsilon > 0$, there are $n > 0$ and $(m_1, m_2) \in \mathbb{Z}^2$ such that $|n(nu + m_1)(nv + m_2)| < \varepsilon^5$. Suppose that one of $|nu + m_1|$ or $|nv + m_2|$ is greater than ε. Say $|nv + m_2| > \varepsilon$, and $|n(nu + m_1)| < \varepsilon^4$. Consider the sequence $qnv \mod 1$. If $q_1 nv$ and $q_2 nv$ differ by less than ε modulo one, then $\langle (q_1 - q_2)nv \rangle < \varepsilon$. Thus there exists an integer $q < 1/\varepsilon$ such that

$$|qn(qn + qm_1)| < \varepsilon^2 \quad \text{and} |qnv + m_2'| < \varepsilon,$$

for some $m_2' \in \mathbb{Z}$. Thus perhaps replacing the integers n, m_1, m_2, we can assume both $|nu + m_1|$ and $|nv + m_2|$ are no more than ε, and that $|n(nu + m_1)(nv + m_2)| < \varepsilon^3$. So there exist $r, s > 0$ such that for $z = (n, m_1, m_2)$, $a(r,s)g(u,v)z$ has entries no more than ε. Mahler's Compactness Criterion then implies A^+x is unbounded. $\square$

EKL measure classification is enough to prove that the subset of $\mathbb{R}^2$ of exceptions to the Littlewood's conjecture is a countable union of compact sets of box dimension zero. In particular, it has Hausdorff dimension zero. We refer the interested reader to (paper EKL).

Three

Non-Divergence in the space of lattices

3.1 Good Functions

Good functions appeared already in the proof of Ratner's Measure Classification Theorem for $SL_2(\mathbb{R})$. There we introduced a large compact subset of generic points and other related sets of large measure. Flowing along u_t, since polynomials of bounded degree are good functions, we could lend most of the time in those large sets.

Good functions on a ball B in a measured metric space are characterised by the property that they are relatively small on B only on a relatively small subset of B. This is quantified in the following definition.

Definition 3.1.1. Let (X, μ, d) be a measure metric space, U an open subset, and C and $\alpha > 0$ be real constants. A function $f : U \to \mathbb{R}$ is said to be (C, α)-*good* if for any ball $B \subset U$,

$$\mu(\{x \in B \,:\, |f(x)| < \varepsilon \sup_{y \in B} |f(y)|\}) < C\varepsilon^{\alpha}\mu(B).$$

This property is scale invariant, and polynomials of bounded degree modulo scalars have a compact smooth parametrisation. Thus they form a family of (C, α)-good functions for some fixed C and α:

Proposition 3.1.2. *Polynomials on the real line and of degree bounded by k are $(k(k+1)^{1/k}, 1/k)$-good.*

Proof. Let $\varepsilon > 0$, B an open interval and f a polynomial of degree no more than k. Denote $\|f\|_B$ the supremum of $|f(B)|$ and suppose

$$\mu(\{x \in B \,:\, |f(x)| < \varepsilon \|f\|_B\}) \geq m.$$

There are $k+1$ points x_i such that $|x_i - x_j| \geq m/k$ whenever $i \neq j$ and such that $|f(x_i)| < \varepsilon \|f\|_B$. By Lagrange Interpolation

$$f(x) = \sum_{i=1}^{k+1} f(x_i) \frac{\prod_{j\neq i}(x - x_j)}{\prod_{j\neq i}(x_i - x_j)}.$$

Taking the supremum and obvious bounds, we obtain

$$\|f\|_B \leq (k+1)\varepsilon \|f\|_B \mu(B)^k \frac{k^k}{m^k},$$

and thus $m \leq k(k+1)^{1/k}\varepsilon^{1/k}\mu(B)$. □

As an exercise, one can verify that if u_t is a unipotent one-parameter subgroup of $SL_n(\mathbb{R})$ and v a vector, then $\|u_t v\|$ is a good function.

