DYNAMICS AND RIGIDITY

SIMION FILIP

Spring 2018

ABSTRACT. Course Notes for Math 253, Spring 2018, Harvard. Please let me know of any inaccuracies.

CONTENTS

Revised APRIL 24, 2018 .

LIST OF CORRECTIONS

1. Introduction

The course will be concerned with maps $X \to X$ from a space to itself. The simplest case is when the maps generate a group like $\mathbb{Z}$ or $\mathbb{R}$, but our context will often have a larger group of maps, such as a Lie group (perhaps p-adic) with $\mathbf{SL}_2\,\mathbb{R}$ a representative example.

Here are two basic examples to keep in mind, in both cases we'll have transformations of the circle $\mathbb{T}^1 := \mathbb{R}/\mathbb{Z}$. For the first, fix an irrational α and consider the translation

$$T_\alpha(x) := x + \alpha \quad \mod 1$$

This action is isometric for the natural metric (see Exercise 9.1.1). All points behave uniformly for this transformation: every orbit is dense, in fact equidistributed in an appropriate sense.

At the opposite extreme is the doubling transformation

$$M_2(x) := 2x \quad \mod 1$$

This action is uniformly expanding on the circle, and the orbit of a point x is determined by its base 2 expansion. The distribution properties of an orbit can be as chaotic as those of a random coin flip.

The goal of the course will be to explore situations when one can make uniform conclusions as in the first example, but when the setup is closer to that of the second example. A key tool is to study invariant probability measures on X. Given a transformation $T : X \to X$ and a measure μ on X, its push-forward is defined by $T_*\mu(A) := \mu(T^{-1}A)$. This action extends the transformation to the space of all (say probability) measures, which is more flexible.

The key point of most statements below is that they hold *for all* orbits of points $x \in X$, and not just for almost all. General ergodic theory can yield statements that hold for almost all points, in a natural measure-theoretic sense. One speaks of rigidity if there is a statement that holds for all orbits, and the possibilities are easily enumerated.

1.1. Abelian Rigidity

The results in this category are less definitive and more than what is currently known is expected to be true. The first instance of the phenomena discussed is the following result of Furstenberg.

1.1.1. Theorem (Furstenberg [Fur67], topological $\times 2, \times 3$ rigidity)**.** *Consider the multiplicative semigroup $\Gamma := \left\langle 2^a \cdot 3^b : a, b \in \mathbb{N} \right\rangle$ acting on the circle $\mathbb{T}^1 := \mathbb{R}/\mathbb{Z}$, by descending the multiplicative action from $\mathbb{R}$. Then the Γ-orbit of any irrational $x \in \mathbb{T}^1$ is dense.*

The following measure rigidity analogue is still open.

1.1.2. Conjecture. *For the same action as above, the only ergodic Γ-invariant probability measures are either atomic, or Lebesgue measure.*

Rudolph's theorem says that the only possible exceptions to the above conjecture must have zero entropy for the action of any element of the semigroup. One interpretation of the above statements is that for an irrational, there should be no relation between the digits of the expansion in base 2 and base 3.

A seemingly unrelated statement is the following question.

1.1.3. Conjecture (Littlewood)**.** *Let $\alpha, \beta \in \mathbb{R}$ be irrational numbers, and for a real number γ denote by $\|\gamma\|$ the distance to the nearest integer. Then*

$$\liminf_{n \to \infty} n \cdot \|n\alpha\| \cdot \|n\beta\| = 0$$

It turns out that this conjecture is equivalent to a statement about dynamics on homogeneous spaces. Here is the connection.

1.1.4. Proposition. *Consider in the space of lattices $\mathbf{SL}_3(\mathbb{R})\big/\mathbf{SL}_3(\mathbb{Z})$ the lattice $\Lambda_{\alpha,\beta}$ spanned by the vectors*

$$\begin{bmatrix} 1 \\ \alpha \\ \beta \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Let A be the group of positive diagonal matrices of determinant 1. Then the orbit $A \cdot \Lambda_{\alpha,\beta}$ is not contained in a compact set if and only if the pair (α, β) satisfies Littlewood's conjecture.

Proof. For a vector $v = (v_1, v_2, v_3) \in \mathbb{R}^3$, the product of its coordinates $v_1 \cdot v_2 \cdot v_3$ is invariant under A.

By Mahler's compactness criterion, a set of lattices is not contained in a compact set if and only if there is a sequence of non-zero vectors in the lattices whose norms go to zero (for a fixed ambient metric). For the case at hand, this is equivalent to there being a sequence of elements $a_i \in A^+$ and vectors $v_i \in a_i \cdot \Lambda_{\alpha,\beta} \setminus \{0\}$ such that $|v_i| \to 0$, where $|v_i|$ denotes their norm for a fixed metric.

By construction, the products of coordinates of v_i goes to zero as well. But v_i and $a_i^{-1} v_i \in \Lambda_{\alpha,\beta}$ have the same products of coordinates. Now an integral vector in $\Lambda_{\alpha,\beta}$ is of the form

$$\begin{bmatrix} k_1 \\ k_1\alpha + k_2 \\ k_1\beta + k_3 \end{bmatrix} \text{ for } k_i \in \mathbb{Z}$$

The claim that there exist k_1, k_2, k_3 such that the product of entries is arbitrarily small is equivalent to Littlewood's conjecture. □

Unfortunately, the conjectured measure or topological rigidity results are not yet known in the generality needed. However, Einsiedler, Katok, and Lindenstrauss proved that the set of possible exceptions to Littlewood's conjecture has Hausdorff dimension 0.

The methods that are currently available do give the following application.

1.1.5. Theorem (Lindenstrauss, Arithmetic Quantum Unique Ergodicity). *Suppose that $\Gamma \subset \mathbf{SL}_2(\mathbb{R})$ is an arithmetic lattice. Let ϕ_i be a sequence of normalized eigenfunctions of the Laplace operator on the hyperbolic surface $\mathbb{H}^2/\Gamma$, with eigenvalues tending to infinity.*

Then the sequence of probability measures $|\phi_i|^2 \,\mathrm{dVol}$ tends weakly to Lebesgue measure.

The result is saying that no subsequence of eigenfunctions can concentrate unevenly on subsets of the surface. Again, by general principles one knows (Shnirelman) that for a sequence of eigenfunctions of positive upper density the statement holds. The key point is to show it for *any* sequence of eigenfunctions.

The relation between Theorem 1.1.5 and measure rigidity is via Hecke operators. The key point is the arithmetic structure of Γ, which says that there exists a $\mathbb{Q}$-algebraic group G such that $\Gamma = G(\mathbb{Z})$ and [1] $\mathbf{SL}_2(\mathbb{R}) = G(\mathbb{R})$.

Then one can lift the situation to a measure on the space

$$G(\mathbb{R}) \times G(\mathbb{Q}_p)/G(\mathbb{Z}[1/p]) \text{ with a left } A(R \times \mathbb{Q}_p) - \text{action}$$

where A is a split Cartan, i.e. a subgroup isomorphic to the diagonal matrices. The action of $A(\mathbb{Q}_p)$ corresponds to the Hecke action.

Unlike in the case of the Littlewood conjecture, one can show that the action of A does have some positive-entropy directions. This then suffices to apply the existing measure rigidity results.

1.2. Unipotent Rigidity

A good introduction to unipotent measure rigidity are the notes of Eskin [Esk10].

The following result, whose statement involves only quadratic forms, was initially attacked by methods of analytic number theory and was

---

[1]This is abusive, there isn't really a $G(\mathbb{Z})$ for a $\mathbb{Q}$-algebraic group, but some finite index version of it

settled for sufficiently large dimensions. Raghunathan made the connection to unipotent flows, and Margulis settled the conjecture.

1.2.1. Theorem (Oppenheim Conjecture). *Suppose that Q is an indefinite quadratic form on $\mathbb{R}^n$ with $n \geq 3$. If Q is not proportional to a rational form, then for any $\varepsilon > 0$ there exists $v \in \mathbb{Z}^n$ such that*

$$0 < |Q(v)| < \varepsilon$$

Equivalently, the values $Q(\mathbb{Z}^n)$ accumulate to 0.

1.2.2. Remark.

(i) If α is a quadratic irrational such that $\alpha^2 \notin \mathbb{Q}$, then the quadratic form

$$x^2 - \alpha^2 y^2$$

is uniformly bounded from below on integer points. Indeed, a quadratic irrational α is badly approximable in the sense that $\exists \varepsilon$ such that

$$\left| \frac{p}{q} - \alpha \right| \geq \frac{\varepsilon}{q^2}$$

Note that if $\alpha^2 \notin \mathbb{Q}$, it is also a quadratic irrational.

(ii) By restricting the variables, it suffices to prove the Oppenheim conjecture for $n = 3$.

1.2.3. Relation to Unipotent Flows. A proof of the Oppenheim conjecture follows from the following result in homogeneous dynamics.

1.2.4. Theorem (Margulis). *Let $H := \mathbf{SO}_{2,1}(\mathbb{R})$ and consider its action on $X := \mathbf{SL}_3 \mathbb{R} / \mathbf{SL}_3 \mathbb{Z}$. Then for any $x \in X$, the orbit Hx is either dense, or $\mathrm{Stab}_x H \cap \mathbf{SL}_3 \mathbb{Z}$ is a lattice in H and hence the orbit is closed.*

The key property of the group $\mathbf{SO}_{2,1}(\mathbb{R})$ is that it is generated by unipotent matrices, which is not true of $\mathbf{SO}_{1,1}(\mathbb{R})$. Theorem 1.2.4 and the Oppenheim conjecture are related by the following observation, due to Raghunathan. The claim that $Q(\mathbb{Z}^n)$ accumulates to zero is equivalent to $Q(g \cdot \mathbb{Z}^n)$ accumulating to zero, for any $g \in \mathbf{SO}(Q)$. This gives a lot more flexibility, and a group action, for checking the desired property. The following statement makes this intuition precise.

1.2.5. Proposition. *Set $G := \mathbf{SL}_3(\mathbb{R})$, $\Gamma := \mathbf{SL}_3(\mathbb{Z})$ and $H := \mathbf{SO}_{2,1}(\mathbb{R})$.*

(i) *The space of unimodular quadratic forms of signature $(2,1)$ on $\mathbb{R}^3$ is naturally identified with $H \backslash G$.*

(ii) *There is a natural equivalence between closed Γ-invariant sets on $H \backslash G$ and closed H-invariant sets on G / Γ.*

(iii) *If a quadratic form $Q \in H \backslash G$ has dense orbit under the right action of Γ, then it satisfies the Oppenheim conjecture.*

(iv) *If a quadratic form $Q \in H \backslash G$ has a stabilizer under the right action of Γ which is a lattice in H, then Q is proportional to a rational quadratic form.*

With the correspondence between closed sets as above, Theorem 1.2.4 implies that either the Γ-orbit of a quadratic from is dense in $H \backslash G$, or the quadratic form is rational. From the density of the Γ-orbit in $H \backslash G$ one can deduce further strengthenings of the property $\exists v \in \mathbb{Z}^n, 0 < |Q(v)| < \varepsilon$.

Proof. For (i), fix the quadratic form

$$Q_0(x_1, x_2, x_3) := x_1^2 + x_2^2 - x_3^2$$

Any other unimodular quadratic form Q on $\mathbb{R}^3$ of signature $(2, 1)$ can be expressed as $Q(x) = Q_0(gx)$ for some $g \in \mathbf{SL}_3 \mathbb{R}$. The stabilizer of Q_0 is exactly H.

The correspondence between closed sets stated in (ii) is valid in greater generality, and is proved by identifying both sets with closed left-H and right-Γ invariant sets on G.

For (iii), it is clear that there is an irrational form Q' such that $\varepsilon/2 > |Q'(v_0)| > 0$ for some fixed $v_0 \in \mathbb{Z}^3$. By assumption, there is a $\gamma \in \Gamma$ such that $|Q(\gamma v_0) - Q'(v_0)| < \varepsilon/4$. Therefore the conclusion of the Oppenheim conjecture is satisfied.

Part (iv) follows from the Borel density theorem, which says that a lattice is Zariski-dense in the ambient real algebraic group. This implies that the real algebraic group $\mathbf{SO}(Q)$ stabilizing the quadratic form Q can be defined over the rationals, and this in turn implies that Q is proportional to a form defined over the rationals. $\square$

1.2.6. Remark (Aside)**.** Mahler's compactness criterion is not needed in the equivalence of Oppenheim with the homogeneous dynamics statement. Indeed one could have $H/H \cap \Gamma$ be non-compact and the orbit still be closed, if the relevant quadratic form represents zero over $\mathbb{Q}$.

The most general results for unipotent rigidity, both topological and measure-theoretic, are due to Ratner.

1.2.7. Theorem (Ratner, topological rigidity)**.** *Suppose that G is a semisimple real algebraic group, Γ is a lattice in G and H is a subgroup which can be generated by unipotent elements. Then for any $x \in G/\Gamma$, there exists a real algebraic group $L \subset G$ such that the orbit closure $\overline{H \cdot x}$ equals $L/(L \cap \Gamma)$ and $L \cap \Gamma$ is a lattice in L.*

The theorem above, in turn, is based on a measure-theoretic rigidity result.

1.2.8. Theorem (Ratner, measure rigidity)**.** *Suppose that G is a semisimple real algebraic group, Γ is a lattice in G and H is a subgroup which can be generated by unipotent elements. Then for any ergodic H-invariant probability measure μ on G/Γ there exists a real algebraic group $L \subset G$, with $L \cap \Gamma$ a lattice in L, such that μ equals the Haar measure on $L\big/(L \cap \Gamma)$.*

In Section 6 we will prove the measure rigidity theorem in two special cases. One is when $H \cong \mathbf{SL}_2\,\mathbb{R}$, and another when H is a horospherical subgroup. This last case contains, for example, the case of $G = \mathbf{SL}_2\,\mathbb{R}$ and $H = U$, a unipotent subgroup.

A distinctive feature of unipotent dynamics is its similarity to the isometric case, as opposed to the exponentially expanding situation. In other words, under the action of a unipotent group on a homogeneous space, points will diverge polynomially.

1.3. Non-Abelian Stiffness

The following elementary to state theorem was proved independently by Benoist–Quint [BQ11] and [BFLM11]. Recall that a matrix $A \in \mathbf{SL}_2\,\mathbb{Z}$ acts naturally on the torus $\mathbb{T}^2 := \mathbb{R}^2/\mathbb{Z}^2$.

1.3.1. Theorem (Benoist–Quint, topological rigidity)**.** *Suppose that $\Gamma \subset \mathbf{SL}_2(\mathbb{Z})$ is a finitely generated group which acts strongly irreducibly on $\mathbb{R}^2$. Then for any point $x \in \mathbb{T}^2$, its Γ-orbit is either finite or dense in $\mathbb{T}^2$.*

Equivalently, the only closed, Γ-invariant sets are either a finite collection of (rational) points, or all of $\mathbb{T}^2$.

A linear representation of a group is *strongly irreducible* if there is not finite collection of subspaces which is invariant by the group action.

1.3.2. Example. If the group Γ contains unipotents, then the result is not difficult. Another case where the problem should be approachable is when say all matrix entries are positive.

However, if say

$$\Gamma = \left\langle \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} -5 & -3 \\ 12 & 7 \end{bmatrix} \right\rangle$$

then the only way to prove the result is via the general methods of Benoist–Quint.

1.3.3. Remark.

(i) The strong irreducibility assumption is necessary. An equivalent one is that Γ is Zariski-dense in $\mathbf{SL}(2)$.

(ii) Note that if Γ is generated by a single hyperbolic matrix, then the result is false. Indeed, then the transformation is Anosov and the orbit closure of a point can have any Hausdorff dimension in the interval $[0, 2]$. This situation is very close to the doubling map of the circle.

The topological result above follows from a measure classification theorem. Since a general group action on a compact metric space doesn't have invariant measures, an extension of the notion is needed.

1.3.4. Definition (Stationary Measure)**.** Suppose that a topological group G acts on a space X. For a probability measure μ on G, a measure ν on X is called μ-*stationary* if $\mu * \nu = \nu$, where

$$\mu * \nu(A) := \int_G \nu(g^{-1}A) \, d\mu(g) = \int_G g_* \nu(A) \, d\mu(g)$$

For a general group G acting on a space X, there might be no invariant measures. However, for a fixed probability measure μ on G, any G-action on a compact space X always has at least one stationary measure.

1.3.5. Theorem (Benoist–Quint, measure rigidity)**.** *Suppose that μ is a finitely supported measure on $\mathbf{SL}_2 \mathbb{Z}$, whose support generates a Zariski-dense subgroup. Then the only ergodic μ-stationary measures on $\mathbb{T}^2$ are either atomic, or Lebesgue measure.*

1.3.6. Remark.

(i) The theorem implies that a μ-stationary measure is invariant by individual group elements in $\mathbf{SL}_2 \mathbb{Z}$. This property is called *stiffness* by Furstenberg [Fur98].

(ii) The topological rigidity theorem follows from the measure rigidity one, combined with a separation technique that will be discussed below.

(iii) There are some parallels between unipotent rigidity and the random walk (or non-abelian) rigidity results of Benoist–Quint. For example, the non-divergence and separation techniques are similar in both cases (see [EM04]).

1.4. Higher-rank lattices and Super-Rigidity

Of the results presented in the introduction, these are perhaps the earliest. The similarity with abelian actions comes from an extensive

use, in the proofs, of higher rank properties and in particular the existence of isometric directions.

1.4.1. Example. Consider the quadratic form

$$Q := x_1^2 + \cdots + x_p^2 - \sqrt{2}(x_{p+1}^2 + \cdots + x_{p+q}^2).$$

Let $\mathbf{SO}(Q)$ be the group of matrices preserving the quadratic form, viewed as a collection of equations on the entries. The matrices which satisfy the equations and have entries in a ring R are denoted by $\mathbf{SO}(Q)(R)$.

A theorem of Borel and Harish–Chandra (in this case, due to Siegel?) implies that $\mathbf{SO}(Q)(\mathbb{Z}[\sqrt{2}])$ is a lattice in $\mathbf{SO}(Q)(\mathbb{R}) \cong \mathbf{SO}_{p,q}(\mathbb{R})$. This is a typical example of an arithmetic lattice. If $p \geq q \geq 2$ then $\mathbf{SO}_{p,q}(\mathbb{R})$ is a higher-rank Lie group.

To see discreteness of $\mathbf{SO}(Q)(\mathbb{Z}[\sqrt{2}])$ in the real points, let σ denotes the Galois conjugation on $\mathbb{Z}[\sqrt{2}]$ with $\sigma(\sqrt{2}) = -\sqrt{2}$. Then the Galois-conjugate matrices preserve the Galois-conjugate quadratic form:

$$\sigma(Q) := x_1^2 + \cdots + x_p^2 + \sqrt{2}(x_{p+1}^2 + \cdots + x_{p+q}^2)$$

and $\mathbf{SO}(\sigma(Q))(\mathbb{R})$ is a compact group. Note now that $\mathbf{SO}(Q)(\mathbb{Z}[\sqrt{2}])$ embeds discretely into $\mathbf{SO}(Q)(\mathbb{R}) \times \mathbf{SO}(\sigma(Q))(\mathbb{R})$.

1.4.2. Theorem (Margulis Super-Rigidity)**.** *Suppose that $\Gamma \subset G$ is an irreducible lattice in a higher rank semisimple Lie group. Then any representation $\rho : \Gamma \to H$ with non-compact image to another Lie group extends to a continuous representation of G°, the connected component of the identity of G.*

This theorem holds, and is often used, when the target Lie group is non-Archimedean, e.g. $\mathbf{SL}_n(\mathbb{Z}_p)$. It allows, for example, to characterize lattices in higher rank Lie groups.

1.4.3. Theorem (Margulis Arithmeticity)**.** *Any irreducible lattice in a higher rank real semisimple Lie group is arithmetic.*

Proof. Suppose that Γ is a lattice in G. First, the rigidity of the identity representation inside G implies that any deformation of Γ is conjugated to Γ, therefore Γ has entries in an algebraic number field.

Now, taking again the identity representation, but in the $\mathbb{Q}_p$-points of the ambient $\mathbf{GL}_n$ implies that the denominators of all entries are uniformly bounded. Indeed, this is equivalent to the image being compact, and follows because a connected real Lie group doesn't have non-trivial continuous representations into a p-adic group. This implies that the lattice Γ is indeed arithmetic. $\square$

A related circle of ideas also gives group-theoretic information on Γ.

1.4.4. Theorem (Margulis Normal Subgroup). *Suppose that $\Gamma \subset G$ is a lattice in a higher-rank real Lie group.*

Then the only normal subgroups of Γ are either finite, or finite index.

The theorem is false for lattices in real rank 1 Lie groups. For example, there are lattices isomorphic to free groups in $\mathbf{SL}_2\,\mathbb{R}$. However, it can be salvaged by requiring that the normal subgroup is finitely generated (see Exercise 9.2.6).

2. Ergodic Theorems

Throughout this section, fix a measurable transformation preserving a probability measure, that is

$$T : X \to X \text{ with } \mu \text{ a probability measure and } T_*\mu = \mu$$

where $T_*\mu(A) := \mu(T^{-1}A)$ for all measurable A. For now, X can be any standard Borel space, i.e. a Borel subset of a Hausdorff separable metric space, with the induced Borel σ-algebra. The transformation T is only required to preserve the σ-algebra, i.e. $T^{-1}A$ is measurable, for any measurable set A.

An approach close to the discussion below can be found in [EW11, §2.6], though it doesn't discuss Kingman's theorem and goes via Vitali covers instead of Besicovitch covers (in dimension 1 it's not a significant difference).

2.1. Birkhoff and Kingman Ergodic Theorems

The pointwise Birkhoff ergodic theorem is one of the most used results in dynamics. It is concerned with the averages of a function $f : X \to \mathbb{R}$ along an orbit, denoted by

$$S_n f(x) := f(x) + f(Tx) + \cdots + f(T^{n-1}x)$$

and are sometimes called Birkhoff sums. Note that $T^*f(x) := f(Tx)$ and T^* preserves measurability of functions. Pull-back is the natural operation on functions, while push-forward is natural for measures.

2.1.1. Theorem (Birkhoff). *For $f \in L^1(\mu)$ the Birkhoff averages $\frac{1}{n}S_n f(x)$ converges for μ-a.e. x to a T-invariant function $\overline{f}$:*

$$\frac{1}{n}S_n f(x) \to \overline{f}$$

Moreover $\overline{f} \in L^1(\mu)$ and the convergence also holds in the L^1-norm:

$$\left\| S_n f - \overline{f} \right\|_{L^1} \to 0.$$

The following variant of the theorem is also useful and follows directly from the usual version. When writing T^i with $i < 0$, assume that T is measurably invertible. For any two reals $\alpha < \beta$ the convergence

$$\frac{1}{N(\beta - \alpha)} \sum_{i=\lfloor \alpha N \rfloor}^{\lfloor \beta N \rfloor} f(T^i x) \to \overline{f}$$

holds, where $\overline{f}$ is the same as in the usual Birkhoff theorem.

Most proofs of this theorem are based on two principles. First, the theorem is clearly true if $f = g - T^* g$, since in this case one gets a telescoping sum (and assuming say that g is bounded). The next step is to have some control on the averaging operators, in the form of control on their maximal size, via the maximal operators.

2.1.2. Proposition (Maximal Inequality, Positivity Version). *Consider the set where the Birkhoff sums become positive at some point:*

$$P := \{x \colon \limsup S_n f > 0\}$$

Then $\int_P f \, d\mu \geq 0$.

The same conclusion holds for any subset $P' \subset P$ which is also T-invariant, by restricting the dynamics to P'.

In many texts (e.g. [EW11]) the definition of P involves $\frac{1}{n} S_n f > 0$ instead of $S_n f > 0$. Note that the set where $\limsup \frac{1}{n} S_n > 0$ is T-invariant and contained in the set where only $\limsup S_n > 0$.

In fact, a slightly stronger version of Proposition 2.1.2 is also true, and will be useful when proving Kingman's theorem.

2.1.3. Proposition (Subadditive Maximal Inequality, Positivity Version). *Consider a sequence of functions $S_n \in L^1(X, \mu)$ satisfying the sub-additivity property*

$$S_{n+m}(x) \leq S_n(T^m x) + S_m(x)$$

Let $P := \{x \colon \limsup S_n > 0\}$. Then $\int_P S_1 \, d\mu \geq 0$.

The same conclusion holds for any subset $P' \subset P$ which is also T-invariant, by restricting the dynamics to P'.

To see the relation to the previous Proposition 2.1.2, just take $S_n := S_n f(x)$. Note that the sub-additivity property goes "the right way" in the sense that if we know S_n gets large, from the subadditive inequality so will S_1.

Proof. Define the maximal function

$$M^+S_n(x) := \max(0, S_1(x), \ldots, S_n(x))$$

Note that $0 \le M^+S_n \le M^+S_{n+1}$. Consider now the set

$$P_n := \left\{ x : M^+S_n(x) > 0 \right\}$$

and note that $P = \cup_n P_n$. On the complement of P_n we have $M^+S_n = 0$.
On P_n we have $M^+S_{n+1} \le M^+S_n(Tx) + S_1(x)$, by subadditivity since

$$M^+S_{n+1}(x) = \sup(0, S_1(x), \ldots, S_{n+1}(x))$$

$$M^+S_n(Tx) + S_1(x) = \sup(S_1(x), S_1(x) + S_n(Tx), \ldots, S_1(x) + S_n(Tx))$$

This gives the chain of inequalities

$$\int_{P_n} S_1 \, d\mu \ge \int_{P_n} (M^+S_{n+1}(x) - M^+S_n(Tx)) d\mu \qquad \text{from sub-additivity}$$

$$\ge \int_{P_n} (M^+S_n(x) - M^+S_n(Tx)) d\mu \qquad \text{since } M^+S_{n+1} \ge M^+S_n$$

$$\ge \int_X (M^+S_n(x) - M^+S_n(Tx)) d\mu = 0$$

where the vanishing of the last integral follows since the integrand is a coboundary. $\qquad \square$

Proof of Theorem 2.1.1. For real numbers $\alpha > \beta$, consider the set

$$E_{\alpha,\beta} = \left\{ x : \limsup \frac{1}{n} S_n f(x) > \alpha > \beta > \liminf \frac{1}{n} S_n f(x) \right\}$$

Note that $E_{\alpha,\beta}$ is T-invariant. By applying the maximal inequality, i.e.
Proposition 2.1.2 for the function $f - \alpha$, it follows that $\int_{E_{\alpha,\beta}} f - \alpha \ge 0$,
and similarly $\int_{E_{\alpha,\beta}} \beta - f \ge 0$. But this gives the contradictory inequality

$$\beta\mu(E_{\alpha,\beta}) \ge \int_{E_{\alpha,\beta}} f \ge \alpha\mu(E_{\alpha,\beta}).$$

To obtain the convergence in L^1 norm, first note that the limit $\overline{f}$
satisfies $\left\|\overline{f}\right\|_{L^1} \le \|f\|_{L^1}$. To show this, first note that it suffices to show
it assuming f is positive. But then by Fatou's lemma

$$\int \overline{f} = \int \lim \frac{1}{n} S_n f \le \lim \frac{1}{n} \int S_n f = \int f$$

Now the L^1-convergence follows for bounded f by the dominates convergence theorem. Any $f \in L^1(\mu)$ can be approximated by bounded
functions, i.e. $\exists f_\varepsilon$ bounded such that $\|f - f_\varepsilon\|_{L^1} \le \varepsilon$. This implies that

$$\lim \left\|\frac{1}{n}S_n f - \overline{f}\right\|_{L^1} \le \lim \left\|\frac{1}{n}S_n f_\varepsilon - \overline{f_\varepsilon}\right\|_{L^1} + \lim \left\|\frac{1}{n}S_n(f - f_\varepsilon) - \overline{f - f_\varepsilon}\right\|_{L^1}$$

$$\le 2\|f - f_\varepsilon\|_{L^1} \le 2\varepsilon$$

$\square$

2.1.4. Theorem (Kingman Subadditive Ergodic Theorem)**.** *Let $S_n \in$ $L^1(\mu)$ be a sequence of functions such satisfying*

$$S_{n+m}(x) \leq S_n(T^m x) + S_m(x)$$

Then the limit $\lim \frac{1}{n} S_n(x)$ *exists for μ-a.e. x and in fact equals* $\inf \frac{1}{n} S_n(x)$.

The limit function is clearly T-invariant.

Proof. The reasoning is almost identical to that in Theorem 2.1.1, but applying Proposition 2.1.3. In both cases, the reason why we can apply the maximal inequality with both $\limsup$ and $\liminf$ is because

$$\liminf_n (\beta - S_n) = -\limsup_n (S_n - \beta).$$

$\square$

2.1.5. Definition (Ergodicity)**.** A measure-preserving transformation $T : (X, \mu) \to (X, \mu)$ is *ergodic* if for any T-invariant set A, either $\mu(A) = 0$ or $\mu(X \setminus A) = 0$.

The definition above makes sense, and is interesting, even for infinite-measure preserving systems. Our focus however will be on probability measures.

For an ergodic system, any T-invariant function is necessarily constant. Therefore, the Birkhoff averages converges, for μ-a.e. x, to the same constant.

The following result related to the Birkhoff theorem is often useful.

2.1.6. Theorem (Atkinson, Furstenberg, Kesten)**.** *Let $f \in L^1(X, \mu)$ be such that for μ-a.e. $x \in X$*

$$\lim_{n \to \infty} f(x) + f(Tx) + \cdots + f(T^n x) = +\infty$$

Then $\int_X f d\mu > 0$ and so one has the stronger statement

$$\lim_{n \to \infty} \frac{1}{n} (f(x) + f(Tx) + \cdots + f(T^n x)) = \left(\int_X f d\mu \right) > 0$$

This result first appeared in the context of random walks on infinite-measure spaces, in this case the skew-product of $X \times \mathbb{R}$ with the transformation $(x, h) \mapsto (Tx, h + f(x))$. The statement is saying that either the random walk is drifting linearly to $+\infty$, or it is recurrent.