3.2 Mahler's Compactness Criterion

Let $G = SL_n(\mathbb{R})$, $\Gamma = SL_n(\mathbb{Z})$ and $X = G/\Gamma$. The space X identifies with the space of lattices in $\mathbb{R}^n$ with base point $\mathbb{Z}^n$.

Proposition 3.2.1 (Mahler's Compactness Criterion). *Suppose $C \subset SL_n(\mathbb{R})$. Its image in X is precompact if and only if* 0 *is* not *an accumulation point of $C\mathbb{Z}^n$.*

Proof. Suppose $C\Gamma$ is precompact. Choose a compact set $C_0 \subset G$ such that $C_0\Gamma \supset C\Gamma$. Since $C_0(\mathbb{Z}^n \setminus \{0\})$ is closed, it contains all its accumulation points. But $0 \notin C_0(\mathbb{Z}^n \setminus \{0\})$.

For the other direction, we assume $n = 2$, the general case following by induction. Let $g_n \in G$ be a sequence such that 0 is not an accumulation point of $\cup g_n(\mathbb{Z}^2 \setminus \{0\})$. We will show that there exists $\gamma_n \in \Gamma$ such that $g_n\gamma_n$ has a convergent subsequence.

Let $v_n \in \mathbb{Z}^2 \setminus \{0\}$ be such that $\|g_n v_n\|$ is minimal, $\pi_n : \mathbb{R}^2 \to \mathbb{R}g_n v_n$ the projection, $\pi_n^\perp$ its orthogonal projection and $w_n \in \mathbb{Z}^2 \setminus \mathbb{R}v_n$ such that $\|\pi_n^\perp(g_n w_n)\|$ is minimal.

Perhaps changing w_n for $w_n + kv_n$, we can assume that $\|\pi_n(g_n w_n)\| \leq \|g_n v_n\|/2$. Then the minimality of $g_n v_n$ implies that $\|\pi_n^\perp(g_n w_n)\| \geq \|g_n v_n\|/2$.

Let γ_n be the matrix with column vectors v_n and w_n. Perhaps changing w_n by $-w_n$, we can assume that γ_n has determinant one. From the equality $1 = det(g_n\gamma_n) = \|\pi_n^\perp(g_n w_n)\|\|g_n v_n\|$, $g_n v_n$ is bounded. Passing to a subsequence, we can assume $g_n v_n \to v \neq 0$. In turn, $g_n w_n$ must remain bounded, and we can assume $g_n w_n \to w \notin \mathbb{R}v$. Let g be the matrix whose column vectors are v and w. Then $g_n\gamma_n \to g$. □

Definition 3.2.2. $X(\varepsilon)$ is the compact subset of lattices in $\mathbb{R}^n$ whose shortest nonzero vector has length no less than ε.

3.3 Protected Points

Suppose u_t is a unipotent one-parameter subgroup of G. Then $\|u_t v\|$ is a good function. Thus either $u_t v$ is a fixed vector, or it is small only for a small time. However in higher rank, before it becomes big, another vector can become small. Hence below we consider not only vectors, but primitive subgroups Δ of lattices, which forms a poset of finite rank.

Let $\mathcal{P}(\mathbb{Z}^n)$ be the poset of primitive subgroups of $\mathbb{Z}^n$ under inclusion.

Definition 3.3.1. Given an interval $B \subset \mathbb{R}$, $\mathcal{P}$ a sub-poset of $\mathcal{P}(\mathbb{Z}^n)$, a map $h : B \to G$ and $0 < \varepsilon < \rho < 1$, we say that $x \in B$ is *(ε/ρ)-protected relative to $\mathcal{P}$* if there exists a flag $\mathcal{F} = \Delta_1 \subset ... \subset \Delta_l$ such that

$$(M1)\ \varepsilon\rho^{rk(\Delta)-1} \leq \|h(x)\Delta\| \leq \rho^{rk(\Delta)}\ \forall\Delta \in \mathcal{F}$$
$$(M2)\ \|h(x)\Delta\| \geq \rho^{rk(\Delta)}\ \forall\Delta \in \mathcal{P} \setminus \mathcal{F} \text{ comparable with } \mathcal{F}$$