Proof. Let $S_i(x)$ denote the Birkhoff sums of f as before. Assume by contradiction that $\int f d\mu = 0$. Suppose we knew that $\forall \varepsilon > 0$ we have

$$\mu(\{x : |S_i(x)| > \varepsilon, \forall i \geq 1\}) = 0$$

Taking the preimage under T^p, this would imply that for any $p \in \mathbb{N}$ and $\forall \varepsilon > 0$ we have

$$\mu(\{x \colon |S_{i+p}(x) - S_p(x)| > \varepsilon, \, \forall i \geq 1\}) = 0$$

This follows since the measure μ is T-invariant, and if the latter set had positive measure, so would the former. Finally, the last condition for all p and a sequence of $\varepsilon \to 0$ implies, by taking unions, that

$$\mu(\{x \colon |S_{i+p}(x) - S_p(x)| > \varepsilon, \, \forall i \geq 1, \text{for some } p, \varepsilon > 0\}) = 0$$

But this is a contradiction to the claim that $S_i(x) \to \infty$ for μ-a.e. x.

Let us now show that $\forall \varepsilon > 0$ we have

$$\mu(\{x \colon |S_i(x)| > \varepsilon, \, \forall i \geq 1\}) = 0$$

Consider $B_n(\varepsilon, x)$ to be the union of ε-balls around $S_1(x), \ldots, S_n(x)$ in $\mathbb{R}$. By the assumption that $\int f = 0$ it follows from the Birkhoff ergodic theorem applied to f that $\lim_n \frac{1}{n}|B_n(\varepsilon, x)| = 0$ for μ-a.e. x, where $|B|$ denotes the Lebesgue measure of B, and therefore

$$(2.1.7) \qquad\qquad \frac{1}{n}\int_X |B_n(\varepsilon, x)| d\mu(x) \to 0$$

Now geometrically it is clear that

$$|B_{n+1}(\varepsilon, x)| - |B_n(\varepsilon, Tx)| \geq 2\varepsilon \mathbf{1}_{\{x \colon |S_i - S_1| > \varepsilon, \forall i = 2, \ldots n\}}$$

where $\mathbf{1}_E$ denotes the indicator function of a set E. Integrating, and using the T-invariance of the integral of $|B_n|$, gives that

$$\int |B_{n+1}| d\mu - \int |B_n| d\mu \geq 2\varepsilon \mu(\{x \colon |S_i - S_1| > \varepsilon, \forall i = 2, \ldots n\})$$

Finally, using that

$$\mu(\{x \colon |S_i - S_1| > \varepsilon, \forall i = 2, \ldots n\}) = \mu(\{x \colon |S_{i-1}| > \varepsilon, \forall i = 1, \ldots n - 1\})$$

combined with Eqn. (2.1.7) implies that the required set has measure zero. $\qquad\square$

Throughout the proof, we used several times the invariance or equivariance of various sets under the dynamics. We also used the fact that if certain quantities that look like Birkhoff sums (the $|B_n(\varepsilon, x)|$) don't grow at a linear speed, then their difference, on average, must also be getting arbitrarily small. Note that the quantity $|B_n(\varepsilon, x)|$ in fact satisfies the sub-additivity property that appears in Proposition 2.1.3.

2.1.8. Definition (Generic points)**.** Fix an ergodic T-invariant measure μ. A point $x \in X$ is *generic for μ and T* if there exists a countable set of continuous functions $f_i \in C^0(X)$, dense for the C^0 topology, such that the Birkhoff theorem holds for x and each f_i.

Generic points form a set of full μ-measure by the Birkhoff theorem. The Birkhoff theorem holds for x and any continuous function, by approximating it in the C^0 topology.

2.2. L^2 ergodic theorems

The proofs of ergodic theorems in the averaged, or integral sense, is shorter than that of the pointwise versions. However, most applications to rigidity theorems require the pointwise ergodic theorems. First, some terminology.

The pull-back operator T^* acting on L^2 functions is called the Koopman operator. Functions which can be written as $f = g - T^*g$ are called coboundaries, and we'll use this notation for general linear operators. The next theorem is quite general and applies in particular to T^* acting on $L^2(\mu)$.

2.2.1. Theorem (von Neumann ergodic theorem). *Let U be a unitary operator on a Hilbert space H.*

  (i) *The closure of the space of coboundaries in H is the orthogonal complement to the space of functions which are invariant under U:*

$$H = H^U \oplus \overline{Img(1 - U)}$$

  (ii) *Let $P : H \to H$ denote the projection onto the T-invariant functions. Then for any $f \in H$ we have*

$$\frac{1}{n}\left(f + Uf + \cdots + U^{n-1}f\right) \to P(f)$$

Proof. For the first part, note that an U-invariant function is also $U^\dagger$-invariant, since

$$
\begin{aligned}
0 = \|Uf - f\|^2 &= 2\|f\|^2 - 2(\langle Uf, f\rangle + \langle f, Uf\rangle) \\
&= 2\|f\|^2 - 2(\langle U^\dagger f, f\rangle + \langle f, U^\dagger f\rangle) \\
&= \|U^\dagger f - f\|
\end{aligned}
$$

So the space orthogonal to U-invariant functions is the same as that of $U^\dagger$-invariant functions. But we have

$$
\begin{aligned}
\langle f, g - Ug\rangle &= 0 \quad \forall g \in H \\
\Leftrightarrow \langle f, g\rangle &= \langle U^\dagger f, g\rangle \quad \forall g \in H \\
\Leftrightarrow f &= U^\dagger f
\end{aligned}
$$

For the second part, there are two options. One is to note that for coboundaries and for invariant functions the result is obvious, and the result is stable under approximating in the L^2 norm.

The other option is to note that if $f_n := \frac{1}{n}(f + Uf + \cdots + U^{n-1}f)$ then

$$\|Uf_n - f_n\| \to 0$$

and so one doesn't really need to characterize the orthogonal complement to the invariant functions. □

2.2.2. Remark. The L^2 convergence $\|S_n f - f\|_{L^2} \to 0$ implies the L^1 convergence $\|S_n f - f\|_{L^1} \to 0$ by Cauchy–Schwarz.

2.3. Coverings and Maximal Theorems

There is an approach to ergodic theorems which is essentially equivalent to the one described above, but is based on techniques (and terminology) closer to classical analysis. Two key concepts in the process are covering theorems and maximal inequalities, which also appear say in the Lebesgue differentiation theorem.

2.3.1. Theorem (Besicovitch covering theorem). *There exists a universal constant C_n with the following property. Let Ω be a bounded set in $\mathbb{R}^n$, and for each $\omega \in \Omega$ assume given a ball B_ω centered at ω. Then there exists a countable subset of these balls $\{B_{\omega_i}\}$ which cover Ω, and such that for any $\omega \in \Omega$, the number of selected balls covering it is bounded by C_n:*

$$\# \{i \colon \omega \in B_{\omega_i}\} \leq C_n$$

2.3.2. Remark.
  (i) The Birkhoff ergodic theorem will only require the result for $n = 1$, when C_1 can be taken to be 2.
  (ii) The theorem implies the result for $\mathbb{Z}^n$ instead of $\mathbb{R}^n$, which is the case used below.

2.3.3. Corollary (weak L^1 bound). *Let μ be a locally finite measure on $\mathbb{R}^n$ and $f \in L^1(\mu)$. Define the maximal function*

$$M_* f(x) := \sup_{r>0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} f \, d\mu$$

Then for $\lambda > 0$ we have the bound

$$\mu(x \colon M_* f(x) \geq \lambda) \leq \frac{C_n}{\lambda} \|f\|_{L^1(\mu)}$$

Proof. By a cutoff, we can assume that all objects in question are in a bounded set in $\mathbb{R}^n$ (note that the maximal function decays at ∞ at

least like $\frac{1}{\lambda} \|f\|_{L^1(\mu)}$). Now the set $\{x \colon M_* f(x) \geq \lambda\}$ can be covered, via Theorem 2.3.1 by a collection of balls $B(x_i, r_i)$ such that

$$\mu(B(x_i, r_i)) \leq \frac{1}{\lambda} \int_{B(x_i, r_i)} f d\mu \leq \frac{1}{\lambda} \int_{B(x_i, r_i)} |f| d\mu$$

Summing over all the balls and using the bound on the overlap gives the conclusion. $\square$

The maximal functions we already encountered before (with a different normalization), defined as

$$(2.3.4) \qquad \begin{aligned} M_n f &:= \sup_{1 \leq i \leq n} \frac{1}{i} S_i f \\ M_* f &:= \sup_n M_n f \end{aligned}$$

The proof of the Birkhoff ergodic theorem is now split into several parts.

2.3.5. Proposition (Birkhoff via Maximal Functions)**.**

(i) *Suppose that $f = h + g - T^* g$ where h is T-invariant and $g \in L^\infty(\mu)$. Then the Birkhoff ergodic theorem holds for f.*
(ii) *Functions as above are dense in $L^1(\mu)$ for the L^1 norm.*
(iii) *Any $f \in L^1(\mu)$ satisfies the following maximal inequality:*

$$\mu(x \colon M_* f \geq \lambda) \leq \frac{2}{\lambda} \|f\|_{L^1}$$

(iv) *If the Birkhoff ergodic theorem holds for a dense in L^1 set of functions, it holds for all L^1 functions.*

The maximal inequality should be compared with Chebyshev's inequality, which says that for any $f \in L^1(\mu)$ one has

$$\mu(x \colon |f(x)| \geq \lambda) \leq \frac{1}{\lambda} \|f\|_{L^1}$$

Proof. Part (i) is straightforward, and part (ii) follows from the von Neumann ergodic Theorem 2.2.1, since the functions described are dense in $L^2(\mu)$ and therefore also in $L^1(\mu)$. Part (iii) is the key point and will be proved below. For part (iv) one again shows that the $\limsup$ of $S_n f$ is a well-defined L^1 function, by approximating f in the L^1 norm by a function in the dense set where Birkhoff holds and using the maximal inequality.

For the proof of the maximal inequality, some reductions first. Define

$$\begin{aligned} B(\lambda) &:= \{x \colon M_* f(x) \geq \lambda\} \\ B(l, \lambda) &:= \{x \colon M_l f(x) \geq \lambda\} \end{aligned}$$

It is clear that $B(\lambda) = \cup B(l, \lambda)$ and it suffices to prove that

$$\mu(B(l, \lambda)) \leq \frac{2}{\lambda} \|f\|_{L^1}$$

independently of l. Now take N large and consider $X \times \{0, 1, \ldots N - 1\}$ equipped with the measure $\tilde{\mu} := \mu \otimes (\text{counting})$. Define now $F(x, i) := f(T^i x)$ on this new space and consider the set

$$\tilde{B}(l, \lambda) := \bigcup_{i=0,\ldots n-1} (T^{-i} B(l, \lambda)) \times \{i\}$$

The key point now is that for a fixed x we have the bound:

$$(2.3.6) \qquad \#\tilde{B}(x, l, \lambda) \leq \frac{2}{\lambda} \|f\|_{L^1(\{x\} \times \{0,\ldots N-1\})} + 2l$$

which follows by applying Corollary 2.3.3, with the extra $2l$ coming from the fact that we are working on a bounded set and the covering might overspill into an l-neighborhood of $\{0, \ldots, N - 1\}$ in $\mathbb{Z}$.

Now integrating Eqn. (2.3.6) over $x \in X$ for the measure μ gives

$$N \cdot \mu(B(l, \lambda)) \leq \frac{2N}{\lambda} \|f\|_{L^1(\mu)} + 2l$$

Dividing by N and sending $N \to \infty$ gives the result. $\qquad \square$

Proof of Besicovitch Theorem 2.3.1. First, we can assume that the radii of the balls are bounded say by M, since otherwise one sufficiently large ball will cover the entire set Ω (recall that Ω is bounded). Divide now the radii r_ω into scales:

$$R_k := \left\{ \omega \colon \frac{M}{2^{k+1}} \leq r_\omega \leq \frac{M}{2^k} \right\}$$

We will select finitely many points first from R_0, then from R_1, and so on, and the order of selection will be remembered.

When selecting points in R_k, take first an ω which is not covered by any of the balls previously selected. Repeat the same process until it cannot continue any longer with points from R_k. Only finitely many ω_i will be selected, since the distance between any two ω, ω' satisfies $d(\omega, \omega') \geq 2^{-k-1} M$. Indeed if say ω was selected first, then ω' must be outside the ball $B(\omega, r_\omega)$ to be also selected.

At the end of the process, the union of selected balls $B(\omega_i, r_{\omega_i})$ will cover the entire set Ω. Indeed, each ω belongs to one of the scales R_k, and if ω was not selected at that stage, then it must have been covered by some of the previous balls already.

It therefore suffices to check that the selected balls have uniformly bounded overlap. Assume that $0 \in \Omega$ (the origin) and let us bound the number of selected balls containing it.

First, there is a uniform bound A_1 on the number of selected balls which contain 0 and with $r \le r_\omega \le 10 \cdot r$, for some $r > 0$. Indeed, the distance between any two centers is at least r, and any of the balls $B(\omega_i, 10 \cdot r)$ contains 0. The bound A_1 on the number of such centers is independent of r.

It suffices now to check that there is another constant A_2 such that there are at most A_2 selected balls containing 0, and such that their radii are far apart: for any two $\omega \ne \omega'$ we have either $r_\omega > 10 r_{\omega'}$ or $r_{\omega'} > 10 r_\omega$. Assuming this, the balls containing 0 have radii in at most A_2 intervals of the form $[r, 10r]$ (for some collection of r's), so the total number of balls is at most $A_1 \times A_2$.

Suppose therefore that ω, ω' satisfy $r_\omega > 10 r_{\omega'}$ and $0 \in B(\omega, r_\omega) \cap B(\omega', r_{\omega'})$. Note that ω was necessarily selected before ω', and that $\omega' \notin B(\omega, r_\omega)$ by construction. This implies that the angle between ω and ω', as viewed from 0, is at least $\frac{1}{100}$ say. The bound A_2 then comes from the maximal number of $\frac{1}{100}$-separated sets on the unit sphere in $\mathbb{R}^n$. $\square$

2.4. Filling Scheme

To be filled in. A proof of Birkhoff, as well as the stronger version of Atkinson-Furstenberg-Kesten.

The approximate cohomological equation from the filling scheme.

Point of ergodic theorem: it's the Birkhoff sums that matter, not the averages.

2.5. Oseledets Theorem

An extended overview is in [Fil]. Below is a discussion based on Kingman's subadditive ergodic theorem (Theorem 2.1.4).

2.5.1. Cocycles. Let $T : (X, \mu) \to (X, \mu)$ be a measure-preserving system, and suppose that $E \to X$ is a vector bundle, perhaps with only measurable transition functions, and perhaps defined only μ-a.e. The lift of T to a *cocycle*[2] on E is a measurable map, denoted by abuse of notation also $T : E \to E$, such that on the fibers of E it acts by linear transformations. The fiber over $x \in X$ will be denoted E_x and the linear transformations are

$$T_x : E_x \to E_{Tx}$$

---

[2]This terminology is unfortunately standard in dynamical systems.

I E is equipped with a metric, a natural integrability condition on the cocycle is

$$(2.5.2) \qquad \int_X \log \|T_x\| \, d\mu(x) < +\infty$$

In fact one can relax the condition to the integrability of $\log^+ \|T_x\| := \max(0, \log \|T_x\|)$, by allowing some of the quantities below (e.g. the Lyapunov exponents) to be $-\infty$.

2.5.3. Theorem (Oseledets Multiplicative Ergodic Theorem)**.** *Suppose that $T : (E \to X, \mu) \to (E \to X, \mu)$ is a cocycle satisfying the integrability condition from Eqn. (2.5.2). Assume that μ is ergodic for T.*

Then there exist reals $\lambda_k < \lambda_{k-1} < \cdots < \lambda_1$, called Lyapunov exponents, and a filtration of E by T-invariant measurable subbundles:

$$\{0\} \subsetneq E^{\leq \lambda_k} \subsetneq \cdots \subsetneq E^{\leq \lambda_1} = E$$

such that for μ-a.e. x, if $v \in E_x^{\leq \lambda_i} \setminus E_x^{\leq \lambda_{i+1}}$ then

$$\lim_{n \to \infty} \frac{1}{n} \log \|T^n v\| = \lambda_i.$$

2.5.4. Remark.

    (i) The Lyapunov exponents and filtration are essentially independent of the choice of metric on E. For example, changing the metric pointwise by a bounded amount does not affect the limits, and in fact changing it by a factor which is log-integrable is also allowed.

    (ii) There is a two-sided version of the theorem, when T is an invertible map of X. Applying Theorem 2.5.3 to T^{-1} and appropriately intersecting the resulting filtrations gives a direct sum decomposition

$$E = E^{\lambda_1} \oplus \cdots \oplus E^{\lambda_k}$$

    into measurable subbundles such that if $v \in E^{\lambda_i}$ then

$$\lim_{n \to \pm\infty} \frac{1}{n} \log \|T^n v\| = \lambda_i$$

    Note that if say $\lambda_i > 0$ then this means into the future the vector v is expanded at an exponential rate λ_i, while into the past it is contracted at the same exponential rate.

2.5.5. Example.

    (i) If E is one-dimensional, then the Oseledets theorem reduces to the Birkhoff ergodic theorem.

(ii) If M is a smooth manifold and $f : M \to M$ is a smooth diffeormorphism with an invariant measure μ, then one can take $E := TM$, the tangent bundle of the manifold. In this case, there will in fact also be stable and unstable manifolds defined for μ-a.e. point corresponding to the Lyapunov subspaces.

2.5.6. Singular Values. Let $A : V \to W$ be a linear map between finite-dimensional real vector spaces, each equipped with a positive-definite inner product. Let $A^\dagger : W \to V$ denote the adjoint of A, and note that $A^\dagger \cdot A : V \to V$ is a positive-semidefinite self-adjoint operator. The spectral theorem implies that A can be diagonalized, with non-negative eigenvalues λ_i^2 and corresponding eigenvectors v_i, which are mutually orthogonal. The same reasoning applies to $A \cdot A^\dagger : W \to W$ and gives eigenvectors w_i. Now if $A^\dagger A v_i = \lambda_i^2 v_i$ then

$$AA^\dagger(Av_i) = A(A^\dagger A v_i) = \lambda_i^2(Av_i)$$

so it follows that A takes the eigenvectors v_i to w_i and moreover preserves the eigenvalues. Note also that $\langle Av_i, Av_i \rangle = \left\langle A^\dagger A v_i, v_i \right\rangle = \lambda_i^2 \langle v_i, v_i \rangle$, so it follows that A expands the norm of v_i by λ_i.

The numbers λ_i are called the *singular values* of A, written usually in increasing order: $\lambda_1 \geq \cdots \geq \lambda_n$. Note that only $\min(\dim V, \dim W)$ of them can be non-zero, and in most situations of interest $\dim V = \dim W$.

The largest singular value is the operator norm of A, indeed

$$\|A\| := \sup_{v \in V \setminus 0} \frac{\|Av\|}{\|v\|} = \lambda_1.$$

Singular values also behave as expected under tensor operations, for example taking an exterior power $\Lambda^p A : \Lambda^p V \to \Lambda^p W$, the norm of this operator will be the sum of the first p singular values:

$$\|\Lambda^p A\| = \lambda_1 + \cdots + \lambda_p$$

Proof of Theorem 2.5.3. The first step will be to see that the singular of the composed maps

$$E_x \xrightarrow{T_x} E_{Tx} \to \cdots \to E_{T^{n-1}x} \xrightarrow{T_{T^{n-1}x}} E_{T^n x}$$

have a limit, when normalized appropriately. The second step will be to identify the subbundles, which will be determined by the eigenvectors.

Consider the functions:

$$S_n(x) := \log \|T_x^{\circ n}\|$$

By assumption S_1 is integrable, and the sequence satisfies the subadditivity relation

$$S_n(x) + S_m(T^n x) \geq S_{n+m}(x)$$

since for a product of linear operators $\|A \cdot B\| \leq \|A\| \cdot \|B\|$. Theorem 2.1.4 implies that $\lim \frac{1}{n} S_n$ exists μ-a.e., which implies that the first singular value of the composed map satisfies:

$$\lim_{n \to \infty} \frac{1}{n} \log \lambda_1(T_x^{\circ n}) = \lambda_1$$

exists. Applying the same reasoning to the exterior power cocycles $\Lambda^p E$ implies that

$$\lim_{n \to \infty} \frac{1}{n} \left(\lambda_1(T_x^{\circ n}) + \cdots + \lambda_p(T_x^{\circ n}) \right) = \lambda_1 + \cdots + \lambda_p.$$

The Oseledets bundles are determined as the limit of the span of the corresponding eigenvectors $v_{i,n}(x) \in E_x$ for the map $T_x^{\circ n} : E_x \to E_{T^n x}$. That the limit does exist follows from a calculation, for which I refer to [Led84, Prop. 3.1]. □

2.6. Examples

2.6.1. Irrational Rotation. Let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ be irrational. Then rotation of the circle by an angle α, denoted R_α, is ergodic. By von Neumann's ergodic theorem, it suffices to check that the only functions in $L^2(S^1)$ which are invariant by R_α are the constants. But the action on Fourier series is by

$$R_\alpha(e^{2\pi\sqrt{-1}k \cdot x}) = e^{2\pi\sqrt{-1}k \cdot \alpha} \cdot e^{2\pi\sqrt{-1}k \cdot x}$$

so if $f = \sum \hat{f}(k) e^{2\pi\sqrt{-1}kx}$ and $f(R_\alpha x) = f(x)$ then f must be a constant. Indeed, $e^{2\pi\sqrt{-1}k \cdot \alpha} \neq 1$ for any $k \in \mathbb{Z} \setminus 0$ since α is irrational.

Note that for the irrational rotation, the Birkhoff ergodic theorem holds for *every* point $x \in S^1$, assuming the test function is continuous. Indeed, the rotation is an isometry, and for *any* point x, there is an arbitrarily close point x' to which the theorem does apply. By continuity of the test function, the Birkhoff averages of x and x' can be made arbitrarily close. The same argument implies the conclusion for any x and the test function equal to the indicator of an interval, which will be used in §2.6.2 below.

The explanation is that R_α is uniquely ergodic – there exists only one probability measure invariant by R_α. In fact, Lebesgue measure is the only *distribution* invariant by R_α, where distributions are viewed as functionals on smooth functions.

2.6.2. First digits of powers of 2. Consider

$$\{1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, \ldots\}$$

Which first digits occur more frequently? Note that the first digit of 2^n is i if and only if $i \cdot 10^k \leq 2^n < (i + 1) \cdot 10^k$ for some k. Taking logarithms, this is equivalent to

$$\log_{10} i \leq (n \cdot \log_{10} 2) \mod 1 < \log_{10}(i + 1)$$

Now the interval $[0, 1)$ is divided up into 9 intervals of the form $[\log_{10} i, \log_{10}(i + 1))$ and the first digit of 2^n is determined by which interval $n \cdot \log_{10} 2$ lands in.

Now $\log_{10} 2$ is irrational and the ergodic theorem then determines the frequency of first digits.

2.6.3. Bernoulli shift. Consider the space $\Omega := \{0, 1\}^{\mathbb{N}}$ equipped with the distance function

$$d(\omega, \omega') = 2^{-i} \quad \text{where}$$
$$\omega = \omega_0 \omega_1 \cdots \omega_i \omega_{i+1} \cdots$$
$$\omega' = \omega_0 \omega_1 \cdots \omega_i \omega'_{i+1} \cdots$$
$$\text{and } \omega_{i+1} \neq \omega'_{i+1}$$

Sets of the form

$$C_{x_0 \cdots x_k} := \{\omega \colon \omega_0 = x_0, \ldots, \omega_k = x_k\}$$

are called cylinder sets and generate the Borel σ-algebra. The cylinder sets are also norm balls with appropriate radius and center anywhere in the set.

The (left) shift transformation is defined to be:

$$S(\omega_0 \omega_1 \omega_2 \cdots) := \omega_1 \omega_2 \cdots$$

The preimage of a cylinder set satisfies:

$$S^{-1}(C_{x_0, \ldots, x_k}) = C_{0, x_0, \ldots, x_k} \cup C_{1, x_0, \ldots, x_k}$$

We can equip Ω with the product measure $\mu := \nu^{\otimes \mathbb{N}}$ where $\nu(0) := \frac{1}{2}$ and $\nu(1) := \frac{1}{2}$ and the shift S will preserve the measure. With this measure, the system satisfies a stronger property than ergodicity.

2.6.4. Definition (Mixing)**.** A probability measure preserving system $T : (X, \mu) \to (X, \mu)$ is *mixing* if for any two measurable sets $A, B \subset X$:

$$\lim_{n \to \infty} \mu(A \cap T^{-n}B) \to \mu(A) \cdot \mu(B)$$

Equivalently, for any $f, g \in L^2(\mu)$:

$$\lim_{n \to \infty} \int_X (T^*)^n f \cdot g \, d\mu \to \left(\int f d\mu \right) \left(\int g d\mu \right)$$

2.6.5. Remark.

(i) Mixing implies ergodicity, since the only positive reals x with $x^2 = x$ are $0, 1$.

(ii) It is enough to check mixing for a dense subset of functions in L^2 to obtain it in general, whereas for ergodicity one needs to consider all functions, or equivalently all measurable sets.

2.6.6. Proposition. *The Bernoulli shift* $S : (\Omega, \mu) \to (\Omega, \mu)$ *is mixing.*

Proof. It is enough to check mixing for cylinder sets. But if $C_{x_1,\dots,x_k}, C_{y_1,\dots,y_l}$ are two such, then for $n \geq l$:

$$(T^{-n}C_{x_1,\dots,x_k}) \cap C_{y_1,\dots,y_l} = C_{y_1,\dots,y_l,\bullet,x_1,\dots,x_k}$$

where $\bullet$ above signifies that the values of $\omega_{l+1}, \cdots, \omega_n$ can be anything in $\{0, 1\}$. By definition of μ, the mixing property therefore holds. $\square$

2.6.7. Conjugacy to doubling map. Consider the map

$$h : \Omega \to S^1$$

$$\omega_0 \omega_1 \cdots \mapsto \frac{\omega_0}{2} + \frac{\omega_1}{2^2} + \cdots + \frac{\omega_i}{2^{i+1}} + \cdots$$

This map is bijective when restricted to the irrationals on the unit circle, and it is $2 : 1$ on the rationals. Moreover it satisfies $h \circ S = M_2 \circ h$ where $M_2(x) := 2x \mod 1$ is the doubling map. Furthermore the pushed-forward measure $h_*(\mu)$ is Lebesgue measure on the circle.

Therefore, the two dynamical systems (S, Ω, μ) and $(M_2, S^1, dLeb)$ are measure-theoretically isomorphic. As a corollary, Lebesgue measure is ergodic, and even mixing, for the doubling map. The Birkhoff theorem then implies that for Lebesgue-a.e. $x \in S^1$, the orbit is uniformly distributed on the circle, in particular the base 2 digits of x have $0, 1$ with equal frequency. But for specific x, uniform distribution can fail.

2.6.8. Homogeneous spaces. For G a semisimple Lie group with lattice Γ, the quotient G/Γ has a finite Haar probability measure. We will see in §4.2 that the action on G/Γ of any non-compact 1-parameter subgroup of G is ergodic. In fact the stronger mixing property

$$\mu((g \cdot A) \cap B) \xrightarrow{g \to \infty} \mu(A) \cdot \mu(B) \quad \forall A, B \text{ measurable}$$

holds, where $g \to \infty$ means that g leaves every compact set of G.

3. Algebraic Groups and Ergodic Theory

References:

(i) Zimmer [Zim84]

(ii) Witte Morris [WM15]
(iii) Benoist's notes [Ben08]
(iv) Einsiedler–Ward [EW11]

3.1. Algebraic Groups: the basics

We will only be concerned with affine algebraic groups. The "classical" approach to the theory over a field k is to take an algebraic closure $\overline{k}$ and for any intermediate subfield $k \subset l \subset \overline{k}$ take the l-points inside the $\overline{k}$ points. This is unsatisfactory in many ways, not the least because it requires fixing once and for all an algebraic closure and hides available morphisms. The approach below is to take the defining equations as the basic invariant. Some familiarity with the Zariski topology and commutative algebra is assumed below.

A modern reference is [Mil17].

3.1.1. The base field. Throughout the discussions below, fix a field k. It can be $\mathbb{R}, \mathbb{C}$ or $\mathbb{Q}_p$ and even perhaps $\mathbb{F}_q((t))$, which are *local fields*. Interesting arithmetic questions appear when $k = \mathbb{Q}$.

3.1.2. Definition (Affine Variety). An affine variety over k, denoted $\mathbf{V}/k$, is the spectrum of a finitely generated k-algebra A. The *k-points* of $\mathbf{V}$ will be denoted $\mathbf{V}(k)$:

$$\mathbf{V}(k) := \mathrm{Hom}(A, k) \text{ as ring homomorphisms.}$$

More generally, for any k-algebra R we have the R-points

$$\mathbf{V}(R) := \mathrm{Hom}(A, R) \text{ as ring homomorphisms.}$$

A morphism of affine k-varieties $\mathrm{Spec}(A) \to \mathrm{Spec}(B)$ is the same as a ring homomorphism $B \to A$.

The algebra A is called *the ring of regular functions* on $\mathbf{V}$ and will be denoted $k[\mathbf{V}]$.

If $k[\mathbf{V}]$ is given explicitly using generators and relations:

$$k[\mathbf{V}] = k[x_1, \ldots, x_n] / (f_1(x_\bullet), \ldots, f_k(x_\bullet))$$

then a k-point is the same as an assignment $x_i \in k$ compatible with all the imposed relations. Similarly, if

$$k[\mathbf{W}] = k[y_1, \ldots, y_m] / (g_1(y_\bullet), \ldots, g_l(y_\bullet))$$

then a morphism $h : \mathbf{V} \to \mathbf{W}$ is the same as a ring map $k[\mathbf{W}] \to k[\mathbf{V}]$, i.e. collection of maps

$$y_i = h_i(x_1, \ldots, x_n) \quad i = 1 \ldots m$$

which is to be interpreted as: the point with coordinates $(x_1, \ldots, x_n)$ goes to the point with y-coordinates $(h_1(x_\bullet), \ldots, h_m(x_\bullet))$.

3.1.3. Remark. A field extension $k \subset l$ gives a natural map

$$\mathbf{V}(k) \to \mathbf{V}(l)$$

between the k-points and l-points of a variety.

For a k-variety $\mathbf{V}$ and a subring $R \subset k$, it doesn't make sense to speak of $\mathbf{V}(R)$ unless some structure of R-variety has been specified. In the case $\mathbb{Z} \subset \mathbb{Q}$, there is always some way to pick a $\mathbb{Z}$-structure and we will implicitly do so below when writing $\mathbf{V}(\mathbb{Z})$ for a $\mathbb{Q}$-variety $\mathbf{V}$.