Proposition 3.3.2. *Let $0 < \varepsilon < \rho < 1$, then for $x \in B$ (ε/ρ)-protected relative to $\mathcal{P}(\mathbb{Z}^n)$, one has $h(x)\mathbb{Z}^n \in X(\varepsilon)$.*

Proof. Let $\Delta_0 \subset ... \subset \Delta_l$ be the flag protecting x concatenated with $\{0\}$ and $\mathbb{Z}^n$. For any nonzero integral vector v, there is some j such that $v \in \Delta_j \setminus \Delta_{j-1}$. Let $\Delta = \mathbb{R}(\Delta_{j-1} + \mathbb{Z}v) \cap \mathbb{Z}^n$, which is a primitive subgroup between Δ_{j-1} and Δ_j. Whether Δ is in the flag or not, properties M above leads to the inequality $\|h(x)\Delta\| \geq \varepsilon\rho^{rk(\Delta_{j-1})}$. On the other hand, submultiplicativity of norms give us $\|h(x)\Delta\| \leq \|h(x)\Delta_{j-1}\|\|h(x)v\|$. Combining them and using properties M, we obtain $\|h(x)v\| \geq \varepsilon$. $\square$

3.4 Measure of Non-Protected Points

Protected points are those for which $h(x)\Gamma$ have no short vectors, hence remains in a compact region. For non-divergence to occur, we must have most points being protected. The following is a quantitative version of this.

Theorem 3.4.1. *Let $\mathcal{P}$ be a sub-poset of $\mathcal{P}(\mathbb{Z}^n)$ of length k. Assume that there are constants C, α and ρ such that for any $\Delta \in \mathcal{P}$,*

i) $\|h(x)\Delta\|$ *is (C, α)-good on B,*

ii) $\sup_{x\in B} \|h(x)\Delta\| \geq \rho^{rk(\Delta)}$.

Then for any $\varepsilon < \rho$,

$$\mu(\{x \in B \,:\, x \textit{ is not } (\varepsilon/\rho)\textit{-protected relative to } \mathcal{P}\}) \leq k2^k C(\frac{\varepsilon}{\rho})^\alpha \mu(B).$$

Proof. For $k = 0$, all points are protected. We now proceed by induction.

Let $y \in B$, $S(y) := \{\Delta \in \mathcal{P} : \|h(y)\Delta\| < \rho^{rk(\Delta)}\}$. By discreteness of $\mathbb{Z}^n$, $S(y)$ is a finite set. If empty, $\mathcal{F} = \emptyset$ can be used to protect y for any ε. Let $E := \{y \in B : S(y) \neq \emptyset\}$.

For $y \in E$ and $\Delta \in S(y)$, let $B_{\Delta,y}$ be the maximum interval on which $\|h(x)\Delta\| < \rho^{rk(\Delta)}$. Thus $\sup_{x \in B_{\Delta,y}} \|h(x)\Delta\| = \rho^{rk(\Delta)}$.

For $y \in E$, choose $\Delta_y \in S(y)$ such that $B_y = B_{\Delta_y,y}$ is maximum, hence contains all other $B_{\Delta,y}$. Thus,

$$\forall y \in E \; \forall \Delta \in \mathcal{P}, \sup_{x \in B_y} \|h(x)\Delta\| \geq \rho^{rk(\Delta)}.$$

For $y \in E$, let $\mathcal{P}_y$ be the poset of $\Delta \in \mathcal{P} \setminus \{\Delta_y\}$ which are comparable to Δ_y. It has rank $k - 1$. We can therefore apply induction with $B = B_y$ to obtain

$$\mu(\{x \in B_y : x \text{ not } (\varepsilon/\rho)\text{-protected relative to } \mathcal{P}_y\}) \leq (k-1)2^{k-1}C(\frac{\varepsilon}{\rho})^\alpha \mu(B_y).$$