3.1.4. Definition (Affine Algebraic Group). An affine algebraic group over k is an affine variety G over k, equipped with (algebraic) morphisms

$$m : G \times G \to G, \quad i : G \to G, \quad e \in G(k)$$

satisfying the axioms of multiplication, inverse element, and identity in a group.

3.1.5. Example.

(i) The additive group $\mathbb{G}_a$ is given by

$$k[\mathbb{G}_a] = k[t]$$

with group operation $m : k[t] \to k[t_1, t_2]$ given by $m(t) = t_1 + t_2$.

(ii) The multiplicative group $\mathbb{G}_m$ is given by

$$k[\mathbb{G}_m] = k[t, t^{-1}]$$

with group operation $m : k[t, t^{-1}] \to k[t_1, t_1^{-1}, t_2, t_2^{-1}]$ given by $m(t) = t_1 \cdot t_2$.

(iii) The general linear group $k[\mathbf{GL}_n] = k[x_{i,j}, \det^{-1}]$ where det is the determinant in the variables $x_{i,j}$ viewed as the entries of an $n \times n$ matrix.

3.1.6. Restriction of Scalars. The following construction, due to Weil, produces a broad range of examples. It is related to, but fundamentally different from, *extending* scalars by considering an algebraic group over k as one over l, for an extension $k \subset l$.

Suppose that $k \subset l$ is a field extension and $\mathbf{V}/l$ is an algebraic variety over l. The *Weil restriction of scalars* $\mathrm{Res}_k^l \mathbf{V}$, when it exists, is characterized by the condition that for any k-algebra R, its R-points satisfy:

$$\mathrm{Res}_k^l \mathbf{V}(R) = \mathbf{V}(R \otimes_k l)$$

In particular $\mathrm{Res}_k^l \mathbf{V}(k) = \mathbf{V}(l)$. For finite extensions of fields, the restriction of scalars group always exists (see Exercise 9.2.1). When $\mathbf{V}$ is an affine algebraic group, so is the restriction of scalars.

3.1.7. Example (Deligne Torus). Let $\mathbb{S} := \mathrm{Res}_{\mathbb{R}}^{\mathbb{C}} \mathbb{G}_m$ be the restriction of scalars of the multiplicative group from $\mathbb{C}$ to $\mathbb{R}$. It is an $\mathbb{R}$-algebraic group, with the property that $\mathbb{S}(\mathbb{R}) = \mathbb{G}_m(\mathbb{C}) = \mathbb{C}^{\times}$.

To compute it directly, let $t = a + b\sqrt{-1}, t^{-1} = c + d\sqrt{-1}$, so the relation $t \cdot t^{-1} = 1$ becomes:

$$(a + b\sqrt{-1})(c + d\sqrt{-1}) = 1$$

and so

$$ac - bd = 1$$
$$ad + bc = 0$$

One can check that $\mathbb{R}[a, b, c, d]/(ac - bd = 1, ad + bc = 0)$ is isomorphic to $\mathbb{R}[a, b, \frac{1}{a^2+b^2}]$ using the relation

$$(a + b\sqrt{-1})^{-1} = \frac{a - b\sqrt{-1}}{a^2 + b^2}$$

to read off c, d. Note that $\mathbb{S}$ can also be viewed as the matrix group

$$\left\{ \begin{bmatrix} a & b \\ -b & a \end{bmatrix} : a^2 + b^2 \neq 0 \right\}$$

with the group operation coming from matrix multiplication.

3.1.8. Example. Let $\mathbb{Q} = k \subset l = \mathbb{Q}(\alpha)/(\alpha^2 - 2)$ and consider the quadratic form over l:

$$Q = x_1^2 + \cdots + x_m^2 - \alpha(x_{m+1}^2 + \cdots + x_{m+n}^2)$$

Then $\mathbf{SO}(Q)$ is defined to be the l-linear group of all linear transformations of $\mathbb{A}^{m+n}$ that preserve Q. Let $\mathbf{G} := \mathrm{Res}_k^l \mathbf{SO}(Q)$ be the restriction of scalars group.

We have then than $\mathbf{G}(\mathbb{Q}) = \mathbf{SO}(Q)(\mathbb{Q}(\alpha))$ and (abusively) $\mathbf{G}(\mathbb{Z}) = \mathbf{SO}(Q)(\mathbb{Z}[\alpha])$. To compute the $\mathbb{R}$-points, recall that $\mathbf{G}(\mathbb{R}) = \mathbf{SO}(Q)(\mathbb{Q}(\alpha) \otimes_{\mathbb{Q}} \mathbb{R}$.

Since $\mathbb{Q}(\alpha) \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{R} \oplus \mathbb{R}$ with $\alpha \mapsto \sqrt{2} \oplus -\sqrt{2}$, it follows that

$$\mathbf{G}(\mathbb{R}) = \mathbf{SO}(Q_1)(\mathbb{R}) \times \mathbf{SO}(Q_2)(\mathbb{R}) \cong \mathbf{SO}_{m,n}(\mathbb{R}) \times \mathbf{SO}_{m+n}(\mathbb{R})$$

where

$$Q_1 = x_1^2 + \cdots + x_m^2 - \sqrt{2}(x_{m+1}^2 + \cdots + x_{m+n}^2)$$
$$Q_2 = x_1^2 + \cdots + x_m^2 + \sqrt{2}(x_{m+1}^2 + \cdots + x_{m+n}^2)$$

In particular $\mathbf{SO}(Q)(\mathbb{Z}[\sqrt{2}]) \to \mathbf{SO}(Q_1)(\mathbb{R}) \cong \mathbf{SO}_{m,n}(\mathbb{R})$ maps as a discrete subgroup, since it maps discretely into the product with $\mathbf{SO}(Q_2)$, and this second factor is compact.

3.2. Algebraic Actions of Algebraic Groups

Unlike the actions of interest in ergodic theory, the algebraic actions of algebraic groups have a simple orbit structure and dynamics.

3.2.1. Actions on varieties. Suppose that $\mathbf{X}$ is an (affine) algebraic variety and $\mathbf{G}$ is an affine algebraic group. An action of $\mathbf{G}$ on $\mathbf{X}$ is an algebraic map

$$\mathbf{G} \times \mathbf{X} \to \mathbf{X}$$

subject to the usual associativity conditions. A special case is a linear representation of $\mathbf{G}$ on $\mathbf{V}$, i.e. an action $\mathbf{G} \times \mathbf{V} \to \mathbf{V}$ fixing the origin and respecting the vector space axioms on $\mathbf{V}$.

3.2.2. Theorem (Borel–Serre, Countable separability of orbits)**.** *Suppose that k is a local field and $\mathbf{G} \curvearrowright \mathbf{X}$ is an algebraic action of an algebraic group on an algebraic manifold. Equip $\mathbf{G}(k)$ and $\mathbf{X}(k)$ with their natural topologies, and the quotient space $Q := \mathbf{X}(k)/\mathbf{G}(k)$ with the quotient topology.*

Then Q satisfies the T_0 separability axiom, i.e. for any two points $q_1, q_2 \in Q$, there exists an open set $U \subset Q$ containing one but not the other.

3.2.3. Example. Consider the action of $\mathbb{G}_m \curvearrowright \mathbb{A}^2$ given by $t \cdot (x_1, x_2) = (t \cdot x_1, t^{-1} x_2)$. The orbits are contained in the fibers of the map

$$\mathbb{A}^2 \to \mathbb{A}^1 \quad (x_1, x_2) \mapsto x_1 \cdot x_2$$

The quotient space Q also maps to $\mathbb{A}^1$, isomorphically outside the origin, and with three orbits over the origin, corresponding to the two axes and the origin. The space Q is the line with a double point at the origin, plus an even fatter point also at the origin whose neighborhood contains the two double points.

3.2.4. Example. Suppose that $\mathbf{H} \subset \mathbf{G}$ is an algebraic subgroup. Then the quotient $\mathbf{V} := \mathbf{G}/\mathbf{H}$ exists as an algebraic variety, and there is a natural injective map

$$\mathbf{G}(k)/\mathbf{H}(k) \hookrightarrow \mathbf{V}(k)$$

which need not be surjective. However, the natural action $\mathbf{G} \curvearrowright \mathbf{V}$ is such that $\mathbf{G}(k)$ acts with finitely many orbits on $\mathbf{V}(k)$ when k is a local field. This is a cohomological finiteness result due to Borel & Serre.

Take for example $\mathbf{G} = \mathbb{G}_m$ and $\mathbf{H} = \ker(M_2)$ where $M_2(t) = t^2$ is the doubling map $\mathbb{G}_m \to \mathbb{G}_m$. Since the doubling map is surjective (as a map of algebraic groups) we have $\mathbf{G}/\mathbf{H} \cong \mathbb{G}_m$. Now $\mathbf{G}(\mathbb{R}) = \mathbb{R}^\times$, $\mathbf{H}(\mathbb{R}) = \{\pm 1\}$ and the map $\mathbf{G}(\mathbb{R})/\mathbf{H}(\mathbb{R}) \to \mathbb{G}_m(\mathbb{R})$ is $t \mapsto t^2$ which has image the positive reals.

The useful tameness properties of algebraic actions are encapsulated in the following definition.

3.2.5. Definition (Tame action, [Zim84, 2.1.8-9]). A space X with σ-algebra $\mathcal{B}$ is called *countably separated* if there exist countably many measurable sets $A_i \in \mathcal{B}$ which separate points, i.e. for any two points there exists an A_i such that one of the points is in A_i and the other one isn't.

A Borel action $G \curvearrowright X$ is called a *tame action in the sense of Zimmer* if the quotient space X/G equipped with the quotient σ-algebra is countably separated.

Note that Zimmer calls actions as above *smooth*, not tame.

3.2.6. Proposition (Ergodic vs. Tame Actions).

(i) *Suppose $G \curvearrowright S$ is a tame action. Then any ergodic invariant probability measure is supported on a single orbit.*

(ii) *Suppose that $H \curvearrowright (X, \mu)$ is an ergodic action on a probability space, $G \curvearrowright S$ is a tame action, $\rho : H \to G$ is a group homomorphism and $f : X \to S$ is a measurable equivariant map, i.e.*

$$f(h \cdot x) = \rho(h) \cdot f(x) \text{ for } \mu - a.e. \ x.$$

Then there exists a single G-orbit $O \subset S$ such that the image of f is contained in O for μ-a.e. x.

Proof. Note that (i) implies (ii) since f defines a pushed-forward measure $f_*\mu$ which satisfies the assumptions in (i).

Suppose therefore that μ is a G-ergodic measure on S. Let A_i be the countable sequence of measurable sets which separate the points of $G\backslash S$ and let $p : S \to G\backslash S$ be the quotient map. By ergodicity of the G-action, the G-invariant sets $p^{-1}(A_i)$ have measure 0 or 1. Let I_1, I_0 be the indexes of those that have measure 1, resp. 0 and set

$$O := \left(\bigcap_{i \in I_1} p^{-1}(A_i) \right) \cap \left(\bigcap_{i \in I_0} S \setminus p^{-1}(A_i) \right)$$

be their intersection. Then $\mu(O) = 1$ by construction and it suffices to check that O consists of a single orbit. If not, there exist two disjoint orbits $O_1, O_2 \subset O$ and by the tameness assumption, there must exist an A_j such that $p^{-1}(A_j)$ separates them. Either $p^{-1}(A_j)$ or its complement must have measure 1 and this contradicts the construction of O. $\qquad \square$

3.2.7. Proposition. *Suppose that $G \curvearrowright X$ is a continuous action of a locally compact group on a locally compact Hausdorff space, such that*

every orbit is locally closed. Then the action is tame in the sense of Definition 3.2.5.

Proof. The tameness of the action will follow if we show that for any G-orbit $O = G \cdot x$, with orbit closure $\overline{O}$, the set $\overline{O} \setminus O$ is closed. For then, given two disjoint orbits O_1, O_2, either O_2 is contained in $\overline{O_1}$ in which case the closed G-invariant set $\overline{O_1} \setminus O_1$ provides the separation, or O_2 is not contained in $\overline{O_1}$, in which case $\overline{O_1}$ does it.

To check that $\overline{O} \setminus O$ is closed, note first that any $y \in X \setminus \overline{O}$ is contained in an open set disjoint from $\overline{O}$, since X is Hausdorff. So it suffices to find for every $y \in \mathcal{O}$ an open set containing it but disjoint from $\overline{O} \setminus O$. But since $\mathcal{O}$ is locally closed, let $U \ni y$ be a neighborhood such that $O \cap U = K \cap U$ for a closed set K. It then follows that $\overline{O} \cap U = K \cap U$, so $\overline{O} \setminus O$ is disjoint from U. $\qquad\square$

3.2.8. Theorem (Algebraic actions on measures are tame, Margulis, Zimmer). *Suppose that $G \curvearrowright X$ is an algebraic action over a local field k, i.e. both of G, X are affine algebraic and the action is also. Then the action of $G(k)$ on the space $\mathcal{M}^1(X(k))$ of probability measures on $X(k)$ has locally closed orbits, in particular it is tame.*

3.2.9. Theorem (Stabilizers of measures for algebraic actions).

3.3. Chevalley's theorems

3.3.1. Definition (Constructible set). Let X be a topological space. The *constructible sets* are the smallest collection of sets that contains the opens, and is closed under taking complements and finite unions.

It turns out (see Exercise 9.2.5) that constructible sets are finite unions of locally closed sets.

The following classical result is proved using elimination theory.

3.3.2. Theorem. *A polynomial map $f : \mathbb{C}^n \to \mathbb{C}^m$ takes, in the Zariski topology, constructible sets to constructible sets.*

A much stronger version is true, but one should be careful in using it over non-algebraically closed fields.

3.3.3. Theorem (Chevalley on Constructible Sets). *Let $f : \mathbf{X} \to \mathbf{Y}$ be a finitely presented morphism of schemes. Then the image of any constructible set is constructible.*

The example of $x \mapsto x^2$ shows that being a real point in *not* definable as a subset of $\mathbb{A}^1$.

3.3.4. Representations. Affine algebraic groups have different embeddings into matrix groups. By definition, a *representation* of G is a map $\rho : \mathbf{G} \to \mathbf{GL}(\mathbf{V})$ where $\mathbf{V}$ is a vector space and $\mathbf{GL}(\mathbf{V})$ is the group of invertible linear transformations of $\mathbf{V}$.

3.3.5. Proposition (Representations as Comodules)**.** *A linear representation of* $\mathbf{G}$ *on a vector space* $\mathbf{V}$ *is equivalent to any of the following:*

  (i) *An action* $\mathbf{G} \times \mathbf{V} \to \mathbf{V}$ *respecting the additive group structure on* $\mathbf{V}$ *and preserving* $0 \in \mathbf{V}$.
  (ii) *A map* $\Delta : k[\mathbf{V}] \to k[\mathbf{G}] \otimes k[\mathbf{V}]$ *subject to the axioms of a comodule over* $k[\mathbf{G}]$.

Proof. The axioms of a comodule a picked so that the above is true. FiXme: Fill in details. $\qquad \square$

A natural source of representations are the regular functions on G, i.e. $k[\mathbf{G}]$, but this space is typically infinite-dimensional. The next result shows that it is, in fact, a union of finite-dimensional subrepresentations.

3.3.6. Proposition (Finite-dimensional representations)**.**

  (i) *The multiplication* $m : \mathbf{G} \times \mathbf{G} \to \mathbf{G}$ *induces a comodule structure on* $k[\mathbf{G}]$ *over* $k[\mathbf{G}]$ *via the map on regular functions*

$$\Delta : k[\mathbf{G}] \to k[\mathbf{G}] \otimes k[\mathbf{G}]$$

  (ii) *Any element* $f \in k[\mathbf{G}]$ *sits inside a finite-dimensional subcomodule* $\mathbf{V} \subset k[\mathbf{G}]$ *and therefore there is a finite-dimensional representation*

$$\mathbf{G} \to \mathbf{GL}(\mathbf{V})$$

  (iii) *Any irreducible finite-dimensional representation of* $\mathbf{G}$ *arises as a subrepresentation in* $k[\mathbf{G}]$ *via the previous construction.*

Proof. Part (i) is true by the definition of group structure. For part (ii) fix a vector space basis $\{v_\alpha\}$ of $k[\mathbf{G}]$ (typically infinite). Write

$$\Delta(v) = \sum_i v_i' \otimes v_i''$$

with v_i', v_i'' in the basis. Set $\mathbf{V} := span\,\{v_i'\}$ and it suffices to check that $\Delta(V) \subset k[\mathbf{G}] \otimes V$.

For this, recall that the associativity of the group operation translates into $(\Delta \otimes \mathbf{1}) \circ \Delta = (\mathbf{1} \otimes \Delta) \circ \Delta$. Apply this identity to $\Delta(v)$ to find:

$$\sum_i \Delta(v_i') \otimes v_i'' = \sum_i v_i' \otimes \Delta(v_i'')$$

Since the v_i'' are distinct elements of a basis of $k[\mathbf{G}]$ it follows, by expressing in the same basis the right-hand side, that $\Delta(v_i')$ is in the linear span of $\sum_i v_i' \otimes h_i$ for some $h_i \in k[\mathbf{G}]$.

For (iii) let $v \in \mathbf{V}$ be a vector such that its $\mathbf{G}$-orbit spans $\mathbf{V}$. A basis of linear functions $\xi_i \in \mathbf{V}^\vee$ pull back to regular functions on $\mathbf{G}$, via $\xi_i \mapsto (g \mapsto \xi_i(g \cdot v))$. So the dual representation occurs in $k[\mathbf{G}]$, so $\mathbf{V}$ can be reconstructed as a representation from the procedure in part (ii). $\square$

3.3.7. Theorem (Chevalley Theorem on Subgroups). *For any algebraic subgroup $\mathbf{H} \subset \mathbf{G}$ of an affine algebraic group, there exists a linear representation $\mathbf{G} \to \mathbf{GL}(\mathbf{V})$ and a line $l \subset \mathbb{P}(\mathbf{V})$ such that*

$$\mathbf{H} = \{g \in \mathbf{G} \colon g \cdot l = l\}$$

are the defining equations of $\mathbf{H}$.

Proof. Let $I_\mathbf{H} \subset k[\mathbf{G}]$ be the ideal of functions determining $\mathbf{H}$. By the Noetherian property of $k[\mathbf{G}]$, there exists a finite-dimensional subspace $I_\mathbf{H}' \subset I_\mathbf{H}$ which generates the ideal. By Proposition 3.3.6 there exists $\mathbf{V}' \supset I_\mathbf{H}'$ such that $\mathbf{V}'$ is $\mathbf{G}$-invariant, hence gives a $\mathbf{G}$-representation. Let $\mathbf{V} := \Lambda^p \mathbf{V}'$ be the exterior power representation with $p = \dim I_\mathbf{H}'$, so that $\Lambda^p I_\mathbf{H}'$ is a line $l \subset \mathbb{P}(\mathbf{V})$. Then $\mathbf{H}$ can be defined as the subgroup of G fixing l. $\square$

3.3.8. Corollary (Existence of Quotients). *Given an algebraic subgroup $\mathbf{H} \subset \mathbf{G}$, there exists a variety $\mathbf{Q} := \mathbf{G}/\mathbf{H}$ equipped with a $\mathbf{G}$-action and a basepoint $q \in \mathbf{Q}(k)$ such that for any other $\mathbf{G}$-variety $\mathbf{Y}$ with a basepoint $y \in \mathbf{Y}(k)$ such that $\mathrm{Stab}_\mathbf{G}\, y \supset \mathbf{H}$, there is a unique morphism of $\mathbf{G}$-varieties*

$$\mathbf{Q} \to \mathbf{Y} \quad q \mapsto y.$$

Proof. Identify $\mathbf{Q}$ with the orbit $\mathbf{H} \cdot l \subset \mathbb{P}(V)$ provided by Theorem 3.3.7. $\square$

3.4. A bit of structure theory of algebraic groups

Fix a field k, with an algebraic closure $\overline{k} \subset k$, and a k-vector space $\mathbf{V}$.

3.4.1. Definition (Nilpotent, Unipotent, Semisimple). An element $g \in \mathbf{End}(\mathbf{V})$ is called:

- *nilpotent* if $g^N = 0$ for some $N \geq 0$.
- *unipotent* if $g - 1$ is nilpotent.
- *semisimple* if it can be diagonalized over $\overline{k}$.

3.4.2. Lemma (Jordan Decomposition). *Suppose* $\mathbf{G} \subset \mathbf{GL}(\mathbf{V})$ *is a* k-*algebraic group (and* k *is a perfect field, i.e.* $\bar{k} = \bar{k}^{sep}$).

(i) *If* $g \in \mathbf{End}(\mathbf{V})$ *is some element, then it has a canonical additive Jordan decomposition*

$$g = g_s + g_n$$

with g_s, g_n *semisimple, respectively nilpotent, and* $g_s g_n = g_n g_s$.

(ii) *If* $g \in \mathbf{G}(k)$, *then it has a canonical multiplicative Jordan decomposition*

$$g = g_s \cdot g_u$$

with g_s, g_u *semisimple, respectively unipotent, and* $g_s g_u = g_u g_s$. *Moreover* $s_m, u_m \in \mathbf{G}(k)$.

In both situations, there exists polynomials $p_s(t), p_u(t)$ *such that the semisimple, resp. unipotent parts, equal* $p_s(g)$, *resp.* $p_u(g)$.

Proof. For both parts, it suffices to check the claim assuming that $k = \bar{k}$ since the decompositions are canonical, therefore invariant under the Galois action.

Then there is a canonical direct-sum decomposition $\mathbf{V} = \oplus \mathbf{V}_\lambda$ into spaces where $(g - \lambda)^N$ acts trivially, for $N \gg 0$. On the spaces $\mathbf{V}_\lambda$, the action of $(g - \lambda)$ induces a filtration and acts nilpotently, giving the desired additive Jordan decomposition: $g = \lambda + (g - \lambda)$.

For the multiplicative decomposition, since $g \in \mathbf{G}(k)$ is invertible, in its additive decomposition $g = g_s + g_n$ the element g_s will be invertible. So $g = g_s(1 + g_s^{-1} g_n)$ provides the desired multiplicative decomposition.

To check that if $g = g_s \cdot g_u$ then $g_s, g_u \in \mathbf{G}(k)$, note from part (i) that each of g_s, g_u is a polynomial in g. So each of them preserves the ideal of functions defining $\mathbf{G}$, since g does. So each of them is in $\mathbf{G}$.   □

3.4.3. Example (Imperfect fields). Suppose that char $k = 2$ and k is imperfect, e.g. $k = \mathbb{F}_2((t))$, and let $a \in k \setminus k^2$, e.g $t \in \mathbb{F}_2((t))$. Then the matrix $\begin{bmatrix} 0 & 1 \\ a & 0 \end{bmatrix}$ has semisimple part $\begin{bmatrix} \sqrt{a} & 0 \\ 0 & \sqrt{a} \end{bmatrix}$ but this matrix is not defined over k but only over a quadratic inseparable extension.

3.4.4. Proposition. *If* $g \in \mathbf{G}(k)$ *is semisimple (resp. unipotent) then for any linear representation* $\rho : \mathbf{G} \to \mathbf{GL}(\mathbf{V})$, $\rho(g)$ *is also semisimple (resp. unipotent).*

Proof. It suffices to work over an algebraically closed field. One checks that the Zariski closure of $g, g^2, g^3, \dots$ is, up to finite index, isomorphic to a product of $\mathbb{G}_m$ (for g semisimple) or $\mathbb{G}_a$ (for g unipotent). Then

one checks that there are no algebraic homomorphisms $\mathbb{G}_m \to \mathbb{G}_a$ or $\mathbb{G}_a \to \mathbb{G}_m$. $\qquad\square$

3.4.5. Corollary. *An element $g \in \mathbf{G}(k)$ is semisimple or unipotent independently of the linear representation, and has a canonical Jordan decomposition $g = g_s \cdot g_u$ with $g_s, g_u \in \mathbf{G}(k)$.*

3.4.6. Definition (Tori)**.**

   (i) A *torus T* is an affine k-algebraic group which, after an extension of scalars to an algebraic closure $\overline{k}$, is isomorphic to several copies of the multiplicative group:

$$T \times_k \overline{k} \cong (\mathbb{G}_m)^N$$

   (ii) A torus is *l-split* if it is isomorphic to $(\mathbb{G}_m)^N$ after extending scalars to $l \supset k$. When the torus is split over the base field, it will be called just *split.*
   (iii) The *l-rank* of a k algebraic group $\mathbf{G}$ is the maximal dimension of an *l*-split subtorus in G.
   (iv) A *character* of a $\mathbf{G}$-algebraic group is an algebraic homomorphism

$$\mathbf{G} \to \mathbb{G}_m$$

   (v) A *cocharacter* of a $\mathbf{G}$-algebraic group is an algebraic homomorphism

$$\mathbb{G}_m \to \mathbf{G}$$

The adjective *geometric* applied to a concept means that the concept is considered over an algebraic closure.

The Deligne torus $\mathbb{S}$ (Example 3.1.7) is an example of a torus which is not $\mathbb{R}$-split. It's ranks are: $\mathrm{rk}_{\mathbb{R}} \mathbb{S} = 1, \mathrm{rk}_{\mathbb{C}} \mathbb{S} = 2$.

3.4.7. Example (Quaternion Algebras)**.** Over a field k, for fixed $a, b \in k^\times$ consider the non-commutative algebra $\mathbb{H}_{a,b}$ generated by the variables i, j subject to the relations:

$$i^2 = a \quad j^2 = b \quad ij = -ji$$

It is equipped with an anti-involution $x \mapsto \overline{x}$ determined by:

$$\overline{y} \quad \overline{i} = -i, \quad \overline{j} = -j \quad \overline{ij} = ji$$

so that it satisfies $\overline{x \cdot y} = \overline{y} \cdot \overline{x}$. The norm $N : \mathbb{H}_{a,b} \to k$ is defined by $N(x) := x \cdot \overline{x}$.

For example, if $k = \mathbb{R}$ then $\mathbb{H}_{-1,-1}$ is isomorphic to the usual quaternions and $N(x)$ is the euclidean norm of the corresponding vector. On the other hand, for any $t \in \mathbb{R}$, the algebra $\mathbb{H}^{1,t}$ is isomorphic to

$\mathbf{Mat}_{2\times 2}(\mathbb{R})$, the algebra of 2×2 matrices, with norm equal to the determinant. Indeed, take $i = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ and $j = \begin{bmatrix} 0 & 1 \\ t & 0 \end{bmatrix}$ to see that the relations are satisfied.

Returning to a general field k, let $\mathbf{G} := \mathbb{H}^1_{a,b}$ denote the group of unit norm quaternions. This is a k-algebraic group determined by the equation $x \cdot \bar{x} = 1$. Take now positive integers a, b and note that $\mathbf{G}(\mathbb{R}) \cong \mathbf{SL}_2(\mathbb{R})$. Moreover $\mathbf{G}(\mathbb{Z}) \subset \mathbf{G}(\mathbb{R})$ will be a lattice, by a general theorem of Borel & Harish-Chandra (Theorem 3.4.9). If $\mathbb{H}^1_{a,b}$ is not isomorphic over $\mathbb{Q}$ to 2×2 matrices, then $\mathbf{G}$ will have $\mathbb{Q}$-rank 0 but $\mathbb{R}$-rank 1.

3.4.8. More terminology for algebraic groups. The subject of algebraic groups is vast. Below are some of the important concepts that will be used later. A k-algebraic group $\mathbf{G}$ is:

- *unipotent* if all its elements are unipotent.
- *reductive* if any linear representation is reductive[3], i.e. any invariant subspace has a complement.
- *semisimple* if it is reductive and with finite center.

For example $\mathbf{GL}_n$ is reductive, but not semisimple, whereas $\mathbf{SL}_n$ is both.

- *k-isotropic* if it contains a k-split subtorus.
- *k-anisotropic* if it isn't isotropic.

These last two notions depend heavily on the field k for which the question is asked.

3.4.9. Theorem (Borel–Harish-Chandra, Godement)**.** *Let $\mathbf{G}$ be a $\mathbb{Q}$-algebraic group.*

(i) *If $\mathbf{G}$ doesn't have any nontrivial $\mathbb{Q}$-character, then $\mathbf{G}(\mathbb{Z})$ is a lattice in $\mathbf{G}(\mathbb{R})$.*

(ii) *If $\mathbf{G}$ doesn't have any nontrivial $\mathbb{Q}$-cocharacter, then $\mathbf{G}(\mathbb{Z})$ is a* cocompact *lattice in $\mathbf{G}(\mathbb{R})$.*

3.4.10. Remark.

(i) Since Theorem 3.4.9 applies to all $\mathbb{Q}$-algebraic groups, it can be applied to unipotent ones, semisimple, or a combination.

(ii) Theorem 3.4.9 contains the Dirichlet unit theorem. Namely, let $K \supset \mathbb{Q}$ be a finite extension and $\mathbf{G} := \mathrm{Res}^K_\mathbb{Q} \mathbb{G}_m$ be the restriction of scalars algebraic group. This groups does have a non-trivial character $N : \mathbf{G} \to \mathbb{G}_m$ given by the norm, but

———————
[3]One usually says "semisimple representation" but it can be confusing in this context

the kernel $\mathbf{G}^1 := \ker N$ has no characters. The theorem then implies that $\mathbf{G}^1(\mathbb{Z}) = \mathcal{O}_K^{\times}$ (the units in the ring of integers $\mathcal{O}_K$) are a lattice in $\mathbf{G}^1(\mathbb{R}) \cong (\mathbb{G}_m(\mathbb{R})^r \times \mathbb{G}_m(\mathbb{C})^s)^1$, where $(-)^1$ denotes the unit-norm elements.

3.5. Lie Algebras

Associated to any algebraic group $\mathbf{G}$ there is the tangent space at the identity $\mathfrak{g} := T_e\mathbf{G}$ which is a k-vector space equipped with the structure of a Lie algebra, i.e. it has a bracket $[,] : \Lambda^2\mathfrak{g} \to \mathfrak{g}$ satisfying the Jacobi identity:

$$[X, [Y, Z]] = [[X, Y], Z] + [Y, [X, Z]] \quad \forall X, Y, Z \in \mathfrak{g}.$$

The Lie algebra can be viewed as a module over itself, and it gives the *adjoint action*:

$$\mathrm{ad}_X(Y) := [X, Y]$$

so that the Jacobi identity can be rewritten as

$$\mathrm{ad}_X([Y, Z]) = [\mathrm{ad}_X(Y), Z] + [Y, \mathrm{ad}_X(Z)].$$

The map $\mathrm{ad} : \mathfrak{g} \to \mathbf{End}(\mathfrak{g})$ is called the adjoint representation.

3.5.1. Definition (Center, Extension)**.** Let $\mathfrak{g}$ be a Lie algebra. Its *center* $Z(\mathfrak{g})$ is the kernel of $\mathrm{ad} : \mathfrak{g} \to \mathbf{End}(\mathfrak{g})$, i.e. the subspace of elements that commute with all other elements.

An *extension* of Lie algebras is a short exact sequence

$$0 \to \mathfrak{g}' \to \mathfrak{g} \to \mathfrak{g}'' \to 0$$

with each map respecting the Lie bracket. The extension is *central* if $\mathfrak{g}' \subset Z(\mathfrak{g})$.

3.5.2. Definition (Abelian, Nilpotent, Solvable)**.** The Lie algebra $\mathfrak{g}$ is

- *abelian* if the Lie bracket $[,]$ is identically 0.
- *nilpotent* if it is obtained by successive central extension of abelian Lie algebras.
- *solvable* if it is obtained by successive extensions of abelian Lie algebras.

3.5.3. Definition (Killing form)**.** The symmetric bilinear form

$$B(X, Y) := \mathrm{tr}(\mathrm{ad}_X \circ \mathrm{ad}_Y)$$

is called the Killing form.

It is invariant under the action of $\mathfrak{g}$:

$$B(\mathrm{ad}_X Y, Z) + B(Y, \mathrm{ad}_X Z) = 0$$

Its radical $\mathfrak{r} \subset \mathfrak{g}$ is an ideal (i.e. $[\mathfrak{g}, \mathfrak{r}] \subset \mathfrak{r}$) and is the maximal solvable ideal.

3.5.4. Theorem (Definition of Semisimple)**.** *The following are equivalent:*

  (i) *There is no non-trivial abelian ideal in $\mathfrak{g}$.*
  (ii) *The Killing form is non-degenerate.*
  (iii) *There is a decomposition $\mathfrak{g} = \oplus_i \mathfrak{g}_i$ with each $\mathfrak{g}_i$ a simple ideal.*

A Lie algebra $\mathfrak{g}$ satisfying the above theorem is called *semisimple*. If $\mathfrak{g}$ has no non-trivial ideals, it is called *simple*.

For the rest of this section, let $\mathfrak{g}$ be a *simple* Lie algebra.

3.5.5. Definition (Cartan subalgebra)**.** A *Cartan subalgebra* $\mathfrak{h} \subset \mathfrak{g}$ is any maximal abelian subalgebra consisting of semisimple elements.

3.5.6. The case $k = \overline{k}$. Suppose that k is algebraically closed. Then an elegant classification of simple Lie algebras is available. Namely, let $\mathfrak{h} \subset \mathfrak{g}$ be a Cartan subalgebra. For any $\alpha \in \mathfrak{h}^\vee$ define

$$\mathfrak{g}_\alpha := \{X \colon \mathrm{ad}_H(X) = \alpha(H) \cdot H, \forall H \in \mathfrak{h}\}$$

The set $\Delta := \{\alpha \in \mathfrak{h}^\vee \setminus 0 \colon \mathfrak{g}_\alpha \neq 0\}$ is called the set of *roots* of $\mathfrak{g}$. Some key properties are listed below:

  • We have $[\mathfrak{g}_\alpha, \mathfrak{g}_\beta] \subset \mathfrak{g}_{\alpha+\beta}$.
  • We have $\dim \mathfrak{g}_\alpha = 1, \forall \alpha \in \Delta$.
  • We have $\Delta = -\Delta$ and for any $\alpha \in \Delta$, the only other root proportional to α is $-\alpha$.
  • The Killing form B is non-degenerate on $\mathfrak{h}$ and so induces an isomorphism $\mathfrak{h}^\vee \to \mathfrak{h}$. For $\alpha \in \Delta \subset \mathfrak{h}^\vee$ let $H_\alpha \in \mathfrak{h}^\vee$ be the corresponding element. Then

    $$[X_\alpha, X_{-\alpha}] = B(X_\alpha, X_{-\alpha})H_\alpha, \forall X_{\pm\alpha} \in \mathfrak{g}_{\pm\alpha}$$

    so $X_\alpha, X_{-\alpha}, H_\alpha$ span a Lie subalgebra isomorphic to $\mathfrak{sl}_2 \subset \mathfrak{g}$.

3.5.7. Real Lie Algebras. Over $\mathbb{R}$ the classification of Lie algebras is not as simple as over $\mathbb{C}$, yet it is manageable. Given a real Lie algebra $\mathfrak{g}$, one first complexifies it to $\mathfrak{g}_\mathbb{C} := \mathfrak{g} \otimes_\mathbb{R} \mathbb{C}$ and notes that $\mathfrak{g}_\mathbb{C}$ has an anti-involution coming from complex conjugation.

3.5.8. $\mathfrak{sl}_2$. The Lie algebra with generators H, E, F and relations

$$[H, E] = 2E \quad [H, F] = 2F \quad [E, F] = H$$

is also the Lie algebra of traceless 2×2 matrices

$$\mathfrak{sl}_2 = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} : a + d = 0 \right\}$$

with the identifications

$$H = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad E = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad F = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Any triple of elements H, E, F in a Lie algebra $\mathfrak{g}$ satisfying the above commutation relations is called an $\mathfrak{sl}_2$-*triple*.

3.5.9. Theorem (Jacobson–Morozov). *Any nilpotent element $E \in \mathfrak{g}$ in a semisimple Lie algebra is part of some $\mathfrak{sl}_2$-triple H, E, F.*

Any two such triples are conjugate under the action of the centralizer $Z(E)$ in the adjoint group of the Lie algebra.

Recall that the adjoint group of a Lie algebra is $\mathrm{Ad}(\mathfrak{g}) \subset \mathrm{Aut}(\mathfrak{g})$ is a connected subgroup with Lie algebra naturally identified with $\mathrm{ad}(\mathfrak{g})$. Here $\mathrm{ad}(\mathfrak{g})$ is the image of the Lie algebra acting on itself in the adjoint representation.

3.5.10. $\mathfrak{sl}_3$. Consider now

$$\mathfrak{sl}_3 := \{M \in \mathbf{Mat}_{3 \times 3} \colon \operatorname{tr} M = 0\}$$

and let

$$\mathfrak{h} := \left\{ t := \begin{bmatrix} t_1 & 0 & 0 \\ 0 & t_2 & 0 \\ 0 & 0 & t_3 \end{bmatrix} : t_1 + t_2 + t_3 = 0 \right\}$$

be a Cartan subalgebra. Define the linear functionals $\alpha_1, \alpha_2 \in \mathfrak{h}^\vee$ by

$$\alpha_1(t_1, t_2, t_3) = t_1 - t_2 \quad \alpha_2(t_1, t_2, t_3) = t_2 - t_3$$

Then the roots are

$$\Delta = \{\pm\alpha_1, \pm\alpha_2, \pm(\alpha_1 + \alpha_2)\}$$

The coroots (identified via the Killing form) are then:

$$H_{\alpha_1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad H_{\alpha_2} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

so that $\mathfrak{h} := span \langle H_{\alpha_1}, H_{\alpha_2} \rangle$.

A fundamental fact used frequently in establishing rigidity in higher rank dynamics is that there are elements of $\mathfrak{h}$ which act trivially on certain root spaces, and moreover there are many such. For example, $H_{\alpha_1} + 2H_{\alpha_2}$ acts trivially on the root space $\mathfrak{g}_{\alpha_1}$.

3.6. Real semisimple Lie algebras and groups

Fix a real semisimple Lie algebra $\mathfrak{g}$ and let $\mathfrak{g}_{\mathbb{C}} := \mathfrak{g} \otimes_{\mathbb{R}} \mathbb{C}$ be its complexification. For most concepts defined below, any two can be conjugated into one another by the action of $\mathrm{Aut}^{\circ}(\mathfrak{g})$, the connected component of the identity of Lie algebra automorphisms inside $\mathbf{GL}(\mathfrak{g})$ (also known as the adjoint group).

3.6.1. Definition (Cartan[4] involution and decomposition)**.** A Lie algebra automorphism $\sigma : \mathfrak{g} \to \mathfrak{g}$ is a *Cartan involution* if $\sigma^2 = 1$ and the modified Killing form $B_{\sigma}(X, Y) := -B(\sigma X, Y)$ is positive-definite.

The resulting *Cartan decomposition* is

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{q}$$

where $\sigma|_{\mathfrak{k}} = \mathbf{1}$ and $\sigma|_{\mathfrak{q}} = -\mathbf{1}$.

3.6.2. Example.
   (i) Take $\mathfrak{g} = \mathfrak{sl}_n \mathbb{R}$ and Cartan involution $X \mapsto -X^t$. Then $\mathfrak{k} = \mathfrak{so}_n \mathbb{R}$ and $\mathfrak{q}$ is the set of symmetric matrices.
   (ii) Take $\mathfrak{g} = \mathfrak{so}_n \mathbb{R}$ with trivial Cartan involution. It is a fact that the Killing form is negative-definite if and only if $\mathfrak{g}$ comes from a compact Lie group.
   (iii) Any semisimple Lie algebra $\mathfrak{g}$ admits some Cartan involution σ. Moreover it admits an embedding $\mathfrak{g} \hookrightarrow \mathfrak{sl}_N \mathbb{R}$ such that σ agrees with $X \mapsto -X^t$.

The definition of Cartan subalgebra in Definition 3.5.5 extends to the real case. The following properties are useful:
   (i) The complexification $\mathfrak{h}$ is a Cartan subalgebra in $\mathfrak{g}_{\mathbb{C}}$.
   (ii) One can choose a Cartan subalgebra $\mathfrak{h}$ and Cartan involution σ such that $\sigma(\mathfrak{h}) = \mathfrak{h}$.
   (iii) In the compact case ($\mathfrak{g}$ has negative-definite Killing form) there is a unique conjugacy class of Cartan subalgebra. For a general semisimple Lie algebra, there are only finitely many conjugacy classes of Cartan subalgebras.
   (iv) In $\mathfrak{sl}_2 \mathbb{R}$, the two classes are $\mathfrak{so}_2(\mathbb{R})$ (a compact Cartan) and $\mathfrak{a}$, the diagonal matrices (a non-compact Cartan).

3.6.3. Definition (Split Cartan subalgebra)**.** A *split Cartan subalgebra* is a an abelian subalgebra $\mathfrak{a} \subset \mathfrak{g}$ consisting of hyperbolic elements, and maximal with this property.

Some properties of split Cartan subalgebras:

---

[4]In Lie theory, "Cartan" most often refers to Élie Cartan

(i) Any split Cartan subalgebra can be conjugated to be in $\mathfrak{q}$ (relative to a fixed Cartan involution σ and Cartan decomposition $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{q}$).

(ii) Any two split Cartan subalgebras in $\mathfrak{q}$ can be conjugated by $K \subset \mathrm{Aut}^\circ(\mathfrak{g})$, which is a maximal compact in the adjoint group.

From now on, consider only real semisimple Lie algebras that have a non-trivial split Cartan, i.e. are not compact. The dimension of a split Cartan subalgebra is called the *real rank* of the Lie algebra.

3.6.4. Restricted Root Systems. Once a split Cartan subalgebra $\mathfrak{a} \subset \mathfrak{g}$ is fixed, its adjoint action decomposes the Lie algebra into:

$$\mathfrak{g} = \mathfrak{l} \oplus \bigoplus_{\alpha \in \Sigma} \mathfrak{g}_\alpha$$

where $\alpha \in \mathfrak{a}^\vee$ are the *restricted roots* and $\mathfrak{l}$ is the centralizer of $\mathfrak{a}$. Note that unlike the case of complex Lie algebras, the centralizer of $\mathfrak{a}$ can be larger than $\mathfrak{a}$ itself. It will decompose, necessarily, as $\mathfrak{l} = \mathfrak{a} \oplus (\mathfrak{l} \cap \mathfrak{k})$, i.e. the elements that commute with $\mathfrak{a}$ and are not in $\mathfrak{a}$ already must be in the compact part.

Inside the roots Σ fix a choice of *positive roots* $\Sigma^+ \subset \Sigma$, such that $\Sigma = \Sigma^+ \coprod -\Sigma^+$. Let $\Pi \subset \Sigma$ be the *simple roots* (i.e minimal for the ordering of roots picked); they will give a basis of $\mathfrak{a}^\vee$.

The *Weyl chamber* $\mathfrak{a}^+ \subset \mathfrak{a}$ is defined by

$$\mathfrak{a}^+ = \left\{ a \in \mathfrak{a} \colon \alpha(a) \geq 0, \, \forall \alpha \in \Sigma^+ \right\}$$

It suffices in fact to take impose the inequalities $\alpha(a) \geq 0$ over the simple roots, i.e. $\alpha \in \Pi$.

3.6.5. Parabolic subalgebras. The *minimal parabolic subalgebras* are

$$\mathfrak{u}^+ := \mathfrak{l} \oplus \bigoplus_{\alpha \in \Sigma^+} \mathfrak{g}_\alpha$$

$$\mathfrak{u}^- := \mathfrak{l} \oplus \bigoplus_{\alpha \in \Sigma^-} \mathfrak{g}_\alpha$$

and they are related by the Cartan involution: $\mathfrak{u}^- = \sigma(\mathfrak{u}^+)$ since $\sigma(\mathfrak{g}_\alpha) = \mathfrak{g}_{-\alpha}$.

More generally, let $\theta \subset \Pi$ be a subset of the simple roots. Associated to it are:

- The linear space generated by θ, $span(\theta) \subset \mathfrak{a}^\vee$
- The roots $\Sigma_\theta := \Sigma \cap span(\theta)$ and associated decomposition

$$\Sigma_\theta = \Sigma_\theta^+ \coprod \Sigma_\theta^-$$

- The associated unipotent subalgebra:

$$\mathfrak{u}_\theta^+ := \bigoplus_{\alpha \in \Sigma^+ \setminus \Sigma_\theta^+} \mathfrak{g}_\alpha$$

and its opposite $\mathfrak{u}_\theta^- = \sigma(\mathfrak{u}_\theta^+)$.
- The associated reductive subalgebra:

$$\mathfrak{l}_\theta := \mathfrak{l} \oplus \bigoplus_{\alpha \in \Sigma_\theta} \mathfrak{g}_\alpha$$

- The associated parabolic subalgebra:

$$\mathfrak{p}_\theta := \mathfrak{l}_\theta \oplus \mathfrak{u}_\theta^+$$

3.6.6. Passing to Lie groups. From now on, let G be a finite center, connected, semisimple real Lie group with Lie algebra $\mathfrak{g}$. All of the above notions for Lie algebras extend to the Lie group and give the corresponding Lie subgroups.

The choice of a Cartan involution $\sigma : G \to G$ is equivalent to choosing a maximal compact subgroup $K \subset G$, constructed as the fixed point of the involution. It is a useful fact that the inclusion $K \hookrightarrow G$ is a homotopy equivalence (including the connected components!).

3.6.7. Cartan, or Polar, or KAK Decomposition. The spectral theorem for symmetric matrices implies that any $M \in \mathbf{GL}_n(\mathbb{R})$ can be written as $M = k_1 a k_2$ with $k_i \in O_n(\mathbb{R})$ and a diagonal, with diagonal entries satisfying $a_1 \geq \cdots \geq a_n$. This generalizes to any semisimple Lie group G to give that any $g \in G$ can be written as

$$g = k_1 a k_2 \text{ with } k_i \in K \text{ and } a \in A^+ = \exp(\mathfrak{a}^+)$$

where $\mathfrak{a}^+$ is the positive Weyl chamber. The decomposition is almost unique, the ambiguity being up to elements in K that are normalized by a.

3.6.8. Iwasawa, or Gram–Schmidt, or KAN decomposition. At the level of Lie algebras, one also has the *Iwasawa decomposition*:

$$\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{a} \oplus \mathfrak{u}^+$$

also known as the KAN decomposition (as $\mathfrak{u}^+$ is often denoted $\mathfrak{n}$). It can be exponentiated to show that one has the decomposition:

$$G \cong K \times A \times N \quad \text{(note it is } A \text{ not } A^+\text{)}$$

For $\mathbf{GL}_n \mathbb{R}$ this is just the Gram–Schmidt process.

3.6.9. The associated symmetric space. The quotient G/K has a natural G-invariant Riemannian metric, by identifying the tangent space at the identity coset with $\mathfrak{q}$ and using the Killing form $-B(-,-)$ as the metric. The curvature tensor of the metric is

$$R(X,Y)Z = -\frac{1}{4}[[X,Y],Z]$$

up to some normalization factors. This gives the sectional curvature to be:

$$K(X,Y) = -B(R(X,Y)X,Y) = -B([[X,Y],X],Y)$$
$$= B([X,Y],[X,Y]) \leq 0$$

Note that the sectional curvature vanishes on abelian subspaces.

The quotient G/K is a *symmetric space* because the Cartan involution σ acts on G/K isometrically, fixing the identity coset and acting as $-\mathbf{1}$ on its tangent space.

The Iwasawa decomposition $G = KAN$ gives coordinates on the symmetric space, and the polar decomposition $G = KA^+K$ gives a way to understand the rough geometry at infinity in terms of the positive Weyl chamber $\mathfrak{a}^+$.

3.6.10. The flag manifolds. To each parabolic subalgebra $\mathfrak{p}_\theta$ there is an associated parabolic subgroup $P_\theta \subset G$. The quotients $\mathcal{F}_\theta := G/P_\theta$ are compact homogeneous spaces called flag manifolds. It suffices to check compactness for the minimal parabolic $P_\emptyset$, but it follows from the KAN decomposition plus the fact that $AN \subset P_\emptyset$ and so in particular K acts transitively on $\mathcal{F}_\theta$.

3.6.11. Example (Symplectic Lie algebras)**.** Let $\mathfrak{sp}_{2g}$ be the symplectic algebra of $2g \times 2g$ matrices preserving the symplectic form $J = \begin{bmatrix} 0 & \mathbf{1}_g \\ -\mathbf{1}_g & 0 \end{bmatrix}$. Writing explicitly:

$$\mathfrak{sp}_{2g} = \left\{ X = \begin{bmatrix} A & B \\ C & D \end{bmatrix} : X^t J + JX = 0 \right\}$$

gives the conditions:

$$\mathfrak{sp}_{2g} = \left\{ X = \begin{bmatrix} A & B \\ C & D \end{bmatrix} : A = -D^t, B = B^t, C = C^t \right\}.$$

The Cartan involution is $X \mapsto -X^t$ so that the maximal compact subalgebra is

$$\mathfrak{u}_n = \left\{ X = \begin{bmatrix} A & B \\ -B & -A^t \end{bmatrix} : B = B^t, A = -A^t \right\}$$

which is the Lie algebra of the unitary group, via $X := A + \sqrt{-1}B$ with $X + \overline{X}^t = 0$.

A split Cartan is $\mathfrak{a}$ consisting of diagonal matrices:

$$\mathfrak{a} = \left\{ \begin{bmatrix} A & 0 \\ 0 & -A \end{bmatrix} : A = \begin{bmatrix} x_1 & & \\ & \ddots & \\ & & x_g \end{bmatrix} \right\}$$

The roots will be $\pm(x_i - x_j)$ (occurring in the A-block) and $\pm 2x_i, \pm(x_i + x_j)$ (occurring in the B-block) with $i < j$.

As simple roots one can take:

$$\alpha_1 := x_1 - x_2 \quad \alpha_2 = x_2 - x_3 \quad \ldots \quad \alpha_{g-1} = x_{g-1} - x_g$$
$$\alpha_g = 2x_g$$

where note that there is one distinguished "longest root" α_g. To get the other (positive) roots from the simple ones, note that:

$$x_i - x_j = \alpha_i + \cdots + \alpha_{j-1} \quad 2x_i = 2(\alpha_i + \cdots + \alpha_{g-1}) + \alpha_g \text{ etc.}$$

The positive Weyl chamber is

$$\mathfrak{a}^+ = \left\{ \begin{bmatrix} A & 0 \\ 0 & -A \end{bmatrix} : A = \begin{bmatrix} x_1 & & \\ & \ddots & \\ & & x_g \end{bmatrix} \quad x_1 \geq \cdots \geq x_g \geq 0 \right\}$$

For parabolic subalgebras, let $G := \mathbf{Sp}_{2g}(\mathbb{R})$.

The minimal parabolic $P_\emptyset$ corresponds to the empty set and $G/P_\emptyset$ parametrizes full flags in $\mathbb{R}^{2g}$ which are self-dual under the symplectic form.

For the choice $\theta := \{\alpha_1, \ldots, \alpha_{g-1}\}$ (where all roots but the longest one are included) the parabolic P_θ stabilizes a Lagrangian subspace and so G/P_θ is the Lagrangian Grassmannian.

For the choice $\theta := \{\alpha_2, \ldots, \alpha_g\}$ the parabolic P_θ stabilizes a line in $\mathbb{R}^{2g}$ (and its symplectic orthogonal) so G/P_θ is the projectivization of $\mathbb{R}^{2g}$.

4. Unitary Representations

4.1. Definitions

Let G be a topological group. Whereas for algebraic groups the basic source of functions on G are the algebraic ones, for a topological group many functions come from unitary representations via their matrix coefficients.

4.1.1. Definition (Unitary representation). A *unitary representation* of G on a Hilbert space V is a map $\pi : G \to \mathcal{U}(V)$ to the unitary operators on V such that the map

$$G \times V \to V$$

is continuous. Equivalently, for any $h \in V$ the map $G \to V, g \mapsto g \cdot h$ is continuous.

For any $v, w \in V$ the function $c_{v,w} : G \to \mathbb{R}$ defined by

$$c_{v,w}(g) := \langle gv, w \rangle$$

is called a *matrix coefficient* of the representation.

Fix a unitary representation $\pi : G \to \mathcal{U}(V)$.

4.1.2. Lemma (Identifying Fixed Vectors). *Let $v \in V$ be a vector. Then its stabilizer $\mathrm{Stab}_v G$ can be identified with*

$$\mathrm{Stab}_v G = \left\{ g \colon c_{v,v} = \|v\|^2 \right\}$$

Moreover the matrix coefficient satisfies $|c_{v,v}| \le \|v\|^2$ pointwise.

Proof. That pointwise inequality $|c_{v,v}| \le \|v\|^2$ follows by Cauchy–Schwarz. It also implies that if equality holds, then $gv = cv$ for some $c \in \mathbb{C}$ with $|c| = 1$. Moreover

$$(4.1.3) \quad \|gv - v\|^2 = 2 \cdot (\|v\|^2 - \mathrm{Re}\,\langle gv, v \rangle) = 2(\|v\|^2 - \mathrm{Re}\,c_{v,v}(g))$$

and the equality case of Cauchy–Schwarz implies the claim. $\qquad\square$

4.1.4. Lemma (Mautner). *Suppose that $\pi : G \to \mathcal{U}(V)$ is a unitary representation of a topological group. Suppose that $v \in V$ and $s_n, s'_n \in G$ are stabilizing v, and moreover there exists a sequence $g_n \to g$ such that*

$$\lim s_n g_n s'_n = \mathbf{1} \quad \text{in } G$$

Then g also stabilizes v.

Proof. By [Lemma 4.1.2](#), the stabilizer of v can be identified with the set of $h \in G$ such that $c_{v,v}(h) = \|v\|^2$. One can then take the limit in

$$c_{v,v}(g) = \lim_n c_{v,v}(g_n) = \lim_n c_{v,v}(s_n g_n s'_n) = \|v\|^2$$

since $c_{v,v}$ is a continuous function. $\qquad\square$

4.1.5. Definition (Positive-definite functions). A continuous function $c : G \to \mathbb{C}$ is called *positive-definite* if for any finite choice of $g_i \in G$ and $w_i \in \mathbb{C}$ we have

$$\sum_{i,j} c(g_i^{-1} g_j) \overline{w_i} w_j \ge 0$$

For example, a matrix coefficient $c_{v,v}(g) := \langle gv, v \rangle$ is a positive-definite function since

$$\sum_{i,j} c_{v,v}(g_i^{-1}g_j)w_j\overline{w_i} = \sum_{i,j} w_j\overline{w_i}\left\langle g_i^{-1}g_jv, v \right\rangle$$

$$= \sum_{i,j} \langle w_jg_jv, w_ig_iv \rangle$$

$$= \left\| \sum_i w_jg_jv \right\|^2 \geq 0$$

4.1.6. Proposition (Any positive-definite function is a matrix coefficient). *Suppose that $c : G \to \mathbb{C}$ is a continuous positive-definite function. Then there exists a unitary representation of G on a Hilbert space V and vector $v \in V$ such that $c = c_{v,v}$. Moreover, the representation and the vector are unique up to isomorphism.*

Proof. For every $g \in G$ define the continuous function $c_g(h) = c(g^{-1}h)$. On the linear span of all such continuous functions in $C(G)$, define the inner product, for

$$\psi = \sum_j v_j c_{g_j}$$

$$\phi = \sum_i w_k c_{g_k}$$

$$\langle \psi, \phi \rangle = \sum_{i,j} v_j \cdot \overline{w_k} \cdot c(g_j^{-1}g_k)$$

$$= \sum_k \overline{w_k} \sum_j v_j c_{g_j}(g_k) = \sum_k \overline{w_k} \cdot \psi(g_k)$$

where the last expression shows that the definition is independent of the representation of ψ as a linear combination of c_g's. A similar computation shows independence of the representation of ϕ, using that $c(g^{-1}) = \overline{c(g)}$ for a positive-definite function.

Since c is a positive-definite function, the inner product $\langle -, - \rangle_c$ just defined is positive-definite so let V be completion of the subspace of continuous function constructed using the inner product. It gives the required representation, with $c_1 \in V$ the required vector. $\square$

4.1.7. Induced Representations. Let $\Gamma \subset G$ be a subgroup and $\rho : \Gamma \to \mathcal{U}(V)$ be a unitary representation. Assume that $\Gamma \backslash G$ has a G-invariant locally finite measure μ. Then the induced representation $\pi := \mathrm{Ind}_\Gamma^G \rho$ is defined by

$$\pi = \left\{ f : G \to V : f(\gamma g) = \rho(\gamma^{-1})f(g) \right\}$$

with a natural action of G on the left (note that G acts on the right on the space, so on the left on the functions). The norm is defined by taking

$$\|f\| : \ \Gamma\backslash G \to \mathbb{R}$$

and integrating $\int \|f\|^2 \, d\mu$.

4.2. Mautner Phenomenon and Howe–Moore theorem

The goal of this section will be to prove the following theorem:

4.2.1. Theorem (Howe–Moore). *Let G be a connected real semisimple Lie group, with finite center. Suppose that V is a unitary representation of G such that for any of the simple factors G_i of G, there is no G_i-invariant vector in V. Then $\forall v, w \in V$:*

$$\lim_{g \to \infty} \langle gv, w \rangle \to 0$$

where $g \to \infty$ means that g leaves every compact set in G.

The property $\langle gv, w \rangle \to 0$ is also called *mixing*, or decay of correlations.

4.2.2. Corollary (Ergodicity of actions). *With the same assumptions as above, suppose that G acts on X preserving a probability measure μ, and the action is ergodic. Then any non-compact subgroup also acts ergodically on (X, μ).*

One example is G acting on G/Γ, where ergodicity of the full G-action is immediate and thus implies ergodicity (in fact mixing) of any 1-parameter subgroup.

Another example is $\mathbf{SL}_2(\mathbb{R})$ acting on the moduli space of translation surfaces, where ergodicity of the diagonal subgroup follows form the Hopf argument. Therefore any non-compact 1-parameter subgroup (e.g. the unipotent one) also acts ergodically, and in fact the action is mixing.

We will first prove Theorem 4.2.1 for the case of $\mathbf{SL}_2 \mathbb{R}$. For some notation, consider the subgroups

$$A := \left\{ a_t := \begin{bmatrix} t & 0 \\ 0 & t^{-1} \end{bmatrix} \right\} U^+ = \left\{ u_s := \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \right\} U^- = \left\{ u_s^- = \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix} \right\}$$

4.2.3. Proposition (Invariant vectors for $\mathbf{SL}_2$). *Suppose that V is a unitary representation of $\mathbf{SL}_2 \mathbb{R}$.*

If v is a vector that is A, or U^+, or U^--invariant, then it is $\mathbf{SL}_2 \mathbb{R}$-invariant.

Proof. The proof is based on the Mautner Lemma 4.1.4.

The first case is to show and A-invariant vector is U^+-invariant (and also U^--invariant by a similar argument). For this, take

$$g = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \quad s_t = a_{e^{-t}} = \begin{bmatrix} e^{-t} & 0 \\ 0 & e^t \end{bmatrix}$$

and note that $s_t g s_t^{-1} \to \mathbf{1}$ as $t \to \infty$.

For the case U^+-invariant implies A-invariant, take

$$s_1 = \frac{1-t}{\varepsilon}, s_2 = \frac{1-t^{-1}}{\varepsilon}$$

and compute

$$\begin{bmatrix} 1 & s_1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} t & 0 \\ \varepsilon & t^{-1} \end{bmatrix} \cdot \begin{bmatrix} 1 & s_2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \varepsilon & 1 \end{bmatrix}$$

so as $\varepsilon \to 0$ we conclude that any U^+-invariant vector is also A-invariant, again by the Mautner lemma. $\square$

Second proof of Proposition 4.2.3, after [McM17, §9]. The following is a more geometric way to argue for the implication U^+-invariant then A-invariant. Suppose that v is a U^+-invariant vector and let $c_{v,v} := \langle gv, v \rangle$ be its matrix coefficient. This descends to a function on $\mathbf{SL}_2\,\mathbb{R}/U^+$ which can be identified with $\mathbb{R}^2 \setminus 0$; moreover $c_{v,v}$ is also invariant by the left action of U^+.

So $c_{v,v} : \mathbb{R}^2 \setminus 0 \to \mathbb{C}$ is constant on U^+-orbits, i.e. all the horizontal lines except for the horizontal axis $(\bullet, 0)$, where each point is stabilized by U^+. But $c_{v,v}$ is continuous and constant on the lines $(\bullet, \varepsilon)$ so as $\varepsilon \to 0$ it follows that it is also constant on the line $(\bullet, 0)$, where it equals its value at $(1, 0)$, i.e. $c_{v,v}(\bullet, 0) = \|v\|^2$. But the horizontal axis is the image of A, so v is A-invariant, and so $c_{v,v}$ must also be constant on the orbits of A in $\mathbb{R}^2 \setminus 0$, which are hyperbolas. Thus $c_{v,v}$ is constant everywhere and v is $\mathbf{SL}_2\,\mathbb{R}$-invariant.