For $x \in B_y$, x protected relative to $\mathcal{P}_y$, there exists a flag $\mathcal{F}'$ in $\mathcal{P}_y$ such that

i) $\varepsilon \rho^{rk(\Delta)-1} \leq \|h(x)\Delta\| \leq \rho^{rk(\Delta)}$, for any $\Delta \in \mathcal{F}'$,

ii) $\|h(x)\Delta\| \leq \rho^{rk(\Delta)}$, for any $\Delta \in \mathcal{P}_y \setminus \mathcal{F}'$ comparable to $\mathcal{F}'$.

However, by definition of Δ_y, x is *not* protected relative to $\mathcal{P}$. Let $\mathcal{F} = \mathcal{F}' \cup \{\delta_y\}$. The lower estimate in i) can fail, but on a set of small measure by (C, α)-goodness:

$$\mu(\{x \in B_y : \|h(x)\Delta_y\| < \varepsilon \rho^{rk(\Delta_y)-1}\}) \leq C\varepsilon^\alpha \mu(B_y).$$

To finish the proof, we combine the estimates of the last two paragraphs. Consider $\{B_y : y \in E\}$ and a subcovering $\{B_i\}$ of multiplicity at most two. Then

$$\mu(\{x \in E : \text{not protected relative to } \mathcal{P}\}) \leq \sum_i \mu(\{x \in B_i : \text{not protected relative to } \mathcal{P}\}) \leq k2^{k-1}C\varepsilon^\alpha \sum_i \mu(B_i) \leq k2^k C\varepsilon^\alpha \mu(B).$$

□

3.5 Non-Divergence

Combining all of the above, we obtain the following

Corollary 3.5.1. *Let $B \subset \mathbb{R}$ be an interval, C, α positive constants, $0 < \rho < 1$ and $h : B \to SL_n(\mathbb{R})$ continuous. Assume that for any $\Delta \in \mathcal{P}(\mathbb{Z}^n)$,*

i) $\|h(x)\Delta\|$ *is* (C, α)*-good,*

ii) $\sup_{x \in B} \|h(x)\Delta\| \geq \rho^{rk(\Delta)}$.

Then for any $\varepsilon < \rho$, λ the Lebesgue measure,

$$\lambda(\{x \in B : h(x)\mathbb{Z}^n \notin X(\varepsilon)\}) \leq n2^n C(\frac{\varepsilon}{\rho})^\alpha \mu(B).$$

Thus, effective non-divergence is reduced to verifying the two conditions of the corollary. The second condition is one of linear algebra.

Four

Counting Integral/Rational Points of Homogeneous Varieties

Although many techniques exist for counting, this chapter will be focused on main term counting of lattice points on homogeneous varieties via ergodic theory. There are in large two methods unified by an equidistribution theorem. One is through mixing, which has the advantage of being effective, and the other is through measure classification, which has the advantage of being broader.

The setting is as follows. Let Γ be a lattice in G, and H a unimodular subgroup of G. Suppose we are given a one-parameter increasing family $\{B_R : R \in \mathbb{R}\}$ of bounded measurable subsets of G/H with volume tending to infinity. The counting function is defined by

$$\begin{aligned} F_R : G/\Gamma &\to \mathbb{N} \\ g\Gamma &\mapsto \sum_{\gamma \in \Gamma/\Gamma\cap H} \mathbb{1}_{B_R}(g\gamma H/H). \end{aligned}$$

Question 4.0.2. What is the asymptotic behaviour of the counting function as $R \to \infty$?