The piece of the argument that gave A-invariance from U^+-invariance can be summarized as:

$$c_{v,v}(x, 0) = \lim_{\varepsilon \to 0} c_{v,v}(x, \varepsilon) = \lim_{\varepsilon \to 0} c_{v,v}(1, \varepsilon) = c_{v,v}(1, 0)$$

from which one can design the matrices in the previous proof that are inserted into the Mautner lemma. $\square$

4.2.4. Theorem (Howe–Moore for $\mathbf{SL}_2$). *Suppose that V is a unitary representation of $\mathbf{SL}_2\,\mathbb{R}$.*

Then either V has an invariant vector, or it is mixing, i.e. $\forall v, w \in V$

$$\lim_{g \to \infty} \langle gv, w \rangle \to 0$$

Proof. Suppose by contradiction that there exists v, w and a sequence $g_i \to \infty$ such that the matrix coefficient does not go to 0, e.g. $\mathrm{Re} \langle g_i v, w \rangle \geq \varepsilon > 0$. Let $g_i = k_{i,1} a_i k_{i,2}$ be the KAK decomposition of g_i (see §3.6.7). By passing to a subsequence, assume that $k_{i,1} \to k_1$ and $k_{i,2} \to k_2$, since K is compact. It follows that

$$\varepsilon \leq \liminf \langle k_1 \cdot a_i \cdot k_2 \cdot v, w \rangle = \liminf \left\langle a_i \cdot (k_2 v), k_1^{-1} w \right\rangle$$

so up to replacing v, w, we can assume $\mathrm{Re} \langle a_i v, w \rangle \geq \varepsilon > 0$. Now $a_i v$ is a bounded sequence of vectors, so let v_0 be some weak limit. Then $v_0 \neq 0$ since $\mathrm{Re} \langle v_0, w \rangle \geq 0$ by construction.

It suffices to check that v_0 is U^+-invariant, since by Proposition 4.2.3 it will be $\mathbf{SL}_2 \mathbb{R}$-invariant. For $u \in U^+$, we have

$$\|uv_0 - v_0\| \leq \limsup_i \left\| a_i^{-1}(ua_i v - a_i v) \right\| = \|v - v\| = 0$$

where we used that $a_i^{-1} u a_i \to \mathbf{1}$ as $i \to \infty$. $\qquad\square$

A similar proof will be available for a general semisimple Lie group, once we talk about the KAK decomposition.

4.3. Property (T)

An extensive reference for this topic is the monograph [BdlHV08].

4.3.1. Definition (Almost invariance)**.** Let $K \subset G$ be a compact set and $\varepsilon > 0$ and $\pi : G \to \mathcal{U}(V)$ a unitary representation. A unit vector $v \in V$ is called (K, ε)-*almost invariant* if

$$\|\pi(g)v - v\| \leq \varepsilon \quad \forall g \in K$$

Equivalently, for a non-zero vector v require that

$$\|\pi(g)v - v\| \leq \varepsilon \|v\| \quad \forall g \in K$$

4.3.2. Example (Failure of almost invariance)**.** Consider the action of $\mathbb{R}$ on $L^2(\mathbb{R})$ by translations. This action is mixing, so it has no invariant vectors. But for any $\varepsilon > 0$ and interval $K := [a, b] \subset \mathbb{R}$ there exist (K, ε)-invariant vectors. For this, just take the indicator function of any sufficiently large interval.

On the other hand, Kazhdan discovered that higher-rank Lie groups satisfy a dichotomy - either they have no almost invariant vectors, or they must have actual invariant vectors. Equivalently, if there is an almost invariant vector, there is an actual invariant vector.

4.3.3. Definition (Property (T))**.** The group G has Kazhdan's property (T) if there exists a compact set $K \subset G$ and $\varepsilon > 0$ such that

if a unitary representation $\pi : G \to \mathcal{U}(V)$ has (K, ε)-almost invariant vectors, then it has an invariant vector.

The pair (K, ε) is called a *Kazhdan pair*.

The reason for (T) in the naming is from an interpretation of the definition in the space of unitary representations of G. It turns out that property (T) is equivalent to the trivial unitary representation being an isolated point of the space of all unitary representations, for an appropriate topology.

4.3.4. Definition (Almost-invariant vectors). A unitary representation of G on V is said to *almost have invariant vectors* if $\forall \varepsilon > 0, \forall K \subset G$ compact, there exists a nonzero vector $v \in V$ such that

$$\|gv - v\| \le \varepsilon \|v\| \quad \forall g \in K$$

The definition of property (T) in Definition 4.3.3 requires a single pair (K, ε) to work for all unitary representations, but it turns out that a weaker requirement suffices. Namely, the proposition below implies that property (T) is equivalent to: a unitary representation that almost has invariant vectors, has actual invariant vectors.

4.3.5. Proposition (Equivalent forms of Property (T)). *Let G be a locally compact group. The following properties are equivalent:*

   (i) *G has property (T) in the sense of Definition 4.3.3.*
   (ii) *For any unitary representation of G on a Hilbert space V, if $\forall \varepsilon > 0, \forall K$ compact, there exists a (K, ε)-almost invariant vector, then there exists a G-invariant vector.*
   (iii) *Any sequence of continuous positive-definite functions on G that converges to 1 uniformly on compact sets, converges uniformly to 1 on all of G.*
   (iv) *Every continuous affine isometric action of G on a real Hilbert space has a fixed point.*

Proof. It is clear that (i) implies (ii). For the converse, from Proposition 4.3.6 it follows that G is compactly generated (note that the proof used only the weaker version of property (T) specified in (ii)). Now let K be any compact generating set with non-empty interior. Suppose that V_ε is a unitary representation that has a (K, ε)-almost invariant vector, but no invariant vector. Then as $\varepsilon \to 0$ take $V := \oplus V_\varepsilon$ to be the direct sum representation. Then V almost has invariant vectors, since K is generating and has non-empty interior. By assumption (ii) it must have actual invariant vectors, and thus some V_ε has an invariant vector, which is a contradiction.

The proofs of (iii) and (iv) will be omitted for now, but note that (iv) can be seen as a cohomology vanishing property. For any unitary representation $\pi : G \to \mathcal{U}(V)$, equivalence classes of isometric affine actions of G on V are the same as elements $H^1(G, V)$. $\qquad\square$

4.3.6. Proposition (Some consequences of property (T)). *Suppose that G has property (T).*

(i) *If $G \twoheadrightarrow H$ is a continuous surjection then H also has property (T).*

(ii) *If $\Gamma \subset G$ is a closed subgroup such that G/Γ has a finite G-invariant measure, then Γ also has property (T).*

(iii) *If G is locally compact, then it is compactly generated.*

(iv) *The abelianization of a discrete Γ with property (T), $\Gamma^{ab} := \Gamma/[\Gamma, \Gamma]$, is finite.*

Proof. For part (i), any unitary representation of H is also a unitary representation of G, so if (K, ε) is a Kazhdan pair for G, then the same ε and the image of K in H will be a Kazdan pair for H.

For part (ii), pick first a finite volume fundamental domain $F \subset G$ for the left action of Γ, and normalize the measure such that F has volume 1. Assume that (K, ε) is a Kazhdan pair for G. Then the set $F \cdot K$ has finite volume, and there exist finitely many F-translates $\gamma_i \cdot F$ which cover $F \cdot K$, up to a set of measure of $\varepsilon/2$.

Suppose now that $\rho : \Gamma \to \mathcal{U}(V)$ is a unitary representation of Γ which has a $(\left\{\gamma_i^{-1}\right\}, \varepsilon/2)$-almost invariant vector v. Let $\pi := \mathrm{Ind}_\Gamma^G(\rho)$ be the induced representation to G and consider

$$\phi : G \to V$$

defined as $\phi|_F = v$ and extended by Γ-equivariance to all of G. To check that ϕ is now a (K, ε)-almost invariant vector, fix $k \in K$ and for $f \in F$ write

$$\phi(fk) = \phi(\gamma_i f') = \gamma_i^{-1} v \text{ for some } f' \in F$$

except for a set of f of measure $< \varepsilon/2$. Integrating the assumption that v is $\varepsilon/2$-almost invariant for Γ it follows that ϕ is (K, ε)-almost invariant. Since G has property (T) it follows that there is a G-invariant nonzero function $\psi : G \to V$, i.e. constant, which is moreover Γ-equivariant. The constant value of ψ is the Γ-invariant vector.

For compact generation, take the definition of property (T) to mean that if a unitary representation almost has invariant vectors, then it has invariant vectors. This is certainly implied by Definition 4.3.3. Let now K be any compact set which has interior (can do this by local compactness of G). Let G_K denote the group generated by K;

since K has interior it follows that G/G_K is discrete. The quotient G/G_K equipped with counting measure gives a unitary G-representation $l^2(G/G_K)$ for which the delta-function at the coset eG_K is G_K-invariant. Taking now $V := \oplus l^2(G/G_K)$ as K ranges over an exhausting family of compact sets with interior. The representation almost has invariant vectors (since it has K-invariant ones for any compact K), so it must have an invariant vector. It follows that some summand has an invariant vector, so G/G_K has finite volume, i.e. it is finite (since discrete).

The abelianization is finite since $\Gamma \twoheadrightarrow \Gamma^{ab}$ and an infinite abelian group does not have property (T) (it surjects onto $\mathbb{Z}$, which doesn't have property (T)). □

4.3.7. Corollary (Groups without property (T)).

    (i) *Free groups and surface groups do not have property (T) since their abelianization is infinite.*

    (ii) $\mathbf{SL}_2\,\mathbb{R}$ *does not have property (T) since it contains surface groups (or free groups) as lattices.*

4.3.8. Proposition (Compact Groups have property (T)). *A compact group G has property (T), specifically if a representation has a $(G, \sqrt{2})$-almost invariant vector, then it has an invariant vector.*

Informally, if v is an almost invariant vector for G, then $\int_G gv\,dg$ is going to be an actual invariant vector, provided it is non-vanishing.

Proof. Suppose that v is a $(G, \sqrt{2})$-almost invariant vector, i.e.

$$\|\pi(g)v - v\| \leq \sqrt{2} \quad \forall g \in G$$

Let $G \cdot v$ be its orbit and v_0 an element of minimal norm in the convex hull of the orbit. Then v_0 is G-invariant, so it suffices to check that $v_0 \neq 0$.

Using Eqn. (4.1.3) it follows that

$$2 > 2(\|v\|^2 - \operatorname{Re} c_{v,v}(g)) \quad \forall g \in G$$

Or equivalently $\operatorname{Re} c_{v,v}(g) > 0 \quad \forall g \in G$. But since G is compact, the inequality is uniform, i.e. there exists $\varepsilon > 0$ such that (making $c_{v,v}$ explicit):

$$\operatorname{Re} \langle gv, v \rangle > \varepsilon$$

Since v_0 is in the convex hull of $\{gv\}$ it follows that $\operatorname{Re} \langle v_0, v \rangle > \varepsilon$ which implies $v_0 \neq 0$. □

4.3.9. Theorem (Higher Rank Lie groups have Property (T)). *Suppose that G is a semisimple Lie group, with each simple factor of $\mathbb{R}$-rank at least 2. Then G has property (T).*

The proof of this result uses a relative notion of property (T).

4.3.10. Definition (Relative property (T))**.** Suppose that $H \subset G$ is a closed subgroup. Then the pair (G, H) has *relative property (T)* if whenever a unitary representation of G almost has invariant vectors, it actually has H-invariant vectors.

Theorem 4.3.11 below says that the pair $(\mathbf{SL}_2\,\mathbb{R} \ltimes \mathbb{R}^2, \mathbb{R}^2)$ has relative property (T). Assume this result for the next proof.

Proof of Theorem 4.3.9. Let V be a unitary representation that almost has invariant vectors. The following construction can be applied to any higher-rank group, but consider for simplicity $\mathbf{SL}_3\,\mathbb{R}$. Consider the subgroup

$$\left\{ \begin{bmatrix} A & N \\ 0 & 1 \end{bmatrix} : A \in \mathbf{SL}_2\,\mathbb{R}, N \in \mathbb{R}^2 \right\}$$

and note that by Theorem 4.3.11, N will have an invariant vector v. But by Proposition 4.2.3, v will also be invariant under the copy of $\mathbf{SL}_2\,\mathbb{R}$ embedded as

$$\begin{bmatrix} 1 & 0 \\ 0 & A \end{bmatrix} \quad A \in \mathbf{SL}_2\,\mathbb{R}$$

Similarly it will be invariant under a copy of $\mathbf{SL}_2\,\mathbb{R}$ embedded as

$$\begin{bmatrix} a & 0 & b \\ 0 & 1 & 0 \\ c & 0 & d \end{bmatrix} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathbf{SL}_2\,\mathbb{R}$$

and so the vector v will be invariant under all of $\mathbf{SL}_3\,\mathbb{R}$. $\qquad\square$

4.3.11. Theorem. *The pair* $(\mathbf{SL}_2\,\mathbb{R} \ltimes \mathbb{R}^2, \mathbb{R}^2)$ *has relative property (T).*

Let $V_3 := \mathrm{Sym}^2(\mathbb{R}^2)$ *denote the symmetric power representation of* $\mathbf{SL}_2\,\mathbb{R}$ *on* $\mathbb{R}^3$. *Then* $(\mathbf{SL}_2(\mathbb{R}) \ltimes V_3, V_3)$ *also has relative property (T).*

The result is also true for $\mathbb{R}$ replaced by any local field of characteristic zero, and even in characteristic p is we take $V_3^\vee$ instead of V_3. For an instance of the failure in characteristic 2, see [BdlHV08, Prop. 1.5.5]. The point is that there are non-trivial $\mathbf{SL}_2\,K$-invariant vectors in $V_3^\vee$, or equivalently invariant symmetric bilinear forms. For this, take the one given by the matrix $\begin{bmatrix} 0 & t \\ t & 0 \end{bmatrix}$ which will be invariant under $\mathbf{SL}_2\,K$ in characteristic 2.

Proof of Theorem 4.3.11. Let V be a unitary representation of $\mathbf{SL}_2\,\mathbb{R} \ltimes \mathbb{R}^2$ that almost has invariant vectors under $\mathbf{SL}_2\,\mathbb{R}$. View V as a unitary representation of $\mathbb{R}^2$. Any vector $v \in V$ determines a measure μ_v on

$\check{\mathbb{R}}^2$, which is the Fourier transform of the matrix coefficient $c_{v,v}$. The total mass of μ_v is $\|v\|^2 = c_{v,v}(1)$.

The assertion that v is $\mathbb{R}^2$-invariant is equivalent to μ_v being the delta-mass at the origin: $\mu_v = \|v\|^2 \delta_0$. Suppose, that there exists a compact generating set $K \subset \mathbf{SL}_2 \mathbb{R}$ and a sequence $v_i \in V$, with $\|v_i\| = 1$ such that $\|g v_i - v_i\| \le \varepsilon_i \to 0$ for all $g \in K$. Assume by contradiction that $\mu_{v_i}(0) = 0$ (otherwise one could extract from v_i a non-zero invariant vector). The almost invariance of v_i under K implies the same almost invariance for the measures μ_{v_i} (in the weak topology). The projection map $\mathbb{R}^2 \setminus 0 \to \mathbb{P}^1(\mathbb{R})$ then gives a K-almost invariant probability measure on $\mathbb{P}^1(\mathbb{R})$, and in the limit an actual invariant probability measure, which is a contradiction. $\qquad\square$

4.3.12. Expander Graphs.
A combination of Proposition 4.3.6 and Theorem 4.3.9 implies that $\mathbf{SL}_3(\mathbb{Z})$ has property (T). Let $S := \{A_1, \dots, A_N\}$ be a finite list of generators and for every prime p consider the quotient $\mathbf{SL}_3(\mathbb{Z}) \twoheadrightarrow \mathbf{SL}_3(\mathbb{F}_p) =: G_p$. The *Cayley graph* $\mathcal{C}(G_p, S)$ is defined to have vertices the elements of G_p and edges $g \to g \cdot s$ for all $s \in S$ (one can assume $S = S^{-1}$ to obtain an undirected graph). Note that the degree of each vertex is bounded independently of the prime p.

The group $\mathbf{SL}_3 \mathbb{Z}$ has a unitary representation on $l^2(\mathcal{C}(G_p, S))$ coming from the *right* action on the Cayley graph, specifically $R_h(g) := g \cdot h$. Note that this action does not necessarily preserve the edges of the Cayley graph. Consider now the operator $T_S := \frac{1}{|S|} \sum_{s \in S} R_s$ and its action on $l_0^2(\mathcal{C}(G_p, S))$, the zero-average functions for which there are no invariant vectors. Because $\mathbf{SL}_3 \mathbb{Z}$ has property (T), the operator T_S satisfies $\|(1 - T_S)v\| \ge \varepsilon \|v\|$ for a uniform $\varepsilon > 0$ independent of p. Indeed, there exists $\varepsilon' > 0$ such that for any unit vector $v \in l_0^2(\mathcal{C}(G_p, S))$ there exists $s \in S$ such that $\|R_s \cdot v - v\| \ge \varepsilon'$.

But the property $\|(1 - T_S)v\| \ge \varepsilon \|v\|$ implies that the spectrum of T_S is uniformly bounded away from 1. This, in turn, implies that the random walk generated by S equidistributes at a uniform rate in the Cayley graph.

Alternatively, one can see that the Cayley graph is highly connected, i.e. any partition of the graph into two sets must have a lot of edges going between the sets. To see it, consider a decomposition $\mathcal{C}(G_p, S) = A \coprod B$ into two disjoint sets. The goal is to show that

$$\frac{\#\mathrm{Edges}(A, B)}{\min(|A|, |B|)} \ge \varepsilon$$

with $\varepsilon > 0$ independent of the Cayley graph and sets A, B. For this, one applies Kazhdan's property (T) to the function $f := |B| \cdot \mathbf{1}_A - |A| \cdot \mathbf{1}_B$

which is constant on the two sets. Property (T) implies that there exists $s \in S$ such that $\|sf - f\| \geq \varepsilon \|f\|$, which reads

$$(|A| + |B|) \cdot (\#\mathrm{Edges}_s(A, B)) \geq \varepsilon'(|B|^2|A| + |A|^2|B|)^{1/2}$$

where $\mathrm{Edges}_s(A, B)$ denotes edges of type s between A and B. Rearranging the above expression gives the desired inequality.

In fact one can get expanders from $\mathbf{SL}_2(\mathbb{F}_p)$ as well. One technique is to use lower bounds on the dimensions of representations (see Exercise 9.2.10, which imply that eigenvalues of the Laplace operator must have high multiplicity.

FiXme!

FiXme: Add discussion of cohomology, fixed points for affine actions. Give an application of fixed points, e.g. acting on hyperbolic spaces with fixed points.

4.4. Amenability

There are several equivalent characterizations of amenability. Some are in terms of fixed points for actions, others are more intrinsic and use a notion of averaging on the group. The difference between discrete and continuous groups arises when using the space L^∞; in the non-discrete case, the action on L^∞ is not continuous in the strong topology.

4.4.1. Definition (Means on a measure space)**.** Let (X, μ) be a measure space. A *mean* on $L^\infty(X, \mu)$ is a norm-continuous linear functional $m : L^\infty(X, \mu) \to \mathbb{C}$ satisfying:

- $m(\mathbf{1}) = 1$ where $\mathbf{1}$ denotes a constant function equal to 1 on X.
- $m(f) \geq 0$ if $f \geq 0$.
- $m(\overline{f}) = \overline{m(f)}$.

Denote by $L^\infty(X, \mu)^\vee$ the norm dual of $L^\infty(X, \mu)$.

Let now a locally compact G act (measurably or continuously) on a measure space (X, μ).

4.4.2. Remark.

- (i) The action on $L^\infty(X, \mu)$ will be continuous for the weak-* topology coming from L^1, but not for the norm topology when G is not discrete. Thus the action on the dual of $L^\infty(X, \mu)$ will typically not be continuous.
- (ii) The space $L^1(X, \mu)$ is contained in the dual of $L^\infty(X, \mu)$ but this dual is much larger. A positive function in L^1 gives a mean, but we will see there are a lot more. Nevertheless, using Hahn–Banach it is possible to show that means coming from

L^1 are weak-* dense in the dual of L^∞. This is also true of L^1 in general. One has to work with nets however, since the weak-* topology on the dual of L^∞ is not metrizable.

Equip G with a fixed Haar measure, and for $A \subset G$ denote by $|A|$ its Haar measure. The group G is *amenable* if it satisfies any one of the following properties.

4.4.3. Proposition (Characterizations of amenability). *The following are equivalent:*

- *(Følner) There exists a sequence of measurable sets $A_i \subset G$ with $|A_i| > 0$ and a compact generating set $K \subset G$ such that*

$$\sup_{g \in K} \frac{|gA_i \triangle A_i|}{|A_i|} \to 0$$

  where $A \triangle B$ denotes the symmetric difference.
- *(L^1) The representation of G on $L^1(G)$ almost has invariant vectors.*
- *(L^2) The unitary representation of G on $L^2(G)$ almost has invariant vectors.*
- *(Means) The space $L^\infty(G)$ has an invariant mean.*
- *(Fixed points) For any continuous affine action of G on a compact convex set C in a locally convex Hausdorff topological space E, there exists a fixed point in C.*
- *(Invariant measures) For any continuous action of G on a compact space X, there exists an invariant probability measure.*

Proof.
Using Følner. Suppose that G satisfies the Følner condition. Then the indicator functions of the sets A_i provide a sequence of almost-invariant vectors in L^1.

Let us see why the Følner condition implies the existence of an invariant measure. For this, suppose that G acts continuously on a compact space X. Fix $x_0 \in X$ and a Følner sequence $A_i \subset G$. Denote by $\nu_{A_i} := \frac{1}{|A_i|}(Haar|_{A_i})$ the Haar measure restricted to A_i and normalized to be a probability measure. Set $\mathcal{O} : G \to X$ to be the orbit map $g \mapsto g \cdot x_0$ and $\mu_i := \mathcal{O}_*(\nu_{A_i})$ the pushed-forward probability measure on X. Then any weak limit μ of the μ_i will be a G-invariant probability measure on X. To check G-invariance, pull back any continuous function from X to G and use the Følner property of the sets A_i.

L^p **properties.** If f is an almost invariant L^1 function, then $|f|^{1/2}$ will then be an almost invariant L^2 function. This follows from the

elementary inequality $(a^{1/2} - b^{1/2})^2 \leq |a - b|$, which in turns can be checked directly by writing $a = x^2, b = y^2$ and assuming $a \geq b$.

Suppose now that f is an almost invariant L^2 function, normalized to $\|f\|_{L^2} = 1$. Then $|f|^2$ will be an almost invariant L^1 function, since

$$\left\||g \cdot f|^2 - |f|^2\right\|_{L^1}^2 = \|(|g \cdot f| - |f|) \cdot (|g \cdot f| + |f|)\|_{L^1}^2 \leq \quad \text{by Cauchy–Schwarz}$$

$$\leq \||gf| - |f|\|_{L^2} \cdot \||gf| + |g|\|_{L^2}$$

$$\leq 2\varepsilon$$

So the L^1 and L^2 conditions are equivalent.

Følner from L^1. Suppose that $f \in L^1(G)$ is almost-invariant, i.e. $\|gf - f\|_{L^1} \leq \varepsilon$ for all g in a compact generating set, and $\|f\|_{L^1} = 1$. There is no harm in assuming that $f = |f|$, and approximate further f by a finite linear combination of step functions. By rearranging the step function, assume that $f = \sum w_i \cdot \mathbf{1}_{A_i}$ where $A_i \supset A_{i+1}$ and $\sum w_i|A_i| = 1$. Because the sets are nested, if follows that $\mathbf{1}_{A_i}(x) - \mathbf{1}_{A_i}(gx)$ have the same sign (or perhaps zero) depending on whether $f(x) \lessgtr f(g(x))$.

The condition $\|gf - f\|_{L^1} \leq \varepsilon$ then translates into

$$\sum w_i|A_i \triangle gA_i| = \sum_i w_i|A_i|\frac{|A_i \triangle gA_i|}{|A_i|} \leq \varepsilon$$

so at least one of $\frac{|A_i \triangle gA_i|}{|A_i|} \leq \varepsilon$. When G is finitely generated, the same argument can be applied but first averaging over finitely many generators. The adaptation to the case that G is not discrete requires a bit more work (see [BdlHV08, G.5]).

Means and L^1 functions. Suppose now that there is a sequence $f_i \in L^1(G)$ of almost invariant functions. The sequence $|f_i|$ is also almost invariant, and assume that $\|f_i\|_{L^1} = 1$. Let $m_i \in L^\infty(G)^\vee$ denote the means associated to integrating against $|f_i|$ and let $m \in L^\infty(G)^\vee$ denote any weak-* limit. Then m is G-invariant, since $\forall g \in G, \forall f \in L^\infty(G)$:

$$(g \cdot m)(f) = m(g \cdot f) = \lim_i m_i(g \cdot f) = \lim_i g \cdot m_i(f)$$

$$= \lim_i (m_i(f) - \|g \cdot m_i - m_i\|_{L^1} \cdot \|f\|_{L^\infty})$$

$$= m(f)$$

To pass from an invariant mean to an almost invariant sequence of L^1 functions, one has to work with nets. Indeed, the space $L^\infty(G)^\vee$ with the weak-* topology is not metrisable and this adds further technicalities.

Equivalence of fixed points and invariant measures. Let us now check that the fixed points and invariant measures conditions are equivalent. In one direction, if G acts on a compact space X, it also acts

on the compact convex set of probability measures on X, so if G has the fixed point property on convex sets, it has an invariant probability measure on X. Conversely, if G acts continuously, affinely, on a compact convex set C, then the invariant measures assumption implies that G has an invariant probability measure μ on C. Then the barycenter construction gives a point $bary(\mu) \in C$ which is fixed by G (since it only depends on μ). Recall that $bary(\mu)$ is defined by the property:

$$\int_C \xi(x)d\mu(x) = \xi(bary(\mu)) \quad \forall \xi \text{ continuous, linear function.}$$

Assume now that G has the fixed point property on convex sets. Then the space of means on $L^\infty(G)$ is itself a convex set, compact for the weak-* topology, on which G acts. But the action is not continuous, if G is not discrete! One has to pass through the space $UCB(G)$ of (say left-)uniformly continuous, bounded functions on G, on which the G-action is continuous. For details on this, see [Zim84, §7.2]. $\qquad\square$

4.4.4. Example (Some amenable groups)**.**

(i) Both $\mathbb{Z}$ and $\mathbb{R}$ are amenable, by exhibiting Følner sets.

(ii) A quotient of an amenable group is amenable (by the fixed point property).

(iii) A closed subgroup $H \subset G$ is amenable, if G is amenable. To see it, embed $L^\infty(H) \hookrightarrow L^\infty(G)$ in an H-equivariant way (by picking a section $s : G/H \to G$) and pull back a G-invariant mean.

(iv) In a short exact sequence

$$0 \to H' \to H \to H'' \to 0$$

if H', H'' are amenable, then so is H. This follows from the criterion on fixed points on convex sets. Given an H-action on a compact convex set C, the set of H'-fixed points is itself compact convex, call it C'. Then H'' will act on C' and will have a non-trivial fixed point.

(v) A group is amenable if the union of all of its compactly generated subgroups is amenable (apply the criterion with convex sets).

(vi) If a group has a dense amenable subgroup, then it is amenable (the fixed points in a convex set for the dense group will also be fixed by the whole group).

(vii) Abelian groups are amenable. For finitely-generated ones, this follows from the amenability of $\mathbb{Z}$ and $\mathbb{R}$.

(viii) Solvable groups are amenable, as they are successive extensions of abelian groups.

(ix) Compact groups are amenable. Take for Følner sets the entire group.

(x) Minimal parabolic subgroups in a semisimple Lie group (§3.6.5) are amenable. Indeed, they are extensions of solvable groups by compact subgroups.

(xi) Semisimple Lie groups are not amenable. Indeed, they do not have invariant probability measures when acting on G/P with P a proper parabolic subgroup.

(xii) If a group G is amenable and has property (T), then it must be compact. Indeed $L^2(G)$ will almost have invariant vectors by amenability, so it will have invariant vectors by property (T), therefore G has finite Haar volume.

4.4.5. Suspension construction. Suppose that $\Gamma \subset G$ is a lattice and Γ acts on a space X (measurably, or continuously, or etc.). To obtain an action of G, consider the suspended space

$$X \times_\Gamma G := \Gamma \backslash (X \times G) \ \text{ with } \gamma \cdot (x, g) := (\gamma \cdot x, \gamma \cdot g)$$

which is an X-bundle over $\Gamma \backslash G$. Moreover, G acts on the right on both spaces, equivariantly for the projection. Properties of the Γ-action on X can now be translated into those of the suspended space with a G-action.

An important application of amenability is the following result, which can be interpreted as a generalization of the Oseledets theorem to the setting of semisimple group actions.

4.4.6. Proposition (Boundary maps, Furstenberg)**.** *Let $\Gamma \subset G$ be an irreducible lattice in a semisimple Lie group G, and let $P \subset G$ be a minimal parabolic. Suppose that Γ acts continuously on some compact space X. Then there exists a measurable Γ-equivariant map*

$$\Theta : G/P \to \mathcal{M}^1(X)$$

to the space $\mathcal{M}^1(X)$ of probability measures on X.

Equivalently, there exists a P-invariant measure ν on the suspended space $X \times_\Gamma G$ which projects to Haar measure on $\Gamma \backslash G$.

The space G/P can be interpreted as a boundary at infinity for Γ (or G) and is sometimes called the Furstenberg boundary.

Proof. For the suspended space, the projection map $\pi : X \times_\Gamma G \to \Gamma \backslash G$ is G-equivariant, for the right action of G on both spaces. Equip $\Gamma \backslash G$ with its canonical G-invariant probability measure μ and let

$$\mathcal{M}^1(X \times_\Gamma G, \mu) = \{\nu \text{ prob. measure on } X \times_\Gamma G : \pi_* \nu = \mu\}$$

be the space of probability measures on the total space which project to Haar measure on $\Gamma\backslash G$.

The minimal parabolic $P \subset G$ acts affinely on $\mathcal{M}^1(X \times_\Gamma G, \mu)$, which is a compact convex set in the space of all fiberwise measures. Since P is amenable, there exists a P-invariant measure $\nu \in \mathcal{M}^1(X \times_\Gamma G, \mu)$. Disintegrating ν along the fibers, this gives a family of probability measures ν_t on the fibers X_t above $t \in \Gamma\backslash G$. Lifting now to G, this gives a Γ-equivariant map

$$G \to \mathcal{M}^1(X)$$

which is P-invariant, so descends to the desired map

$$\Theta : G/P \to \mathcal{M}^1(X).$$

$\square$

For a minimal parabolic $P \subset G$, the space G/P can be identified with an appropriate flag manifold for G. For the example of $P \subset \mathbf{SL}_n \mathbb{R}$, the quotient G/P is the space of flags:

$$0 \subsetneq F_1 \subsetneq \cdots \subsetneq F_{n-1} \subsetneq \mathbb{R}^n$$

where $\dim F_i = i$.

4.5. Aside: Unitary representations of abelian groups

Correspondence between positive-definite functions on G and measures on $\check{G}$.

Description of any (separable) unitary representation as a direct integral.

5. Super-Rigidity and Normal Subgroup Theorem

For a semisimple real Lie group G, a lattice $\Gamma \subset G$ is *irreducible* if it projects densely to any simple factor of G.

5.1. Super-Rigidity

5.1.1. Theorem (Margulis super-rigidity[5]). *Let* **G** *be a semisimple, connected,* $\mathbb{R}$*-algebraic group with* $\mathrm{rk}_{\mathbb{R}}\,\mathbf{G} \geq 2$ *and no compact factors. Let* $\Gamma \subset \mathbf{G}(\mathbb{R})$ *be an irreducible lattice.*

Let **H** *be a* k*-simple algebraic group, where* k *is* $\mathbb{R}, \mathbb{C}$ *or* $\mathbb{Q}_p$*, and let*

$$\rho : \Gamma \to \mathbf{H}(k)$$

be a homomorphism with unbounded, Zariski-dense image.

Then ρ *extends to a continuous homomorphism* $\mathbf{G}(\mathbb{R}) \to \mathbf{H}(k)$.

5.1.2. Remark.

(i) The Zariski-density of $\rho(\Gamma)$ is convenient but not necessary, since **H** can always be replaced by the Zariski closure of $\rho(\Gamma)$.

(ii) When $k = \mathbb{Q}_p$ there can be no continuous homomorphism as in the conclusion of the theorem, so there can be no unbounded representation $\Gamma \to \mathbf{H}(\mathbb{Q}_p)$. For example, the inclusion $\mathbf{SL}_n\,\mathbb{Z} \hookrightarrow \mathbf{SL}_n\,\mathbb{Q}_p$ does not extend to a homomorphism from $\mathbf{SL}_n\,\mathbb{R}$, since the image is compact.

(iii) In Example 3.1.8, $\Gamma \subset \mathbf{G}(\mathbb{R})$ where $\mathbf{G}(\mathbb{R}) \cong \mathbf{SO}_{m,n}(\mathbb{R}) \times \mathbf{SO}_{m+n}(\mathbb{R})$. Projecting to the first factors gives a lattice $\Gamma_1 \subset \mathbf{SO}_{m,n}(\mathbb{R})$ and projecting to the second factor gives a (dense!) subgroup $\Gamma_2 \subset \mathbf{SO}_{m+n}(\mathbb{R})$. The group-theoretical isomorphism $\Gamma_1 \cong \Gamma_2$ gives a representation $\Gamma_1 \to \mathbf{SO}_{m+n}(\mathbb{R})$ which does not extend to a homomorphism from $\mathbf{SO}_{m,n}(\mathbb{R})$. Again, the image of Γ is bounded.

(iv) The assumption that **H** is simple implies that it is center-free. To see that it is necessary, take $\mathbf{G} = \mathbf{PGL}_{2n}$ and $\mathbf{H} = \mathbf{SL}_{2n}$ and a finite index subgroup $\Gamma \subset \mathbf{SL}_n\,\mathbb{Z}$ which is torsion free. Although $\mathbf{SL}_{2n}(\mathbb{R}) \hookleftarrow \Gamma \hookrightarrow \mathbf{PGL}_{2n}(\mathbb{R})$, it does not lift to a homomorphism $\mathbf{PGL}_{2n}(\mathbb{R}) \to \mathbf{SL}_{2n}(\mathbb{R})$ (since $\pm 1 \in Z(\mathbf{SL}_{2n}(\mathbb{R}))$).

5.1.3. Proof strategy for super-rigidity. To simplify notation, set $G := \mathbf{G}(\mathbb{R}), H := \mathbf{H}(k)$. Fix a minimal parabolic $P \subset G$, a Cartan involution $\sigma : G \to G$ and thus opposite parabolic $P^- = \sigma(P)$ and let $A \subset P \cup P^-$ be a split Cartan. The higher rank assumption on G implies that $\dim A \geq 2$.

For the target, view $H \subset \mathbf{GL}(V)$ as acting irreducibly and faithfully on some k-vector space V. The representation $\rho : \Gamma \to H$ gives an

---

[5]For the spelling of "super-rigidity" I follow the Chicago Manual of Style, Hyphenation table http://www.chicagomanualofstyle.org/16/images/ch07_tab01.pdf which recommends hyphenating in order to separate two repeating letters.

action of Γ on V and $\mathbb{P}(V)$ as well. The approach below is a hybrid between that in [Ben08] (which is close to [Mar91]) and [Zim84].

Step 1 uses amenability of P and the unboundedness of the image $\rho(\Gamma) \subset H$. The goal is to produce a vector bundle $E \to \Gamma \backslash G$ with a G-action, and a section of E which is invariant under the subgroup A. The vector bundle E will come from the suspension construction applied to the linear representation $\Gamma \to H \to \mathbf{GL}(n)$ for some n.

Specifically, Proposition 4.4.6 gives a Γ-equivariant measurable map

$$\Theta : G/P \to \mathcal{M}^1(\mathbb{P}(V))$$

where $\mathcal{M}^1$ denotes the space of probability measures.

Next, use the unboundedness of Γ in H to upgrade the map Θ to a measurable map

$$\theta : G/P \to H/L$$

where $L \subset H$ is (the k-points of) some proper algebraic subgroup. There are two ways to achieve this, giving slightly different information on L. One is to use the notion of proximal elements and show that L can be assumed to be a parabolic subgroup (this is the approach in [Ben08]).

The route below (following [Zim84]) is to use the tameness of the action on the space of measures. When $k = \mathbb{R}$, the stabilizer of a measure on $\mathbb{P}(V)$ is an algebraic subgroup, while for $k = \mathbb{Q}_p$ the stabilizer is either compact, or contained in some proper algebraic subgroup $L \subset G$.

When $k = \mathbb{R}$, we get by construction $L \subset H$ a proper algebraic subgroup (and we know that H is non-compact by assumption). When $k = \mathbb{Q}_p$ there is an extra lemma in Zimmer.

Applying the contragredient representation (i.e. taking $x \mapsto (x^{-1})^t$ throughout) gives another Γ-equivariant boundary map

$$\theta_- : G/P_- \to \theta(H)/\theta(L) = H/\sigma(L)$$

which can then be interpreted as a dual line-subbundle in the dual representation. Using projections, this gives a non-zero section of $\mathbf{End}(V) \to \Gamma \backslash G$ which is invariant under $P \cap P^-$.

Step 2 uses the higher-rank assumption on $\mathbf{G}$. The approach in Zimmer is to show that the map σ from the previous step is not just measurable but in fact polynomial, and conclude from there. The route below is to produce more invariant sections under larger and larger subgroups of G, until G acts on a finite-dimensional space of sections. This will be the desired extension of the representation.

Moving the sections around gives the finite-dimensional representation we're after.

A little hiccup here - after identifying $G/P \times G/P_-$ with G/A (essentially) need to check that the images are generically transverse. Again, I think an argument a la Zimmer should give it.

There seems to be an interesting relation between showing non-vanishing of Lyapunov exponents and Zimmer tameness and algebraic hull. They're somehow addressing the same issues.

5.2. Proof of step 1 of super-rigidity

With the assumptions and notations from before, the goal of this section is:

5.2.1. Proposition (One invariant section). *There exists a faithful representation $\pi : H \to \mathbf{GL}(W)$, a vector $w \in W \setminus \{0\}$ and a Γ-equivariant map*

$$G/A \to H \cdot w \subset W$$

defined a.e. for the natural Lebesgue measure class on the left-hand side. The action of Γ on the right is via $\pi \circ \rho$.

Equivalently, on the G-equivariant vector bundle

$$W \times_\Gamma G \to \Gamma \backslash G$$

there exists a measurable non-zero A-invariant section, defined a.e. for the Lebesgue measure class on $\Gamma \backslash G$.

The proof will require several steps. First, some preliminaries on Γ-equivariant maps; throughout, Γ acts on the right-hand side by the representation $\rho : \Gamma \to H$.

5.2.2. Proposition (Existence of a factorization). *Suppose that $D \subset G$ is a non-compact subgroup. For any two Γ-equivariant maps*

$$G/D \to H/L_1 \qquad G/D \to H/L_2$$

defined a.e. for Lebesgue measure on G/D (and $L_i \subset H$ algebraic subgroups) there exists a third subgroup $L_3 \subset H$ and injections $L_3 \hookrightarrow L_1$, $L_3 \hookrightarrow L_2$ given by conjugation inside H, such that the relevant diagram commutes.

5.2.3. Definition (Algebraic hull, à la Zimmer). Proposition 5.2.2 implies that there exists a smallest, up to conjugacy, algebraic subgroup $L \subset H$ such that any Γ-equivariant map to an H-homogeneous space factors through

$$G/D \to H/L.$$

This L is called the *algebraic hull* of the action.

Note that the algebraic hull depends on the subgroup $D \subset G$ and if $D_1 \subset D_2$, then the algebraic hull of D_1 is contained in that of D_2.

Proof of Proposition 5.2.2. Consider the Γ-equivariant map

$$G/D \to H/L_1 \times H/L_2.$$

Since D is non-compact, the action of Γ on G/D is ergodic by the Howe–Moore Theorem 4.2.1 (applied to the D-action on $\Gamma\backslash G$). It therefore follows, from the tameness of the H-action on the right-hand side, that the image of G/D is inside a single H-orbit, say of the point $(h_1 L_1, h_2 L_2) \in H/L_1 \times H/L_2$. The stabilizer of this point is $(h_1 L_1 h_1^{-1}) \cap (h_2 L_2 h_2^{-1}) =: L$ which satisfies the requirements. $\square$

Proof of Proposition 5.2.1. By Proposition 4.4.6, there exists a Γ-equivariant map

$$\Theta : G/P \to \mathcal{M}^1(\mathbb{P}(V))$$

into the space of probability measures on the projectivization of a faithful H-representation V. By the tameness of algebraic actions on the space of measures (Theorem 3.2.8), it follows that Θ in fact descends to a single H-orbit:

$$\theta : G/P \to H/L$$

with $L \neq H$, since L is the stabilizer of a probability measure on $\mathbb{P}(V)$ and H cannot stabilize such a measure by assumption[6]. Therefore the algebraic hull of the P-action is non-trivial, so assume that L is minimal with the property that there exists a Γ-equivariant section as above. By one of Chevalley's theorems (Theorem 3.3.7) there exists a representation W of H and a line $l \subset W$ such that $L = \mathrm{Stab}_H(l)$; by shrinking W is necessary, assume that $H \cdot l$ spans W.

For the fixed Cartan involution $\sigma : G \to G$, consider $P^- := \sigma(P)$. First, by Lemma 5.2.4 there exists a Lebesgue-measure class preserving G-equivariant map:

$$G\big/(P \cap P_-) \to G/P \times G\big/P_- .$$

The upshot (NOT FINISHED) is that there exists a Γ-equivariant map

$$G\big/(P \cap P_-) \to \mathbb{P}(W) \times \mathbb{P}(W^\vee)$$

where Γ acts in the standard (not σ-conjugated) ways on both sides. Moreover, the image is a.e. contained in the open set

$$\mathbb{P}(W) \times \mathbb{P}(W^\vee) \supset \mathcal{U}^\pitchfork = \{([v], [\xi]) : \xi(v) \neq 0\}$$

---

[6]In the p-adic case L is not necessarily an algebraic subgroup of H and a further argument is required.

on which there is a map

$$\mathcal{U}^\pitchfork \to \mathbf{End}(W)$$

$$([v], [\xi]) \mapsto \text{projection onto } [v] \text{ along } \ker \xi.$$

The desired Γ-equivariant map is then

$$G/A \to \mathbf{End}(W)$$

whose image is clearly non-zero. $\qquad\qquad\qquad\qquad\qquad\qquad\square$

5.2.4. Lemma (Open cell of the flag manifold)**.** *The natural map*

$$G\big/(P \cap P_-) \to G/P \times G\big/P_-$$

is a G-equivariant isomorphism on a full Lebesgue measure set.

Proof. For $G = \mathbf{SL}_n(\mathbb{R})$ the space on the right-hand side is the space of pairs of flags (F_1, F_2) and the isomorphism is onto the open set where the flags are in transverse position.

The space on the left-hand side is the space of decompositions of $\mathbb{R}^n$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

5.3. Proof of step 2 of super-rigidity

Recall from Proposition 5.2.1 that there exists a representation $\pi : H \to \mathbf{GL}(W)$ which gives rise to a G-equivariant bundle

$$\mathcal{W} := W \times_\Gamma G \to \Gamma/G$$

and with a nonzero A-invariant (measurable) section s_θ. The goal of this section will be:

5.3.1. Proposition (Finite-dimensionality and homomorphisms)**.** *Assume that* $\mathrm{rk}_\mathbb{R} \, G \geq 2$. *Let S denote the space of G-translates of s_θ. Then*

    (i) *S is finite-dimensional.*
    (ii) *The (a priori measurable) representation $G \to \mathbf{GL}(S)$ is continuous.*
    (iii) *The sections in S have continuous representatives.*
    (iv) *The evaluation of the continuous representatives at the identity coset $\Gamma e \in \Gamma/G$ induces a linear map $S \to W$ and a homomorphism $G \to H$ extending $\Gamma \to H$.*

A measurable group homomorphism $G_1 \to G_2$ is always continuous (Exercise 9.2.11). The finite-dimensionality of S will follow from the higher-rank assumption on G and some general facts about finite-dimensionality of spaces of sections.

5.3.2. Proposition (Spaces of sections). *Suppose that $\mathcal{W} \to (X, \mu)$ is a measurable vector bundle, equivariant for an action (on the right) of a group A on both X and $\mathcal{W}$; assume that the measure-class preserving action of A on (X, μ) is ergodic.*

   (i) *The space of A-invariant sections of $\mathcal{W}$ is finite-dimensional.*
  (ii) *Suppose that $A' \subset A \subset Z(A')$ is a sequence of groups such that A' is in the center of $Z(A')$, and A' acts ergodically on (X, μ). Let S be a finite-dimensional, A-invariant space of measurable sections of $\mathcal{W}$. Then the spaces of $Z(A')$-translates of S is also finite-dimensional and A-invariant.*

Note the significant difference between an A-invariant section s (i.e. $s(x \cdot a) = s(x) \, \forall a \in A, x \in X$) and an A-invariant space of sections S (i.e. $\forall s \in S, \forall a \in A$ the pulled-back section satisfies $a^* s \in S$).

5.3.3. Proposition (Successive commutators fill out G). *Suppose that G is a semisimple Lie group with $\mathrm{rk}_{\mathbb{R}} G \geq 2$ and $A \subset G$ is a Cartan subgroup. Then there exists a sequence of non-trivial subgroups $A_i \subset A$ such that their commutators $Z(A_i)$ fill out G, i.e. the map*

$$Z(A_1) \times \cdots \times Z(A_n) \to G$$

is surjective.

Note that $A \subset Z(A_i)$ always, since A is abelian. Assume for the moment the last two propositions (to be proved below).

Proof of Proposition 5.3.1. Start with the space $S_0 := \{s_\theta\}$ which is a 1-dimensional A-invariant space of sections. Proposition 5.3.3 gives a sequence of subgroups $A_i \subset A \subset Z(A_i)$ such that the translates $S_i := Z(A_i) \cdot S_{i-1}$ are still A-invariant spaces of sections, finite-dimensional by Proposition 5.3.2. The final space $S := S_n$ coincides with the space of all G-translates of s_θ and is finite-dimensional by construction.

The (measurable) homomorphism $G \to \mathbf{GL}(S)$ is continuous by Exercise 9.2.11, so fixing a basis $s_1, \ldots, s_d$ of S, there exist continuous functions $c_i(g)$ such that

$$g^* s_\theta = c_1(g)s_1 + \cdots + c_d(g)s_d \text{ where } s_\theta \text{ is the initial section.}$$

Interpret a section of $\mathcal{W}$ as a map $s : G \to W$ which is Γ-equivariant and apply Fubini to the above equation to find that for Lebesgue a.e. $x \in G$:

$$s_\theta(xg) = c_1(g)s_1(x) + \cdots + c_d(g)s_d(x)$$

for Lebesgue a.e. $g \in G$, where the identity is interpreted as an equality of functions $G \to W$. Selecting one such x, it follows that s_θ agrees a.e.

with a continuous section. Since s_θ has a continuous representative, so do all its translates and therefore any function in the space S.

Evaluating at the identity coset $\Gamma e \in \Gamma \backslash G$ gives a linear map $S \to W$. Since the right action of Γ on $\Gamma \backslash G$ fixes the identity coset, the map $S \to W$ is Γ-equivariant, where Γ acts on S by $\Gamma \subset G \to \mathbf{GL}(S)$ and on W by $\Gamma \to H \to \mathbf{GL}(W)$.

We will check that the map $S \to W$ is injective, the image is H-invariant, and the image of G in $\mathbf{GL}(W)$ is contained in H.

First, since $\Gamma \subset G$ is Zariski-dense (by Borel's density theorem) the kernel of the map will be Γ-invariant, hence G-invariant[7] and so any section in the kernel must vanish identically. The image $W' \subset W$ is Γ-invariant, and since $\rho(\Gamma) \subset H$ is Zariski-dense, it follows that the image is also H-invariant. Thus the map $G \to \mathbf{GL}(S) \to \mathbf{GL}(W')$ lands in H and extends $\rho : \Gamma \to H$. $\qquad\square$

Proof of Proposition 5.3.2. The vector bundle $\mathcal{W}$ is finite-dimensional. Given A-invariant sections $s_1, \ldots, s_n$, the set where they are linearly-dependent is A-invariant, hence has either full or null measure. Moreover, if

$$s_1 + f_2 s_2 + \cdots + f_n s_n = 0 \quad \text{for functions } f_i \text{ on } X$$

then the f_i are a.e. unique, A-invariant, and hence constant. Therefore the space of A-invariant sections is at most $\dim \mathcal{W}$-dimensional.

For the second part, recall that S is an A-invariant space of sections and let S' denote the span of its translates under $Z(A')$. View S as a vector space with an A-action and define $\mathcal{S} := S \times X$ to be the trivial vector bundle with non-trivial A-action on the fibers and base.

An element $z \in Z(A')$ determines, by translation, a section $\tau(z) \in \operatorname{Hom}(\mathcal{S}, \mathcal{W})$ which assigns to $s \in S$ the translated section of $\mathcal{W}$:

$$\tau(z)_x(s) = s(x \cdot z^{-1}) \cdot z$$

Since $Z(A')$ commutes with A', it follows that $\tau(z)$ is in fact A'-invariant:

$$\begin{aligned}
\tau(z)_{x \cdot a'}(s) &= s(xa'z^{-1})z = s(xz^{-1}a')z &&\text{by } A'\text{-invariance} \\
&= s(xz^{-1})a'z = s(xz^{-1})za' \\
&= \tau(z)_x(s) \cdot a'
\end{aligned}$$

which is exactly the A'-invariance condition.

By the previous part of the proposition, since A' acts ergodically, it follows that $\tau(z)$ spans a finite-dimensional space of sections in

---

[7] The continuous representation $G \to \mathbf{GL}(S)$ can be assumed algebraic.

$\mathrm{Hom}(\mathcal{S}, \mathcal{W})$ as z ranges over $Z(A')$. It then follows that the space of translates of S by $Z(A')$ is finite-dimensional. □

Proof of Proposition 5.3.3. Let $\alpha_1, \ldots \alpha_r \in \mathfrak{a}^\vee$ be the roots of G. Since $\mathrm{rk}_\mathbb{R} G \geq 2$ the subspaces $\ker \alpha_i$ are non-trivial and determine a non-trivial subgroup $A_i \subsetneq A$. Moreover $Z(A_i)$ contains non-trivial unipotent elements by construction. Therefore $X := Z(A_1) \cdots Z(A_n)$ contains both a minimal parabolic and its opposite. Therefore $X \cdot X$ contains a dense open in G, and so $X \cdot X \cdot X \cdot X$ covers all of G. □

5.4. Proof of super-rigidity

Assume the setting of Theorem 5.1.1 and the notation and setup of §5.1.3.

FiXme: I need to see where I'm using that Γ has non-trivial image, when H has no compact factors.

5.4.1. Step 1: the boundary maps. This is the step which uses the unboundedness assumption of $\rho : \Gamma \to H$. When H is real, assume that H has no compact factors by quotienting them out. It is still true that ρ has unbounded image. In fact, at this point it suffices to have non-trivial image. When H is p-adic, a further argument will be required.

By Proposition 4.4.6 there exists a Γ-equivariant map

$$\Sigma : G/P \to \mathcal{M}^1(\mathbb{P}(V))$$

where V is a faithful H-representation. The left action of Γ on G/P is ergodic, since the right action of P on $\Gamma \backslash G$ is ergodic by Howe–Moore Theorem 4.2.1. Furthermore, the action of H on $\mathcal{M}^1(\mathbb{P}(V))$ is tame in the sense of Zimmer (Definition 3.2.5) by Theorem 3.2.8, so that by Proposition 3.2.6 the image of Σ lies in a single H-orbit (a.e. for the natural measure on G/P).

For a semisimple real Lie group H without compact factors acting linearly on $\mathbb{P}(V)$, the stabilizer of any measure is a proper algebraic subgroup $L \subset H$ by [Zim84, 3.2.19]. When H is p-adic, either the stabilizer is contained in a proper algebraic subgroup, or the stabilizer is compact. An extra argument ([Zim84, 5.1.9]) is required to show that in this case the image $\rho(\Gamma) \subset H$ has compact closure, contradicting the unboundedness assumption.

5.5. Arithmeticity of lattices

A consequence of super-rigidity, applied to both real and p-adic target Lie groups, implies that lattices in higher-rank semisimple Lie groups are arithmetic.

5.5.1. Corollary (Arithmeticity of Lattices)**.** *Suppose that $\Gamma \subset G$ is a lattice in a real algebraic semisimple group with $\mathrm{rk}_{\mathbb{R}} G \geq 2$. Then Γ is an arithmetic lattice.*

For the sketch of proof below, take an arithmetic lattice to be one that can be obtained, up to finite index, from an embedding $G \hookrightarrow \mathbf{SL}_n(\mathbb{R})$; i.e such that $G \cap \mathbf{SL}_n(\mathbb{Z})$ agrees with Γ up to a finite index in both groups.

Sketch of proof of Corollary 5.5.1. Super-rigidity applied to $\Gamma \to G$ implies that the space of deformations of the representation $\Gamma \to G$ is trivial. In other words, any other representation $\rho : \Gamma \to G$ is a conjugate of the standard embedding. This implies that $\Gamma \subset G$ can be conjugated to have algebraic entries, for if not, a transcendental parameter would give non-trivial deformations. Specifically, the point $[id]$ in the character variety $\mathrm{Hom}(\Gamma, G)//G$ is isolated, so it must have coordinates in $\overline{\mathbb{Q}}$, therefore it has a lift to $\mathrm{Hom}(\Gamma, G)$ with coordinates in $\overline{\mathbb{Q}}$.

Take now all the Galois conjugates $\sigma(\Gamma) \subset G^\sigma$ to get an embedding $\Gamma \to G \times \cdots \times G^\sigma \subset \mathbf{SL}_n(\mathbb{R})$ which is Galois-invariant, hence $\Gamma \subset \mathbf{SL}_n(\mathbb{Q}) \cap G \times \cdots G^\sigma$; note that all the G^σ will be compact, again since $\sigma(\Gamma)$ is Zariski-dense and non-compactness of G^σ would contradict super-rigidity.

To replace $\mathbf{SL}_n(\mathbb{Q})$ by $\mathbf{SL}_n(\mathbb{Z})$ it now suffices to give a uniform bound on the p-adic valuation of all elements of Γ. But this follows from super-rigidity applied to the map $\Gamma \to \mathbf{SL}_n(\mathbb{Q}) \to \mathbf{SL}_n(\mathbb{Q}_p)$, which implies that the image of Γ is bounded, hence so are the p-adic valuations. (One also needs to know that Γ is finitely generated, which is true since it has property (T)). $\qquad\square$

5.6. Margulis Normal Subgroup Theorem

5.6.1. Theorem (Margulis normal subgroup theorem)**.** *Suppose that $\Gamma \subset G$ is a lattice in a real algebraic semisimple group with $\mathrm{rk}_{\mathbb{R}} G \geq 2$. Then any normal subgroup $N \subset \Gamma$ is either finite or finite index.*

5.6.2. Remark.

(i) The statement is false for rank 1 groups, for example surface or free groups. The kernel of the abelianization homomorphism provides an infinite index normal subgroup.

(ii) The statement can be recovered for rank 1 groups, if the normal subgroup is additionally assumed to be finitely generated (see Exercise 9.2.6).

(iii) If N is a finite normal subgroup in a lattice $\Gamma \subset G$ with connected, higher rank G, then N is central. Indeed, the N will remain normal in the Zariski closure of Γ, i.e. G, and so will be normal and discrete in G. By connectedness of G, it follows that N is central (it is in the kernel of the adjoint representation).

5.6.3. Corollary (Computing monodromy). *Consider the family of Riemann surfaces, depending on the parameter $t \in \mathbb{A}^1$:*

$$y^2 = f_{2g}(x)(x - t)$$

where $f_{2g}(x)$ is a polynomial of degree $2g$ with no multiple roots. Then the monodromy of this family is a finite-index subgroup of $\mathbf{Sp}_{2g}(\mathbb{Z})$.

This follows from two facts. First, the larger family

$$y^2 = (x - a_1) \cdots (x - a_{2g+1})$$

depending on $a_\bullet \in \mathbb{A}^{2g+1} \setminus (\bigcup_{i,j}\{a_i \neq a_j\})$ has monodromy of finite index in $\mathbf{Sp}_{2g}(\mathbb{Z})$ (a result of A'Campo). The first family arises as a fiber in a fibration of the base space of the second family, so its fundamental group is normal, hence the monodromy is normal in a finite index subgroup of $\mathbf{Sp}_{2g}(\mathbb{Z})$. It is easily verified that the monodromy of the first family is infinite, hence the Margulis normal subgroup theorem implies that it is also of finite index in $\mathbf{Sp}_{2g}(\mathbb{Z})$.

Proof sketch of Theorem 5.6.1. Consider the quotient Γ/N. Since Γ has property (T), so does Γ/N by Proposition 4.3.6. It suffices to prove that Γ/N is amenable, since in that case by Example 4.4.4(xii) it is compact, hence finite.

To check amenability of Γ/N, assume that it acts on a compact metric space X; it suffices to find an invariant probability measure. By Proposition 4.4.6, there exists a Γ-equivariant map $G/P \to \mathcal{M}^1(X)$.

By Theorem 5.6.4 below, the image is isomorphic to some G/Q for a parabolic $Q \subset G$. If $Q = G$ it means that the image of $G/P \to \mathcal{M}^1(X)$ is a single point, hence Γ/N has an invariant probability measure on X, hence it is amenable; this is a contradiction.

If $Q \subsetneq G$, the subgroup $N \subset G$ acts trivially on G/Q by construction (since it acts trivially on X). It follows that N is contained in Q, hence

so is its Zariski-closure $\overline{N}^{Zar}$. But N is normalized by Γ, hence $\overline{N}^{Zar}$ is normalized by G, hence N is in the center of G. $\qquad\square$

5.6.4. Theorem (Characterizations of quotients). *Suppose that Γ acts on a compact space X preserving the measure class of a probability μ. If $P \subset G$ is a minimal parabolic subgroup and there exists a Γ-equivariant map $G/P \to X$, then (X, μ) is isomorphic to G/Q for some $Q \supset P$. In particular Q is parabolic.*

Proof sketch. The original proof uses an argument involving Lebesgue density points. There is another proof, which I learned from Federico Rodriguez-Hertz, which considers the suspension space $X \times_\Gamma G$ with its P-action. It brings in also (relative) entropy. $\qquad\square$

6. Unipotent rigidity

The most complete general results are due to Ratner, in the original proof of Raghunathan's conjecture [Rat91]. Subsequent influential simplifications were introduced by Margulis–Tomanov [MT94] using entropy.

[Esk10] has a discussion of one of the first non-trivial cases beyond just the horocycle flow, for the case $\mathbf{SL}(2) \ltimes \mathbb{A}^2$.

6.1. Horocycle flow rigidity

For the next two definitions, G is an arbitrary group acting continuously on a metric space X.

6.1.1. Definition (Minimality). The action of G on X is *minimal* if there are no other closed G-invariant sets. A closed nonempty G-invariant set $Y \subset X$ such that the G-action on Y is minimal is called a *minimal set*.

When X is compact, there always exists at least one minimal $Y \subset X$. Indeed, the intersection of any number of non-empty closed G-invariant sets is still closed, G-invariant, and nonempty by compactness of X. Therefore, by Zorn's lemma, a minimal set exists.

One can think of minimality as a topological analogue of ergodicity.

6.1.2. Definition (Unique ergodicity). The action of G on X is *uniquely ergodic* if there exists a unique G-invariant probability measure on X.

This definition is best suited for amenable groups, which are guaranteed to have at least one invariant measure; we will specialize below to

$\mathbb{Z}$. When G is amenable, the support of the unique G-invariant probability measure will also be a minimal set; a smaller closed G-invariant subset would carry another probability measure.

6.1.3. Example (Examples of uniquely ergodic systems)**.**

  (i) Irrational rotations on the circle.
  (ii) More generally, translations on compact abelian groups with a dense orbit.
  (iii) The map $x \mapsto \frac{1}{2}x$ on $[0, 1]$.
  (iv) The horocycle flow on a compact quotient $\mathbf{SL}_2 \mathbb{R}/\Gamma$, to be proved below.