To get a heuristic for the expected answer, let us assume that H is trivial. One can tile G with fundamental domains for Γ, each tile containing a single element of Γ. The tiling covers more or less well the sets B_R, with possible boundary nuisance. Ignoring the latter, one would obtain $F_R(e) = \#\Gamma \cap B_R$ to be about the volume of B_R divided by the volume of a fundamental domain. For H nontrivial, for the heuristic to apply, we require $\Gamma \cap H$ to be a lattice in H. Then one expects that

$$F_R(g\Gamma) \sim \frac{\mu_H(H/H\cap\Gamma)}{\mu_G(G/\Gamma)} vol(B_R),$$

where μ_G is a Haar measure on G/Γ, μ_H a Haar measure on $H/H\cap\Gamma$, vol is a compatible G-invariant measure on G/H, and $f(R) \sim g(R)$ means that the limit as $R \to \infty$ of the ratio is one. Note that the right-hand side does not depend on the choice of normalisation of the Haar measures, and we will therefore assume μ_G and μ_H to be probability measures.

4.1 From Counting to Equidistribution

Note that the expected answer is independent of $g \in G$, but this is not always the case. In particular, we wish to assume at least quasi-invariance of the sets B_R, so that there is some smooth relation between the family of sets B_R and the group G. We note that in most natural examples, this condition is met easily, but nonetheless is far from being trivial. For example, lifting the sets B_R from G/H to G will in general destroy quasi-invariance, so that the stabiliser H does play an important role in the counting problems.

Definition 4.1.1. The family of sets B_R is *well-rounded* if for any $\varepsilon > 0$, there exists a neighbourhood U in G of e such that

i) $\bigcup_{g\in U} gB_R \subset B_{(1+\varepsilon)R}$,

ii) $\bigcap_{g\in U} gB_R \supset B_{(1-\varepsilon)R}$,

iii) and there exists a function $\delta(\varepsilon)$ such that $\delta \to 1$ as $\varepsilon \to 0$ and

$$\delta(\varepsilon)^{-1} vol(B_{(1+\varepsilon)R}) \leq vol(B_R) \leq \delta(\varepsilon) vol(B_{(1-\varepsilon)R}).$$

We normalise the counting function: let $\hat{F}_R = F_R/vol(B_R)$.

Proposition 4.1.2. *Suppose the family B_R is well-rounded. If the measures $\hat{F}_R d\mu_G$ converge in the weak-*-topology to a limit which is locally around $g\Gamma/\Gamma$ absolutely continuous with respect to μ_G and with density C, then $F_R(g\Gamma/\Gamma) \sim C\, vol(B_R)$. In particular, if the limit is μ_G, then the heuristic is valid.*

Proof. We use approximation of identity. Given $\varepsilon > 0$, let U be as in the definition of well-roundness. Let $\alpha : G/\Gamma$ be a smooth function such that $\int_{G/\Gamma} \alpha \, d\mu_G = 1$ and $supp(\alpha) \subset U$. Then

$$\frac{\delta^{-1} F_{(1-\varepsilon)R}(e)}{vol(B_{(1-\varepsilon)R})} \leq \frac{F_{(1-\varepsilon)R}(e)}{vol(B_R)} \leq \int \frac{\alpha F_R}{vol(B_R)} d\mu_G \leq \frac{F_{(1+\varepsilon)R}(e)}{vol(B_R)} \leq \frac{\delta F_{(1+\varepsilon)R}(e)}{vol(B_{(1+\varepsilon)R})}.$$

But the limit of the integral being one, we obtain

$$\delta^{-1} \limsup_{R\to\infty} \frac{F_R(e)}{vol(B_R)} \leq 1 \leq \delta \liminf_{R\to\infty} \frac{F_R(e)}{vol(B_R)}.$$

Letting $\varepsilon \to 0$, $\delta \to 1$, and therefore $F_R(e) \sim vol(B_R)$. □