Recall the following concepts from §2.1:

$$S_n f = f + T^* f + \cdots + (T^*)^n f \qquad \text{Birkhoff sums}$$

$$A_n f = \frac{1}{n} S_n f \qquad \text{Birkhoff averages}$$

$$f = h - T^* h \qquad f \text{ is a coboundary}$$

6.1.4. Proposition (Equivalent characterizations of unique ergodicity)**.**
Suppose that $T \colon X \to X$ is a continuous map of a compact metric space. The following are equivalent:

  (i) *T is uniquely ergodic.*
  (ii) *The space of continuous functions decomposes as*

$$C^0(X, \mathbb{C}) = \mathbb{C} \oplus \overline{Img(\mathbf{1} - T^*)}$$

   the direct sum of constants plus closures of coboundaries (compare with von Neumann's version for L^2, Theorem 2.2.1).
  (iii) *For every continuous function f, the Birkhoff averages $A_n f$ converge* uniformly *to a constant.*
  (iv) *For every continuous function f, the Birkhoff averages $A_n f$ converge* pointwise *to a constant $C(f)$.*

There is a further characterization of unique ergodicity, related to equicontinuity. This will be alluded to below.

Proof. For the equivalence of (i) and (ii), note that the space of T-invariant measures is the orthogonal to the closure of the space of coboundaries. If μ_1, μ_2 are two distinct invariant probability measures, then their difference would vanish on $\mathbb{C} \oplus \overline{(\mathbf{1} - T^*)}$, so (ii) implies (i). Conversely, if μ is the unique invariant probability measure, then f belongs to the closure of coboundaries if and only if $\int f d\mu = 0$, from which the decomposition of f follows.

To see that (ii) implies (iii) write $f = c + h - T^*h + g_\varepsilon$ with $\|g_\varepsilon\|_{C^0} \leq \varepsilon$ to find that $\|A_n f - c\|_{C^0} \leq O(\frac{1}{n}) + \varepsilon$ for any $\varepsilon > 0$. It is clear that (iii) implies (iv) so it suffices to deduce (i) from (iv).

Suppose there were two distinct invariant ergodic probability measures μ_1, μ_2. Then there exists a continuous f such that $\int f d\mu_1 \neq \int f d\mu_2$. However, there exists a μ_i-generic point x_i such that the Birkhoff averages satisfy $A_n f(x_i) \to \int f d\mu_i$ and and since all Birkhoff averages converge to the same constant $C(f)$, this is a contradiction. $\square$

6.1.5. The setup. To simplify notation, denote $G := \mathbf{SL}_2\,\mathbb{R}$, $\Gamma \subset G$ a lattice and quotient space $X := \Gamma\backslash G$. The following subgroups:

$$A = \left\{ g_t := \begin{bmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{bmatrix} \right\} N = \left\{ h_s := \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \right\} N^- = \left\{ h_s^- := \begin{bmatrix} 1 & 0 \\ -s & 1 \end{bmatrix} \right\}$$

will play an important role, and our focus will be on the dynamics of N, called the horocycle flow.

6.1.6. Proposition (Closed horocycles). *There exists a closed orbit of N on X if and only if X is not compact, or equivalently Γ has unipotent elements. The closed N-orbits are then identified with Γ-conjugacy classes of unipotent elements in Γ, up to the action of A.*

Proof. If $x = \Gamma g \in \Gamma\backslash G$ is such that $x \cdot h_s = x$ for some $h_s \in N$, it follows that $g h_s g^{-1} \in \Gamma$; the converse is clear. This gives a surjective map from closed horocycle orbits to Γ-conjugacy classes of unipotents in Γ.

If $g_1 h_s g_1^{-1} = g_2 h_{f(s)} g_2^{-1}, \forall s \in N$ then $g_2^{-1} g_1$ normalizes the subgroup N, so up to the N-action we can assume $g_2^{-1} g_1 \in A$, which gives the second part. $\square$

Because A normalizes N, it follows that A acts on the set of N-orbits, and in particular for any $a \in A$ and a minimal set Y for N, $Y \cdot a$ is again N-minimal. In particular, either $Y \cdot a = Y$ or $(Y \cdot a) \cap Y = \emptyset$.

6.1.7. Theorem (Hedlund, minimality of the horocycle flow). *If X is compact, then the N-action is minimal.*

If X is not compact, then the only minimal sets for the N-action are either X itself, or the closed horocycle orbits.

Proof. We formulate the argument in a way similar to the proof of Proposition 4.2.3, following [McM17]. By Remark 6.1.8, it suffices to check that an N-invariant closed set $Y \subset X$ is also A-invariant. Recall that by minimality of Y, for $a \in A$ we know that $Y \cdot a$ is either equal to, or disjoint from Y; the set of $a \in A$ such that $Y \cdot a = Y$ form a closed

subgroup of A. So it suffices to check that there exists a sequence of $A \ni a_i \to 1$ such that $Ya_i = Y$.

Let now $S \subset G$ be the closed set of $g \in G$ such that $Y \cdot g \cap Y \neq \emptyset$. Then S is invariant under N both on the right, and on the left. Let $[S] \subset G/N \cong \mathbb{R}^2$ be its image in the quotient, which is a closed A-invariant set.

Assuming that Y is not a closed N-orbit, it follows that there exist $g_i \in G$ with $g_i \to 1$, and $g_i \notin N$ such that $g_i \in S$. Therefore $[S]$ has an accumulation point at $(1,0) \in \mathbb{R}^2$ (which corresponds to the identity coset). Recall also that the axis $(0, +\infty)$ corresponds to the image of A in the quotient.

If the accumulating points to $(1,0)$ in S come off the horizontal axis, then by N-invariance of $[S]$ it follows that the entire horizontal axis is in $[S]$. Otherwise, the points are coming to $(1,0)$ lying on the horizontal axis, so there exists a sequence of $a_i \to 1$ with $a_i \in A$ as desired. $\square$

6.1.8. Remark. The action of $P = AN$ on X is minimal. Indeed, this is equivalent to the minimality of Γ acting on $G/P = \mathbb{P}^1(\mathbb{R})$. This, in turn, follows because the surface has finite volume, so if there was a non-trivial Γ-invariant open set in the boundary of $\mathbb{H}^2$, one can construct an infinite-volume subset of $\Gamma\backslash\mathbb{H}^2$.

6.1.9. Theorem (Unique ergodicity of the horocycle flow). *Let* $X := \mathbf{SL}_2\,\mathbb{R}/\Gamma$ *be a compact quotient by a discrete subgroup. Then the horocycle flow* $h_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$ *acting on the left is uniquely ergodic.*

The technique in this case is via renormalization of ergodic averages, which is powerful when it applies. There is a more general setting in which this kind of proof works, but the most interesting cases of Ratner's theorems are not approachable by this method. The key point is the relation between the horocycle and geodesic flow:

$$\text{For } g_t = \begin{bmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{bmatrix} \text{ we have } g_t \cdot h_s = h_{se^t} \cdot g_t.$$

The key relation that this induces on the ergodic averages of the horocycle flow is encapsulated in the following:

6.1.10. Lemma (Renormalization of ergodic averages). *Let* $A_T f(x) := \frac{1}{T} \int_0^T f(h_s(x))ds$. *Then defining*

$$AR_t f(x) = \int_0^1 f(g_t h_s x)ds$$

we have $A_T f(x) = (AR_t f)(g_{-t}x)$ *with* $t = \log T$.

The point is that we have equated the behavior of the Birkhoff sums of h_s for large times with the behavior on a bounded interval, but combined with the g_t-action.

Proof. Compute using the definitions and the commutation relations and the notation $T = e^t$:

$$
\begin{aligned}
A_{e^t} f(x) &:= \frac{1}{e^t} \int_0^{e^t} f(h_s(x)) ds \\
&= \int_0^1 f(h_{e^t \cdot s}(x)) ds \\
&= \int_0^1 f(g_t \cdot h_s \cdot g_{-t} x) ds \\
&= A R_t f(g_{-t} x)
\end{aligned}
$$

$\square$

6.1.11. Lemma (Stability of renormalized averages)**.** *For fixed continuous f, the family of functions $\{A R_t f\}_{t \geq 0}$ is equicontinuous.*

This part of the argument uses the *opposite* horocyclic direction to establish.

Proof. Let $h_s^- = \begin{bmatrix} 1 & 0 \\ s & 0 \end{bmatrix}$ $\square$

A proof is in: [EW15].
Do also Coudene's proof. Babillot's proof for case with cusps.
[Ghy92] for non-separability of orbit spaces.

6.2. Non-divergence of unipotent orbits

Do the simplest case.

6.3. Oppenheim conjecture

The presentation follows [Ben08, §12] closely. In fact it suffices to check a weaker statement than Theorem 1.2.4, namely:

6.3.1. Theorem (Margulis)**.** *Let $X := \mathbf{SL}_3\, \mathbb{Z} \backslash \mathbf{SL}_3\, \mathbb{R}$, $H := \mathbf{SO}(2,1)$. Then any bounded[8] H-orbit in X is in fact compact, hence closed.*

---

[8]This is equivalent to: the closure of the orbit is compact.

6.3.2. Some notation. Fix the quadratic form to be

$$Q := x_2^2 - 2x_1 x_3 \quad \text{or in matrix form } Q := \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{bmatrix}.$$

The relevant diagonal subgroup will be

$$A := \left\{ a_t = \begin{bmatrix} e^t & & \\ & 1 & \\ & & e^{-t} \end{bmatrix} \right\}$$

and unipotent subgroups

$$U := \left\{ u_s := \begin{bmatrix} 1 & s & \frac{s^2}{2} \\ & 1 & s \\ & & 1 \end{bmatrix} \right\} \quad V := \left\{ v_s := \begin{bmatrix} 1 & 0 & s \\ & 1 & 0 \\ & & 1 \end{bmatrix} \right\}.$$

Note that $A, U \subset H$ but $V \cap H = e$. Let also $V^{\pm} \subset V$ be the semigroups of elements u_s with $s > 0$ (resp. $s < 0$).

The proof of Theorem 1.2.4 will follow from the following three results. Throughout, assume that $x \cdot H \subset X$ is a bounded H-orbit and let $F := \overline{x \cdot H}$ be the orbit closure, compact by assumption.

6.3.3. Lemma (Too many unipotents). *For any $x \in X$ each of the orbits $x \cdot V^- U A$ and $x \cdot V^+ U A$ are unbounded.*

6.3.4. Lemma (*A*-invariance). *Let $K \subset F = \overline{xH}$ be a minimal U-invariant compact set. Then K is UA-invariant.*

6.3.5. Lemma (*V*-invariance). *Suppose that K is as in Lemma 6.3.4 and $\overline{x \cdot H} \neq x \cdot H$. Then*

$$\text{either } K \cdot V^+ U A \subset F \text{ or } K \cdot V^- U A \subset F$$

The ideas required for Lemma 6.3.4 and Lemma 6.3.5 are quite similar and use the unipotent nature of the dynamics, whereas Lemma 6.3.3 is a general fact.

Proof of Theorem 1.2.4. Suppose that $x \cdot H$ is a bounded but not closed orbit, i.e. $F = \overline{x \cdot H} \neq x \cdot H$. Let $K \subset F$ be a minimal U-invariant closed set. Then by Lemma 6.3.5 one of $x \cdot V^{\pm} U A$ is in F, which by Lemma 6.3.3 implies that F is not bounded; this is a contradiction. $\square$

Checking Lemma 6.3.3 is immediate. Indeed, the freedom to use two unipotent subgroups allows to make a vector have only the last coordinate non-vanishing. The diagonal action then gives the divergent subset.

Proof of Lemma 6.3.3. For any $x \in X$, let $\Lambda_x \subset \mathbb{R}^3$ denote the associated lattice. Pick a vector $w = (w_1, w_2, w_3) \in \Lambda_x$ such that $w_3 \neq 0$ and $w_2^2 - 2w_1w_3 > 0$ for the case of V^+ (or $w_2^2 - 2w_1w_3 < 0$ for V^-); such a vector always exists, since the corresponding cone contains euclidean balls of arbitrarily large radius. Then one can compute

$$
\begin{bmatrix} w_1 & w_2 & w_3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & \frac{w_2^2 - 2w_1w_3}{2w_3^2} \\ & 1 & 0 \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{-w_2}{w_3} & \frac{w_2^2}{2w_3^2} \\ & 1 & \frac{-w_2}{w_3} \\ & & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & w_3 \end{bmatrix}
$$

or in symbols $w \cdot v_{s_w} u_{t_w} = w'$ has only the last coordinate non-zero, where $s_w = \frac{w_2^2 - 2w_1w_3}{2w_3^2}$ and $t_w = \frac{-w_2}{w_3}$. It is now clear that $w' \cdot A$ contains arbitrarily short vectors, so the claim follows by Mahler's criterion. $\square$

The proof of the remaining two lemmas is based on the following property of unipotent actions.

6.3.6. Lemma (Invariant sets for linear unipotent actions). *Let $E \cong \mathbb{R}^d$ be a finite-dimensional real vector space with a 1-parameter action of a group U by unipotent matrices, on the right. Let $F \subset E$ denote the U-fixed points. Then for any subset $D \subset E \setminus F$ and $v_0 \in \overline{D} \cap F$ there exists a nonconstant polynomial path passing through v_0 and contained in $\overline{D \cdot U} \cap F$.*

Informally, the lemma is saying that in the (non-Hausdorff!) space E/U any closed set that intersects the fixed point set $F \subset E/U$ in fact intersects it along a positive-dimensional subset. Note that for unipotent actions, the fixed point set F always contains at least a line.

Proof. Take a sequence $v_n \in D$ such that $v_n \to v_0 \in F$. The orbit $v_n \cdot u_s$ is contained in $E \setminus F$ and is a non-constant polynomial function $\mathbb{R} \to E$, of uniformly bounded degree. There exists $\lambda_n \in \mathbb{R}$ such that

$$
\sup_{s \in [-1,1]} \|v_n \cdot u_{\lambda_n s} - v_0\|^2 = 1
$$

and since $v_n \to v_0$ which is a fixed point for the U-action, it follows that $\lambda_n \to \infty$ (since the action is continuous).

Let $\phi_n(s) := v_n \cdot u_{\lambda_n s}$ viewed as a polynomial map $\phi_n \colon [-1, 1] \to E$ of some uniformly bounded degree. Note that since $\sup_{[-1,1]} \|\phi_n(s) - v_0\| = 1$ and the corresponding space of polynomials is finite-dimensional, it follows that we can extract a convergent subsequence with $\phi_n(s) \to \phi(s)$ uniformly on $[-1, 1]$. Moreover

$$
\phi(0) = \lim \phi_n(0) = v_0 \quad \text{and} \quad \sup_{s \in [-1,1]} \|\phi(s) - v_0\| = 1
$$

since the same is true before the limit. In particular ϕ is non-constant.

Now it is clear that $Img(\phi) \subset \overline{D \cdot U}$ by construction and moreover the image consists of U-fixed points since

$$\phi(s)u_t = \lim \phi_n(s)u_t = \lim v_0 \cdot u_{\lambda_n s}u_t$$
$$= \lim v_0 \cdot u_{\lambda_n\left(s+\frac{t}{\lambda_n}\right)}$$
$$= \lim \phi_n\left(s + \frac{t}{\lambda_n}\right)$$
$$= \phi(s)$$

since $\lambda_n \to \infty$ and the convergence $\phi_n \to \phi$ is uniform.   □

Before starting the proof of the remaining two lemmas, a few observations:

(i) By Chevalley's Theorem 3.3.7 for the two subgroups H, U there exist representations E_H, E_U of $\mathbf{SL}_3\,\mathbb{R}$ and vectors $v_\bullet \in E_\bullet$ such that the corresponding subgroups are defined as the stabilizers of $v_\bullet$.

For H take E_H to be the space of quadratic forms on $\mathbb{R}^3$, with vector v_H corresponding to the quadratic form Q such that $H = \mathbf{SO}(Q)$. For U take $E_U := E_H \oplus \mathbb{R}^3$ (note that $U \subset H$) with $v_U = v_H \oplus \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$.

(ii) The normalizer of U in G is $N_G(U) = UVA$ and so because $U\backslash G \hookrightarrow E_U$ using the vector v_U, it follows that the U-fixed points in E_U contain the VA-orbit.

Proof of Lemma 6.3.4. Let $K \subset F$ be a minimal U-invariant set. Note that because of minimality, for any $g \in N_G(U)$ normalizing U it follows that either $K \cdot g = K$ or $K \cdot g \cap K = \emptyset$ since $K \cdot g$ is also U-invariant and minimal. Finally, note that K cannot be a closed U-orbit because otherwise under the A-action it will be unbounded, but $K \cdot A \subset F$.

Let now

$$T_K := \{g \in G \colon K \cdot g \cap K \neq \emptyset\}$$

which is a closed, left & right U-invariant subset of G. Since as a U-orbit, K is not closed, it follows that there exist $g_i \in T_K \setminus U$ with $g_i \to e$. Indeed, fix $x \in K$ and y such that $y \notin xU$, but by minimality there exist $u_i \in U, g_i \in G \setminus U$ with $g_i \to e$ such that $y \cdot g_i = xu_i$.

Set $L := UVA \cap T_K$; by an earlier remark K is preserved by any element of L, since UVA is the normalizer of U, hence L is a closed subgroup. Our goal is to show L contains A; we already know $U \subset L$.

Now the image of T_K in $U\backslash G \subset E_U$ is a closed, right U-invariant subset. The right U-fixed points in $U\backslash G$ coincide with $U\backslash N_G(U) =$

$U\backslash UVA$. For the sequence $g_i \to e$ constructed above, there are two possibilities:

- either for all sufficiently large i, $g_i \in UVA \setminus U$
- or $g_i \notin UVA$ along a subsequence

Either way it follows that $U\backslash L$ contains a positive-dimensional subset, in the first case by assumption and in the second by Lemma 6.3.6.

Now if L contains V or one of its A conjugates then by Lemma 6.3.3 the set xH would be unbounded, so the only possibility is that L contains A. $\qquad\square$

The next proof follows a similar strategy.

Proof of Lemma 6.3.5. The set K is UA-invariant and closed, $xH \neq \overline{xH}$, and $UA\backslash H$ is compact, so it follows that K cannot be contained in xH. Indeed, there exist $h_i \in H$ such that xh_i accumulates outside xH, so in the coset decomposition can assume $h_i \in UA\backslash H$ converges, so it follows that there is some sequence $u_i a_i \in UA$ such that $xu_i a_i$ converges outside xH.

Define
$$T_{F,K} := \{g \in G \colon Fg \cap K \neq \emptyset\}$$
which is a closed, left-H and right-AU-invariant. There exists a sequence $g_n \in T_{F,K} \setminus H$ converging to e. The justification is similar to the previous: take $y \in K \setminus xH$ and $h_n \in H$ such that $xh_n = yg_n$ with $g_n \to e$.

As before there is an embedding $H\backslash G \hookrightarrow E_H$ and the fixed points for the right U-action are $H\backslash HN_G(U) = H\backslash HUAV \cong V$.

There are again two possibilities for the sequence $g_i \to e$:

- either $g_i \in HV$ for all sufficiently large i
- or $g_i \notin HV$ along a subsequence

In either case the image of $T_{F,K}$ in $H\backslash G$ will contain a sequence accumulating to the identity in $H\backslash HV$, in the second case by Lemma 6.3.6.

Now recall that $T_{F,K}$ was also right A-invariant, and since V is contracted under an appropriate A-conjugation it follows that $T_{F,K}$ will contain either V^+ or V^-, say it's V^+. So $\forall v \in V^+$ we have $Fv \cap K \neq \emptyset$, but the intersection is also U-invariant hence equals K by minimality of K. It follows that $\forall v \in V^+$ we have $Kv \subset F$, and so $KV^+UA \subset F$, which contradicts Lemma 6.3.3. $\qquad\square$

7. Entropy

For a thorough introduction to the subject see [ELW].

7.1. Shannon entropy

The notion of entropy in the context of information theory was introduced by Shannon [Sha48]. The story of the naming is (according to Tribus & McIrvine) quoting Shannon:

> My greatest concern was what to call it. I thought of calling it "information", but the word was overly used, so I decided to call it "uncertainty". When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, "You should call it entropy, for two reasons: In the first place your uncertainty function has been used in statistical mechanics under that name, so it already has a name. In the second place, and more important, nobody knows what entropy really is, so in a debate you will always have the advantage."

7.1.1. Definition (Entropy of a discrete random variable). Let X be a random variable taking finitely many values, i.e. X is a map $X \colon (S, \mu) \to \{1, \ldots, n\}$ from an unspecified probability space (S, μ) to a finite[9] set. The probability that $X = i$ is denoted $\mathbb{P}(X = i)$ and equals $\mu\left(X^{-1}(i)\right)$ The entropy of X is defined[10] by:

$$H(X) := \sum_i \mathbb{P}(X = i) \log \frac{1}{\mathbb{P}(X = i)}$$

Alternatively, if $p_\bullet = (p_1, \ldots, p_n)$ is a probability vector (i.e. $\sum_i p_i = 1$) then

$$H(p_\bullet) := \sum_i p_i \log \frac{1}{p_i}$$

The language of random variables is convenient when there is some independence or correlation properties between several random variables. Throughout a definition or argument using random variables, the same unspecified probability space is used. Throughout, for convenience $[n]$ denotes $\{1, \ldots, n\}$. By convention (and continuity), $x \log \frac{1}{x}$ equals 0 for $x = 0$.

---

[9]A countable set is allowed, but then entropy can be $+\infty$.
[10]By convention, take log to be in base e.

7.1.2. Definition (Conditional entropy)**.** For two random variables X, Y the conditional entropy is defined by

$$H(X|Y) := \sum_j \mathbb{P}(Y = j) \cdot H(X|_{Y=j})$$

$$= \sum_j \mathbb{P}(Y = j) \cdot \sum_i \mathbb{P}(X = i|Y = j) \cdot \log \frac{1}{\mathbb{P}(X = i|Y = j)}$$

where the conditional probability is defined by

$$\mathbb{P}(X = i|Y = j) := \frac{\mathbb{P}(X = i, Y = j)}{\mathbb{P}(Y = j)}$$

7.1.3. Proposition (Basic properties of Shannon entropy)**.** *Fix random variables* $X, Y, Z, X_1, \ldots, X_n$.

(i) Maximal entropy: *If X takes n values then*

$$H(X) \le \log n$$

with equality if and only if X takes each value with the same probability.

(ii) Monotonicity under conditioning:

$$H(X|Y) \le H(X)$$

with equality if and only if X and Y are independent.

(iii) Chain rule: *Viewing (X, Y) as a single random variable, we have:*

$$H(X, Y) = H(Y) + H(X|Y)$$

and more generally

$$H(X_1, \ldots, X_n) = H(X_1) + H(X_2|X_1) + \cdots + H(X_n|X_i : i < n)$$

(iv) Subadditivity:

$$H(X, Y) \le H(X) + H(Y)$$

and more generally

$$H(X_1, \ldots, X_n) \le H(X_1) + \cdots + H(X_n)$$

Proof. The inequalities follow from Jensen's inequality applied to the concave function $x \log \frac{1}{x}$ and the equalities are a straightforward verification of the definitions. $\square$

The following "covering" result is a useful extension of the subadditivity property.

7.1.4. Lemma (Shearer). *Suppose that $\mathcal{F}$ is a family of subsets of $[n]$, such that each element in $[n]$ is contained[11] in at least k members of $\mathcal{F}$. Then for random variables X_i we have*

$$H(X_1, \ldots, X_n) \leq \frac{1}{k} \sum_{F \in \mathcal{F}} H(X_f \colon f \in F)$$

Proof. We will use the monotonicity result from Proposition 7.1.3: entropy decreases when we condition on more variables. Denote for convenience for a set $S \subset [n]$ the random variable $X_S = (X_s \colon s \in S)$. Then for an element $F \in \mathcal{F}$ with $F = \left\{ i_1 < \cdots < i_{|F|} \right\}$ we have

$$H(X_F) = H(X_{i_1}) + H(X_{i_2}|X_{i_1}) + \cdots + H(X_{i_{|F|}}|X_{i_1}, \ldots, X_{i_{|F|-1}})$$
$$\geq H(X_{i_1}|X_{[i_1-1]}) + H(X_{i_2}|X_{[i_2-1]}) + \cdots + H(X_{i_F}|X_{[i_{|F|}-1]})$$

where as before $[i] = [1, \ldots, i]$. Summing the above inequality over all $F \in \mathcal{F}$ and using the k-fold covering property of $\mathcal{F}$ gives

$$\sum_{F \in \mathcal{F}} H(X_F) \geq k \cdot \sum_{i \in [n]} H(X_i|X_{[i-1]})$$
$$= k \cdot H(X_1, \ldots, X_n)$$

where the last line follows from a successive application of the chain rule. □

7.1.5. Corollary (Loomis–Whitney inequality). *Suppose that $A \subset \mathbb{Z}^3$ is a finite set. Denote by $A_x, A_y, A_z \subset \mathbb{Z}^3$ the projections of A along each of the three coordinate axes to the plane. Then*

$$|A|^2 \leq |A_x| \cdot |A_y| \cdot |A_z|$$

Proof. View A as a probability space with uniform measure of each point equal to $\frac{1}{|A|}$. Consider the random variables $X, Y, Z : A \to \mathbb{Z}$ which give the respective coordinates. Then $H(X, Y, Z) = \log |A|$ since the measure is uniform. For individual pairs we have

$$H(X, Y) \leq \log |A_z|$$

since the projected measure might not be uniform (and similarly for other coordinates). Applying Shearer's Lemma 7.1.4 with the obvious cover of $\{X, Y, Z\}$ by three two-element sets gives

$$2 \log |A| = 2 \cdot H(X, Y, Z) \leq H(X, Y) + H(Y, Z) + H(Z, X)$$
$$\leq \log |A_x| + \log |A_y| + \log |A_z|$$

which is the desired inequality. □

See Exercise 9.3.1 for an application.

---

[11]Multiplicities are allowed in $\mathcal{F}$ and also the covering.

7.1.6. Remark. The Loomis–Whitney inequality has many incarnations and variations. Here is one: for measure spaces S_1, S_2, S_3 and functions $f_{i,j} \in L^2(S_i \times S_j)$ we have

$$\|f_{1,2} \cdot f_{2,3} \cdot f_{3,1}\|^2_{L^1(S_1 \times S_2 \times S_3)} \leq \|f_{1,2}\|^2_{L^2(S_{1,2})} \|f_{2,3}\|^2_{L^2(S_{2,3})} \|f_{3,1}\|^2_{L^2(S_{3,1})}$$

A proof is by successive applications of Cauchy–Schwartz, and variants of the above statement hold for more functions and other L^p spaces, with proofs using Hölder's inequality.

7.2. Kolmogorov–Sinai entropy

Throughout (X, μ) is a fixed Borel probability measure space.

7.2.1. Some basic concepts for entropy of partitions. A *partition* ξ of X is a decomposition into measurable disjoint sets

$$X = \coprod A_i$$

with each A_i called an atom of ξ. When ξ has countably many elements, its *entropy* relative to the measure μ is defined as

$$H_\mu(\xi) := \sum_i \mu(A_i) \log\left(\frac{1}{\mu(A_i)}\right)$$

For $x \in X$ denote by $[x]_\xi$ the atom of ξ that contains x.

Then the *information function* is

$$I_\mu(\xi) \colon X \to \mathbb{R}$$

$$I_\mu(\xi)(x) = \log \frac{1}{\mu([x]_\xi)}$$

With this notation the entropy can be computed as

$$H_\mu(\xi) = \int_X I_\mu(\xi)\, d\mu.$$

7.2.2. Operations on partitions. The partitions of X form a lattice in the sense of set theory, i.e. they form a partially ordered set with two operations $\wedge, \vee$. Given two partitions $\xi = \{A_i\}, \eta = \{B_j\}$ of X, we say $\xi \preceq \eta$ if any element of ξ is a union of elements of η. The most important operations are

Join: $\xi \vee \eta := \{A_i \cap B_j\}$ which is the smallest partition larger than both of ξ, η.

Meet: $\xi \wedge \eta$ is the largest partition which is smaller than both of ξ, η.

The join $\xi \vee \eta$ is also called the *common refinement*.

7.2.3. Conditional entropy and information. Given a set $S \subset X$ the associated conditional measure is defined by

$$\mu|_S(A) := \frac{\mu(A \cap S)}{\mu(S)}$$

assuming that $\mu(S) \neq 0$.

Given two partitions ξ, η define the conditional information function by

$$I_\mu(\xi|\eta) := I_\mu(\xi \vee \eta) - I_\mu(\eta)$$

and the conditional entropy by

$$H_\mu(\xi|\eta) := \int_X I_\mu(\xi|\eta) \, d\mu$$

Two partitions ξ, η will be called *independent* if

$$\mu(A_i \cap B_j) = \mu(A_i) \cdot \mu(B_j) \quad \forall A_i \in \xi, B_j \in \eta.$$

Analogously to Proposition 7.1.3 we have:

7.2.4. Proposition (Basic properties of measure-theoretic entropy)**.**
Fix countable partitions $\xi, \eta, Z, X_1, \ldots, X_n$.

(i) Maximal entropy: *If ξ has n atoms then*

$$H_\mu(\xi) \leq \log n$$

with equality if and only if each atom has the same probability.

(ii) Monotonicity under conditioning:

$$H_\mu(\xi|\eta) \leq H_\mu(\xi)$$

with equality if and only if ξ and η are independent. More generally for $\eta_1 \succeq \eta_2$

$$H_\mu(\xi|\eta_1) \leq H_\mu(\xi|\eta_2)$$

(iii) Chain rule:

$$H_\mu(\xi \vee \eta) = H_\mu(\xi) + H_\mu(\xi|\eta)$$

and more generally

$$H_\mu(\xi_1 \vee \ldots \vee \xi_n) = H_\mu(\xi_1) + H_\mu(\xi_2|\xi_1) + \cdots + H_\mu(\xi_n|\xi_i : i < n)$$

(iv) Subadditivity:

$$H_\mu(\xi \vee \eta) \leq H_\mu(\xi) + H_\mu(\eta)$$

and more generally

$$H_\mu(\xi_1 \vee \ldots \vee \xi_n) \leq H_\mu(\xi_1) + \cdots + H_\mu(\xi_n)$$

7.2.5. Introducing Dynamics. Suppose now that $T : X \to X$ is a measurable transformation preserving μ, i.e. $T_*\mu = \mu$. Given a partition $\xi = \{A_i\}$ define

$$T^{-1}\xi := \left\{T^{-1}A_i\right\}$$

to be the preimage partition. It is clear that

$$H_\mu(T^{-1}\xi) = H_\mu(\xi) \text{ and } H_\mu(T^{-1}\xi|T^{-1}\eta) = H_\mu(\xi|\eta).$$

7.2.6. Proposition (Defining entropy). *The sequence*

$$a_n := H_\mu(\xi \vee T^{-1}\xi \cdots \vee T^{-(n-1)}\xi)$$

satisfies $a_n + a_m \leq a_{n+m}$ and therefore the limit

$$h_\mu(T,\xi) := \lim_n \frac{1}{n} H_\mu(\xi \vee T^{-1}\xi \cdots \vee T^{-(n-1)}\xi)$$

exists and equals its infimum (by Exercise 9.1.4).