The measures $\hat{F}_R d\mu_G$ can now be studied using duality: roughly speaking, Γ-orbits on G/H relates to H-orbits on G/Γ. To be more precise, we have a double fibration

$$G/\Gamma \leftarrow G/\Gamma \cap H \rightarrow G/H$$

leading to the following two averaging maps:

$$B_c(G/H) \to L^1(G/\Gamma): \quad \psi \mapsto \psi^\Gamma(g\Gamma) = \sum_{\gamma\in\Gamma/\Gamma\cap H} \psi(g\gamma H/H) \text{ and}$$
$$C_b(G/\Gamma) \to C(G/H): \quad \phi \mapsto \phi^H(gH) = \int_{H/H\cap\Gamma} \phi(gh\Gamma/\Gamma)\, d\mu_H(h).$$

Then for any $\phi \in C_b(G/\Gamma)$ and $\psi \in B_c(G/H)$,

$$< \phi, \psi^\Gamma >_{L^2(G/\Gamma)} = < \phi^H, \psi >_{L^2(G/H)} \quad \text{,that is} \tag{4.1}$$
$$\int_{G/\Gamma} \phi\psi^\Gamma \, d\mu_G = \int_{G/H} \phi^H \psi \, dvol. \tag{4.2}$$

Finally, we note that the counting function F_R was nothing more than $\mathbb{1}_{B_R}^\Gamma$:

Corollary 4.1.3. *The counting estimate is established if one shows that*

$$\frac{1}{vol(B_R)} \int_{B_R} g\mu_H \, dvol(gH) \to \mu_G.$$

4.2 Counting via Mixing

For affine symmetric spaces, the following method works well. We prove that $g\mu_H \to \mu_G$ in two steps: firstly use the wavefront lemma to thicken $gH\Gamma/\Gamma$, secondly deduce equidistribution from mixing.

We record here the facts needed from the theory of symmetric spaces. An affine symmetric space is of the form G/H for a semisimple Lie group G with finite center and a closed symmetric subgroup H, by which we mean that there exists a Lie group involution $\sigma : G \to G$ such that its fix point set is exactly H. We have two decompositions from the structure theory. The first one is a generalisation of the polar decomposition: there exists a maximal compact subgroup K and a maximal abelian group A of diagonalisable elements such that $G = KAH$. The second is a local version of the Iwasawa decomposition: let M be the centraliser of A in K. There exists N a unipotent group made of the positive roots of A such that $N \times A \times M \times H \to G$ is locally an open map. Moreover, A modulo a finite group contracts N.

Lemma 4.2.1 (Wavefront Lemma). *Let G/H be an affine symmetric space. For any neighbourhood V of e in G, there exists a neighbourhood U of e in G such that for all $g \in KA$, $gUH \subset VgH$.*

The proof is almost entirely contained in the above local Iwasawa decomposition and is left as an exercise.

Theorem 4.2.2. *Let G/H be affine symmetric, Γ a lattice in G. Assume the orbit $Y = H\Gamma/\Gamma$ has finite volume. Let μ_G, μ_H be the Haar probability measure on G/Γ and Y respectively. Then as $gH \to \infty$ in G/H, $g\mu_H$ converges to μ_G in the weak-*-topology.*

Proof. Let $g_n \in G$ be such that $g_nH \to \infty$ in G/H. From the KAH decomposition, we can assume $g_n \in KA$. Let ϕ be a continuous compactly supported function on G/Γ and let $\varepsilon > 0$. Choose $V \subset G$ such that $V = V^{-1}$ and for any $v \in V$, $|\phi(vg) - \phi(g)| < \varepsilon$ and let U be as in the wavefront lemma. From Moore's Ergodicity Theorem,

$$\frac{1}{\mu_G(UY)} \int_{g_nUY} \phi \, d\mu_G = \frac{1}{\mu_G(UY)} \int_{G/\Gamma} \mathbb{1}_{UY}(x)\phi(g_nx) \, d\mu_G(x) \to \int_{G/\Gamma} \phi \, d\mu_G.$$

From the wavefront lemma, the lefthand integral is contained in the closed convex hull of

$$\Big\{ \int_{vg_nY} \phi(h) \, d\mu_H(h) \, : v \in V \Big\} = \Big\{ \int_{g_nY} \phi(v^{-1}h) \, d\mu_H(h) \, : v \in V \Big\}.$$

From the uniform continuity of ϕ, the latter integrals are within ε of $\int_{g_n Y} \phi \, d\mu_H$, finishing the proof. $\square$

Example 4.2.3. The hyperbolic plane is G/K for $G = PSL_2(\mathbb{R})$ and $K = SO_2(\mathbb{R})$. We let $\Gamma = PSL_2(\mathbb{Z})$ and B_R be the ball of radius R centered at i. The volume of B_R is $2\pi(\cosh R - 1)$ and is easily seen to be a well-rounded family. The space G/Γ is the unit tangent bundle to the modular surface, and $Y = K\Gamma/\Gamma$ is the base point with unit tangent vectors pointing in all directions. For $g \in A$, gY then is a sphere of radius R with normal unit vectors. The wavefront lemma is a consequence of the triangle inequality on the modular surface, and the fact that the angles between vectors don't vary much is a consequence of negative curvature. We obtain thus that the number of elements $\gamma \in \Gamma$ such that $d(i, \gamma.i) < R$ grows like $2\pi e^R$.

Example 4.2.4. Here is a non-Riemannian example. Let ℓ be a geodesic in the hyperbolic plane with stabiliser H and descending to a closed geodesic on G/Γ, i.e. $H \cap \Gamma \simeq \mathbb{Z}$. Then $\Gamma\ell$ is a locally finite collection of geodesics. We can count the number of them intersecting the balls B_R as above, even if $\Gamma\backslash G/H$ is not Hausdorff. The equidistribution of gY geometrically reads as the parallel at distance t of ℓ becomes equidistributed on the modular surface.

Example 4.2.5. An example of a non symmetric case would be by considering in the previous example a horocycle instead of a geodesic. Here one can mimic the proof above and find both equidistribution and counting estimate. However, it will fall in a more general theorem below using Ratner's measure classification.

Exercises

Chapter 1

Exercise 1. Prove Proposition 1.1.4.

Exercise 2. Prove Jensen's Inequality (Proposition 1.1.5).

Exercise 3. Prove that $\mathbb{E}^{\mathcal{A}} : L^2(X, \mathcal{B}, \mu) \to L^2(X, \mathcal{A}, \mu_{|\mathcal{A}})$ is the orthogonal projection.

Exercise 4. Prove that a nonsingular system being incompressible implies conservative.

Exercise 5. Prove using a diagonal argument that any compact metrisable space with a continuous transformation admits an invariant probability measure.

Exercise 6. Suppose G acts continuously on a compact space X with invariant measure μ. Prove that if it is uniquely ergodic, then every orbit is dense in $supp(\mu)$.

Exercise 7. Show that for G acting properly ergodically on (X, μ), the σ-algebra of invariant sets is not countably-generated.

Exercise 8. Show that for any invariant measure μ, the σ-algebra $\mathcal{I}$ of invariant sets is μ-equivalent to the σ-algebra $\mathcal{J}$ of μ-almost invariant sets, that is the sets A such that for all $g \in G$, $\mu(g^{-1}A \triangle A) = 0$. Show also that $\mathcal{J}$ is maximal with respect to $\subset_\mu$.

Exercise 9. Prove that ergodicity of (X, G, μ) could also have been defined as one of the equivalent statement:

i) Any μ-almost invariant set A has measure zero or full measure.

ii) If A and B are two sets of positive measure, then there exists $g \in G$ such that $\mu(g^{-1}A \cap B) > 0$.

Exercise 10. Give a second proof of Proposition 1.8.2 using Birkhoff Ergodic Theorem instead of the Mean Ergodic Theorem.

Exercise 11.

Exercise 12.

Exercise 13.