Proof. The proposition follows by subadditivity of entropy and the invariance of the measure:

$$
\begin{aligned}
H_\mu(\xi \vee \cdots \vee T^{-(n+m-1)}\xi) &\leq \\
&\leq H_\mu(\xi \vee \cdots \vee T^{-(n-1)}\xi) + H_\mu(T^{-n}\xi \vee \cdots \vee T^{-(n+m-1)}\xi) \\
&= H_\mu(\xi \vee \cdots \vee T^{-(n-1)}\xi) + H_\mu(\xi \vee \cdots \vee T^{-(m-1)}\xi)
\end{aligned}
$$

$\square$

7.2.7. Definition (Kolmogorov–Sinai entropy). The entropy of $T \colon X \to X$ with respect to the measure μ is

$$h_\mu(T) := \sup_{\xi \colon H_\mu(\xi) < \infty} h_\mu(T,\xi)$$

7.2.8. Definition (Partitions and σ-algebras). Denote the Borel σ-algebra on X by $\mathcal{B}$.

- For a partition ξ let $\sigma(\xi)$ denote the coarsest σ-algebra contained in $\mathcal{B}$ for which the atoms of ξ are measurable.
- Two σ-algebras $\mathcal{B}_1, \mathcal{B}_2$ are said to be *equivalent* mod μ if for any $B_1 \in \mathcal{B}_1$ there exists $B_2 \in \mathcal{B}_2$ such that $\mu(B_1 \triangle B_2) = 0$, and the same with the roles of $\mathcal{B}_1, \mathcal{B}_2$ reversed.
- A partition ξ is a *one-sided generator* if

$$\sigma\left(\bigvee_{i \geq 0} T^{-i}\xi\right) = \mathcal{B} \mod \mu$$

- A partition ξ is a *two-sided generator* if T is invertible and

$$\sigma\left(\bigvee_{i=-\infty}^{i=\infty} T^i\xi\right) = \mathcal{B} \mod \mu$$

7.2.9. Theorem (Kolmogorov–Sinai). *If the partition ξ is a one-sided generator then*

$$h_\mu(T) = h_\mu(T,\xi)$$

If T is invertible and ξ is a two-sided generator, then the same holds.

Note that if an invertible transformation has a one-sided generator, then it has zero entropy (see Corollary 7.2.11).

Key to the above and later results are the following properties.

7.2.10. Proposition (Basic properties of Kolmogorov–Sinai entropy). *Fix partitions ξ, η of X.*

- Subadditivity:

$$h_\mu(T,\xi \vee \eta) \leq h_\mu(T,\xi) + h_\mu(T,\eta)$$

- Continuity bound:

$$h_\mu(T,\xi) \leq h_\mu(T,\eta) + H_\mu(\xi|\eta)$$

- Invariance under iterate refinement: *For any $k \geq 1$*

$$h_\mu(T,\xi) = h_\mu\left(T, \bigvee_{i=1}^{k} T^{-i}\xi\right)$$

- Past and Future, finite time: *If T is invertible then for any $k \geq 1$*

$$h_\mu(T,\xi) = h_\mu\left(T, \bigvee_{i=-k}^{k} T^{-i}\xi\right) = h_\mu(T^{-1},\xi)$$

- Entropy of iterates: *For any $k \in \mathbb{Z}$*

$$h_\mu(T^k) = |k| \cdot h_\mu(T)$$

  where for $k < 0$ assume that T is invertible.
- Entropy is the new information, when the past is known: *If T is invertible, then*

$$h_\mu(T,\xi) = \lim_n H_\mu(\xi|T\xi \vee \cdots \vee T^n\xi)$$

Proof. For all properties of h_μ it suffices to check it at the level of finite partitions.

Subadditivity follows immediately from the same property for partitions.

For continuity, apply the formula for conditional entropy on partitions, combined with monotonicity under dropping variables:

$$H_\mu(\xi \vee \cdots T^{-n}\xi) =$$
$$= H_\mu(\eta \vee \cdots \vee T^{-n}\eta) + H_\mu(\xi \vee \cdots \vee T^{-n}\xi | \eta \vee \cdots \vee T^{-n}\eta)$$
$$\leq H_\mu(\eta \vee \cdots \vee T^{-n}\eta) + \sum_{j=0}^{n} H_\mu(T^{-j}\xi | \eta \vee \cdots \vee T^{-n}\eta)$$
$$\leq H_\mu(\eta \vee \cdots \vee T^{-n}\eta) + \sum_{j=0}^{i} H_\mu(T^{-j}\xi | T^{-j}\eta)$$
$$\leq H_\mu(\eta \vee \cdots \vee T^{-n}\eta) + n \cdot H_\mu(\xi | \eta)$$

That $h_\mu(T) = h_\mu(T^{-1})$ follows from the finite time version, which itself follows from invariance:

$$H_\mu(\xi \vee \cdots \vee T^{-k}\xi) = H_\mu\left(T^m(\xi \vee \cdots \vee T^{-k}\xi)\right)$$

for any $m \in \mathbb{Z}$. The formula for entropy of iterates for positive and negative values follows from knowing it for T and T^{-1} and the definitions.

To check the formula for entropy, conditioned on the past, it is best to do the proof using conditional entropy relative to a σ-algebra, but for now just write

$$H_\mu(\xi \vee \cdots \vee T^{-n}\xi) = H_\mu(\xi) + H_\mu(T^{-1}\xi | \xi) + \cdots + H_\mu(T^{-n}\xi | \xi \vee \cdots T^{-n+1}\xi)$$
$$= H_\mu(\xi) + H_\mu(\xi | T\xi) + \cdots + H_\mu(\xi | T\xi \vee \cdots \vee T^n\xi)$$

Since the terms appearing above are monotonically decreasing it follows that

$$h_\mu(T, \xi) = \lim_n H_\mu(\xi | T\xi \vee \cdots \vee T^n\xi)$$

$\square$

7.2.11. Corollary (Zero entropy if the past determines the future). *Suppose that T is invertible and has a one-sided finite entropy generator ξ. Then $h_\mu(T) = 0$.*

Proof. This follows from the formula

$$h_\mu(T, \xi) = \lim_n H_\mu(\xi | T\xi \vee \cdots \vee T^n\xi)$$

and the fact that the partition that's conditioned is arbitrarily fine (approaches $\mathcal{B}$) so the conditional entropy approaches 0. $\square$

The σ-algebra appearing in the statement encodes the past trajectory of a point, as observed by the partition ξ. So the assumption is saying that knowing the past of a point determines the point, and hence also the future of the point.

Proof sketch of Theorem 7.2.9. Suppose that ξ is a generating partition. Then for any other finite partition η and prescribed $\varepsilon > 0$ there exists a k such that $H(\eta|\xi \vee \cdots \vee T^{-k}\xi) \leq \varepsilon$, hence by the continuity property of entropy in Proposition 7.2.10 the result follows. □

7.3. Conditional measures and Entropy

Fix for this section a standard Borel space $(X, \mathcal{B}, \mu)$, with $\mathcal{B}$ the Borel σ-algebra and μ a probability measure. We will be interested in sub-σ-algebras $\mathcal{A} \subseteq \mathcal{B}$, the simplest example coming from measurable maps $X \to Y$ and taking the preimage of the σ-algebra on Y.

7.3.1. Definition (Countably generated σ-algebra). A σ-algebra $\mathcal{A}$ is *countably generated* if there exist countably many sets $A_i \in \mathcal{A}$ such that the smallest σ-algebra containing all the A_i is $\mathcal{A}$.

The *atoms* of a countably-generated σ-algebra $\mathcal{A}$ are the smallest sets belonging to $\mathcal{A}$, specifically for $x \in X$ define

$$[x]_{\mathcal{A}} := \bigcap_{A_i \,:\, x \in A_i} A_i$$

where $\{A_i\}$ are a countable list of generators. Note that $[x]_{\mathcal{A}}$ is in $\mathcal{A}$ and is the smallest element of $\mathcal{A}$ containing x. The atoms form a partition of X.

Many interesting σ-algebras are not countably generated (see Exercise 9.3.2).

7.3.2. Partitions vs. σ-algebras. This point of view is further expanded in [CK12]. Let $\mathcal{P}$ denote the set of all partitions of X into measurable sets, and $\mathcal{S}$ the set of all sub-σ-algebras of $\mathcal{B}$. There are natural maps

$$\mathcal{P} \underset{\alpha}{\overset{\sigma}{\rightleftarrows}} \mathcal{S}$$

with $\sigma(\xi)$ as the coarsest σ-algebra in $\mathcal{B}$ such that all atoms of ξ are $\sigma(\xi)$-measurable (see Definition 7.2.8).

The map from σ-algebras to partitions is a bit more delicate. If $\mathcal{A}$ is countably generated, then $\alpha(\mathcal{A})$ is the partition into the atoms of $\mathcal{A}$ (see Definition 7.3.1). In general, there exists (by [EL10, Prop. 5.8]) a countably generated σ-algebra $\tilde{\mathcal{A}}$ which is equivalent mod μ to $\mathcal{A}$ and one takes $\alpha(\tilde{\mathcal{A}})$. The resulting partition is independent of choice of $\tilde{\mathcal{A}}$, up to an appropriate mod μ equivalence on partitions.

7.3.3. Proposition (Conditional expectations). *Let $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra and $1 \leq p \leq \infty$. There exists a* conditional expectation

operator

$$\mathbf{E}_\mu(\bullet|\mathcal{A})\colon L^p(X,\mathcal{B},\mu) \to L^p(X,\mathcal{A},\mu)$$

defined uniquely by the condition

$$\int_A \mathbf{E}_\mu(f|\mathcal{A})\,d\mu = \int_A f\,d\mu \quad \forall A \in \mathcal{A}, \forall f \in L^p(X,\mathcal{B},\mu).$$

To construct the conditional expectation, define it first on L^2 using orthogonal projections and check that the defining requirement extends it continuously to all other L^p spaces. For further properties of conditional expectation see [EL10, §4].

7.3.4. Proposition (Conditional measures, [EL10, Thm. 5.9])**.** *Let $\mathcal{A} \subset \mathcal{B}$ be a sub-σ-algebra. Then there exists a full measure set $X' \subset X$ and $\forall x \in X'$ probability measures $\mu_x^{\mathcal{A}}$, depending measurably on x, such that:*

- Compatibility with expectations: *For $f \in L^1(X,\mathcal{B},\mu)$ we have*

$$\mathbf{E}_\mu(f|\mathcal{A})(x) = \int_X f\,d\mu_x^{\mathcal{A}}$$

  for a.e. $x \in X'$.
- Compatibility with equivalence: *If two σ-algebras $\mathcal{A}, \mathcal{A}'$ are equivalent $\mod \mu$ then $\mu_x^{\mathcal{A}} = \mu_x^{\mathcal{A}'}$ for a.e. x.*
- Support on atoms: *If $\mathcal{A}$ is countably generated then for a.e. x:*

$$\mu_x^{\mathcal{A}}([x]_{\mathcal{A}}) = 1 \text{ and } \mu_x^{\mathcal{A}} = \mu_y^{\mathcal{A}} \text{ if } [x]_{\mathcal{A}} = [y]_{\mathcal{A}}.$$

7.4. Local properties of entropy

Fix an ergodic probability measure preserving system $T\colon (X,\mu) \to (X,\mu)$.

7.4.1. Theorem (Shannon–McMillan–Breiman)**.** *If ξ is a finite entropy partition of X and $\xi_n := \xi \vee \cdots \vee T^{-(n-1)}\xi$ then*

$$\frac{1}{n} \log \frac{1}{\mu\left([x]_{\xi_n}\right)} \to h_\mu(T,\xi)$$

for μ-a.e. $x \in X$, and in $L^1(X,\mu)$.

7.4.2. Definition (Bowen balls)**.** When X is a metric space, define

$$B(x,\varepsilon,n) = \Big\{y\colon \operatorname{dist}(T^i x, T^i y) \le \varepsilon, \forall i = 0\ldots n-1\Big\}$$

to be the set of all points which are within ε of x up to time n.

7.4.3. Theorem (Brin–Katok local entropy formula)**.** *If $B(x, \varepsilon, n)$ denotes a Bowen ball as above, then*

$$\sup_{\varepsilon > 0} \lim_{n \to \infty} \frac{1}{n} \log \frac{1}{\mu(B(x, \varepsilon, n))} = h_\mu(T)$$

for μ-.a.e x.

8. Abelian rigidity

8.1. Furstenberg's $\times 2 \times 3$ Theorem

Let $S^1 = \mathbb{R}/\mathbb{Z}$ be the unit circle and $M_k \colon S^1 \to S^1$ denote the multiplication by k map. To fix notation, let $A = \langle M_2, M_3 \rangle$ denote the semigroup generated by the two transformations of interest.

8.1.1. Theorem (Furstenberg $\times 2 \times 3$)**.** *Suppose that $K \subset S^1$ is a closed A-invariant set. Then either K is a finite union of rational points, or K is all of S^1.*

Proof. The proof will consist of two steps.

First, if K contains an accumulation point at $0 \in S^1$, or at any other rational point, then K is all of S^1 by Proposition 8.1.2. To produce a rational accumulation point, suppose K contains an irrational point x. Then there is an A-minimal set $M \subset K$ in the closure $\overline{Ax}$ and by Proposition 8.1.4 M contains a rational point, to which a subsequence in Ax accumulates. □

8.1.2. Proposition (Accumulations at 0)**.** *Suppose that $K \subset S^1$ is a closed A-invariant set which accumulates to 0, or to any other rational point. Then $K = S^1$.*

Proof. The difference between a rational point and 0 is not important for the proof (just pass to a subsemigroup of A that will fix the rational point).

Let $S := \operatorname{Log} A \subset \mathbb{R}$, so that by Lemma 8.1.3 $\frac{a_{i+1}}{a_i} \to_{i \to \infty} 1$ with $\{\ldots < a_i < a_{i+1} \ldots\} = A$ the ordering of A; the possibility that $S \subset p \cdot \mathbb{Z}$ is excluded since $2, 3$ are relatively prime.

It now follows that for any sequence $K \ni x_i \to 0$, the orbits Ax_i become dense in a fixed (perhaps one-sided) neighborhood of 0. □

8.1.3. Lemma (Shrinking gaps for additive semigroups)**.** *Suppose that $S \subset \mathbb{R}_{>0}$ is a discrete finitely generated additive semigroup, i.e. $S + S \subset S$ and S is discrete. Then either there exists $p \in \mathbb{R}$ such that $S \subset p \cdot \mathbb{Z}$*

or, writing the elements of S in order:

$$S = \{s_1 < \ldots < s_n < \ldots\}$$

we have

$$\lim_n |s_{n+1} - s_n| \to 0.$$

One can dispense with the finitely-generated assumption and take rational, but ε-dense orbits in the proof below.

Proof. Assume that $s_1 = 1$ and consider the projection $\pi(S) \subset \mathbb{R}/\mathbb{Z}$ of the semigroup. It is either discrete (in which case $S \subset \frac{1}{N}\mathbb{Z}$ for some $N > 0$) or dense. This implies that there exists $s_0 \in S$ which is irrational, so at least we have $\langle \mathbb{N}, \mathbb{N} \cdot s_0 \rangle \subset S$. For a given $\varepsilon > 0$ let n be such that the projections $s_0, 2s_0, \ldots, ns_0$ are ε-dense on the circle. Then it is clear that $s_0 + \mathbb{N}, 2s_0 + \mathbb{N}, \ldots, ns_0 + \mathbb{N}$ are ε-dense in the segment $[ns_0 + 2, \infty)$. $\qquad\square$

8.1.4. Proposition (Disjointness and minimal sets). *Suppose $M \subset S^1$ is a closed, minimal A-invariant set.*

(i) *If K is any other A-invariant closed set and $M + K = S^1$ then $K = S^1$.*

(ii) *The minimal set M contains a rational point, hence consists of finitely many rational points.*

Proof. The more delicate is part (i). To see (ii) assuming (i) note that $(-M)$ is also a closed A-invariant set, and if M is not discrete then $M - M$ contains accumulation points at 0, hence by Proposition 8.1.2 is equal to S^1. It follows that $(-M) = S^1$ which contradicts its minimality.

To prove (i), the technique is to "thin out" the minimal set M while maintaining the equation $M + K = S^1$; in the limit M can be reduced to a single point, hence implying $K = S^1$.

For $n \in \mathbb{N}$ let $S^1[n]$ denote the n-torsion points on S^1 and $A[n] \subset A$ be the subsemigroup which acts trivially on $S^1[n]$. Take a sequence $n_i \to \infty$ whose elements are all relatively prime to elements in A, so that $A[n]$ has finite index in A and such that $A[n_i] \supset A[n_{i+1}]$ and $\cap A[n_i] = 1$. For example $n_i = 5^i$ works.

Freeze for the moment $n = n_i$.
Step 1: Finding a thinner minimal set. Then M might no longer be $A[n]$-minimal, so let $M[n] \subset M$ be an $A[n]$-minimal subset. Pick finitely many representatives $s_i \in A$ for all cosets in $A/A[n]$. Then $\cup s_i \cdot M[n] = M$ since the union is clearly closed and A-invariant. Using

now the assumption it follows that

$$S^1 = \bigcup_i (s_i \cdot M[n] + K)$$

But from the Baire category theorem, one of the finitely many closed sets on the right must contain an open set, which by relabeling we can assume is $M[n] + K$ (all that's needed is that $M[n]$ is $A[n]$-minimal). But $M[n] + K$ is also $A[n]$-invariant, hence it is all of S^1.

Step 2: Establishing approximate density. Let $\tau_n \in S^1[n]$ be a torsion point and $x \in M[n]$ arbitrary. We now check that $\tau_n \in x + K$.

By construction $\tau_n = x' + k'$ for some $x' \in M[n], k' \in K$. But since $M[n]$ is $A[n]$-minimal, there exists $a_i \in A[n]$ such that $a_i x' \to x$. Recalling that $A[n]$ acts trivially on the n-torsion points gives:

$$\tau_n = x' + k' = a_i(x' + k') \to_{i \to \infty} x + k$$

where $k \in K$ is some accumulation point of $a_i k'$.

Step 3: Intersecting the minimal sets. Unfreeze $n = n_i$ and let $M[n_i] \supseteq M[n_{i+1}]$ be a nested sequence of minimal sets as in the previous step. Note that because $A[n_i] \supset A[n_{i+1}]$ the construction of $M[n_{i+1}]$ in the previous step can be arranged to have the nesting property.

Since it is a nested family of compact sets, there is some nontrivial $x \in \cap_i M[n_i]$. By construction, for any torsion point $\tau_{n_i} \in S^1[n_i]$ we have some k_{n_i} such that $x + k_{n_i} = \tau_{n_i}$, so in fact $x + K = S^1$, hence $K = S^1$. $\qquad\square$

8.1.5. Remark.

(i) In fact Furstenberg in [Fur67] proves the statement in Proposition 8.1.4 under more general assumptions: the circle is replaced by a more general torus, and A is a more general finitely-generated semigroup.

(ii) Since eventually one concludes that M consists of finitely many rational points, the sets $M[n_i]$ in the above proof eventually stabilize.

8.2. Lyons–Rudolph theorem

The main result of this section was proved by Lyons [Lyo88] and then extended by Rudolph [Rud90].

8.2.1. Theorem ($\times 2 \times 3$ weak measure classification)**.** *Suppose that μ is a probability measure on S^1, invariant by M_2.*

(i) *If μ is also invariant under translation by $\frac{1}{2}$ then μ is Lebesgue measure.*

(ii) *If μ is also invariant and ergodic under M_3, and has positive entropy under M_2, then μ is Lebesgue.*

8.2.2. Fourier coefficients of invariant measures. Suppose that μ is a M_p-invariant measure, i.e. $(M_p)_*\mu = \mu$. With the abbreviation $e_k(x) := \exp(2\pi\sqrt{-1}kx)$, the invariance of the measure gives the following information on Fourier coefficients:

$$\hat{\mu}(k) = \langle e_k, \mu \rangle = \langle e_k, (M_p)_*\mu \rangle$$
$$= \langle M_p^* e_k, \mu \rangle = \langle e_{pk}, \mu \rangle$$
$$= \hat{\mu}(pk)$$

Let U_α denote translation by α on the circle. If μ is U_α invariant, this gives for the Fourier coefficients

$$\hat{\mu}(k) = \langle e_k, \mu \rangle = \langle e_k, (U_\alpha)_*\mu \rangle$$
$$= \langle U_\alpha^* e_k, \mu \rangle = e(k\alpha) \langle e_k, \mu \rangle$$
$$= e(k\alpha)\hat{\mu}(k)$$

Proof of Theorem 8.2.1. If μ is $U_{1/2}$-invariant, then its Fourier transform satisfies

$$\hat{\mu}(k) = (-1)^k \hat{\mu}(k)$$

so all the odd Fourier coefficients vanish. But M_2-invariance gives $\hat{\mu}(k) = \hat{\mu}(2k)$, so in fact all non-zero coefficients of μ must vanish. This implies that μ is Lebesgue measure.

For part (ii), it suffices to show that μ is $U_{1/2}$-invariant. The key property of $U_{1/2}$ is that it commutes with M_3. Therefore $(U_{1/2})_*\mu$ is also an M_3-invariant measure. It suffices to check that there are two points $x, x + 1/2$ which are μ-generic (see Definition 2.1.8) for the Birkhoff theorem applied to M_3. By M_3-ergodicity of μ, this will imply that μ and $(U_{1/2})_*\mu$ in fact agree.

Now we use the positive entropy of μ for the action of M_2. Let G be the set of M_3-Birkhoff generic points (this uses the M_3-ergodicity of μ). Then G has full μ-measure and $G' := M_2^{-1}(G) \cap G$ also has full μ-measure. If $G' \cap U_{1/2}(G') = \emptyset$ then M_2 is invertible when restricted to G'.

But M_2 has a one-sided generator for the action on S^1 (and hence on G') by taking the partition $[0, 1/2), [1/2, 1)$. By Corollary 7.2.11 it follows that M_2 has zero entropy, which is a contradiction. □

8.3. S-arithmetic point of view

Some parts of this section follow [Ber16, Ch. 9], with adaptations. The main goal is Theorem 8.3.10, which is analogous to Theorem 8.2.1. The proof is a pretext to introduce a richer language and provide a geometric interpretation of some of the concepts.

8.3.1. Some rings and fields. The integers $\mathbb{Z}$ can be localized as a ring:

$$\mathbb{Z}[1/S] := \left\{ \frac{a}{(p_1 \cdots p_k)^n} : a \in \mathbb{Z}, n \in \mathbb{N} \right\}$$

with $S = \{p_1, \cdots, p_k\}$ a set of primes. The completions of $\mathbb{Q}$ for the various norms give the local fields $\mathbb{R}, \mathbb{Q}_p$, and for finite primes we also have the rings $\mathbb{Z}_p \subset \mathbb{Q}_p$. Let $|-|$ and $|-|_p$ denote the archimedean and p-adic norms on $\mathbb{Q}$, extended when possible to the relevant fields containing $\mathbb{Q}$.

For convenience of notation let

$$\mathbb{Q}_S := \mathbb{Q}_{p_1} \times \cdots \times \mathbb{Q}_{p_k}$$

where more generally, view ∞ as an allowed prime in S, with $\mathbb{Q}_\infty := \mathbb{R}$. From now on, always assume that S contains ∞, but keep the notation that $\mathbb{Z}[1/S]$ is the ring where only the finite primes are inverted.

8.3.2. The tree. Associated to each prime p there is the identification

$$\mathbb{Q}_p \Big/ \mathbb{Z}_p = \mathbb{Z}\left[\tfrac{1}{p}\right] \Big/ \mathbb{Z} =: T_p$$

where T_p carries the topology induced from $\mathbb{Q}_p$:

$$\operatorname{dist}\left(\frac{a}{p^n}, \frac{b}{p^m} \right) = \left| \frac{a}{p^n} - \frac{b}{p^m} \right|_p = p^{-\operatorname{ord}_p\left(\frac{a}{p^n} - \frac{b}{p^m} \right)}$$

There is a natural filtration

$$T_p^{(N)} := \left(\tfrac{1}{p^N}\mathbb{Z}\right)\Big/\mathbb{Z} \quad \text{with } T_p^{(N)} \subset T_p^{(N+1)} \subset \cdots$$

and moreover each $T_p^{(N)}$ is a group, as is their union T_p. The natural multiplication by p, denoted M_p, satisfies $M_p(T_p^{(N+1)}) = T_p^{(N)}$ and gives a short exact sequence of groups

$$0 \to T_p^{(1)} \hookrightarrow T_p^{(N+1)} \xrightarrow{M_p} T_p^{(N)} \to 0 \quad \forall N \geq 0$$

so that the fibers of multiplication by M_p are identified with $T_p^{(1)}$-cosets.

As a metric space T_p can be embedded isometrically[12] as the set of leaves of the following tree. At level N the set of vertices is the set of

---

[12]The lengths of the edges between levels should be such that level N is at height p^N.

orbits of $T_p^{(N)}$ acting on T_p, so that at level 0 the vertices are just T_p. Connect by an edge orbit $\mathcal{O}_N$ at level N with orbit $\mathcal{O}_{N+1}$ at level $N+1$ if $\mathcal{O}_N \subset \mathcal{O}_{N+1}$.

8.3.3. The solenoid. For now, fix $S = \{\infty, p\}$ for some prime p.

8.3.4. Proposition (The circle as factor)**.**

   (i) *The diagonal embedding $\mathbb{Z}[1/p] \hookrightarrow \mathbb{R} \times \mathbb{Q}_p$ is discrete and the quotient space*

$$X_S := \mathbb{Z}[1/p] \backslash \mathbb{R} \times \mathbb{Q}_p$$

      is compact and carries a natural finite Haar measure.

  (ii) *The space X_S carries an invertible action of multiplication by p, denoted M_p, preserving the natural Haar measure.*

 (iii) *There is a natural quotient map*

$$X_S \to X_S/\mathbb{Z}_p \xrightarrow{\sim} \mathbb{Z} \backslash \mathbb{R}$$

      which is equivariant for multiplication by p (but not by p^{-1}) and the additive action of $\mathbb{Z}[1/p]/\mathbb{Z} = T_p$, where on X_S it only acts on the $\mathbb{Q}_p$-factor.

Note that multiplication by p on $\mathbb{R} \times \mathbb{Q}_p$ is expanding in the $\mathbb{R}$ factor and contracting in the $\mathbb{Q}_p$ factor, so the situation is akin to that of a matrix in $\mathbf{SL}_2 \mathbb{Z}$ acting on the 2-torus. Note also that $\mathbb{R} \times \mathbb{Q}_p$ can be viewed as the $\mathbb{Q}_S$-points of the additive group, and the quotient is by the S-integral lattice.

Proof. Discreteness follows since if $x \in \mathbb{R}$ and $0 \neq |\frac{a_i}{p^{n_i}} - x| \to 0$ then necessarily $n_i \to \infty$. For compactness, any $(x, \frac{a}{p^k})$ with $x \in \mathbb{R}, a \in \mathbb{Z}_p$ can be brought, using a single element of $\mathbb{Z}[1/p]$, to the region where $|x| \leq 1, k \leq 0$.

Invertibility of M_p is clear since it preserves both $\mathbb{Z}[1/p]$ and $\mathbb{R} \times \mathbb{Q}_p$. It also preserves Haar measure on $\mathbb{R} \times \mathbb{Q}_p$, since Haar measure is a product of Haar measure on the two factors and M_p expands in the $\mathbb{R}$-direction by a factor of p and contracts by the same factor in the $\mathbb{Q}_p$-direction.

Finally, for identifying the quotient, recall the natural isomorphism $\mathbb{Q}_p/\mathbb{Z}_p \leftarrow \mathbb{Z}[1/p]/\mathbb{Z}$ and note the following calculation

$$\begin{aligned} X_S = \mathbb{Z}[1/p] \backslash \mathbb{R} \times \mathbb{Q}_p &\twoheadrightarrow \mathbb{Z}[1/p] \backslash \mathbb{R} \times (\mathbb{Q}_p/\mathbb{Z}_p) \\ &\xrightarrow{\sim} \mathbb{Z}[1/p] \backslash \mathbb{R} \times (\mathbb{Z}[1/p]/\mathbb{Z}) \\ &\xrightarrow{\sim} \mathbb{Z} \backslash \mathbb{R} \end{aligned}$$

where the isomorphism on the last line has comes from the map

$$\mathbb{Z}[1/p] \backslash \mathbb{R} \times (\mathbb{Z}[1/p]/\mathbb{Z}) \to \mathbb{Z} \backslash \mathbb{R}$$
$$(x, t) \mapsto x - t$$

which is well-defined. The claim about the additive action of T_p is immediate. $\qquad \square$

8.3.5. Conditional measures on p-adic leaves. The group T_p acts on $S^1 := \mathbb{Z} \backslash \mathbb{R}$, as do any of its subgroups $T_p^{(N)}$. Let ξ_N be the partition of S^1 into $T_p^{(N)}$-orbits, or equivalently the partition into fibers of the map $M_p^N \colon S^1 \to S^1$. Note that the associated σ-algebra $\sigma(\xi_N)$ is countably generated since the Borel σ-algebra on S^1 is.

Suppose now that μ is any measure on S^1. By Proposition 7.3.4 there exist conditional measures $\mu_{[x]}^N$ on S^1, independent of x in a $T_p^{(N)}$-orbit, satisfying

$$(8.3.6) \qquad \int_{S^1} f \, d\mu = \int_{S^1/M_p^N} \left(\int_{x + T_p^{(N)}} f \, d\mu_x^N \right) d(M_{p,*}^N \mu)$$

Viewed on $x + T_p^{(N)} \subset S^1$, the measure $\mu_{[x]}^N$ is independent of basepoint x, but centering the orbit at x, it gives a measure μ_x^N on $T_p^{(N)}$ which can depend on the choice of x in the orbit. Note that the measures under discussion are supported on finite sets.

8.3.7. Proposition (Normalized conditional measures on infinite leaves).

(i) *For μ-a.e. x we have $\mu_x^N(\{0\}) \neq 0$, i.e. the conditional measure is non-trivial at the center.*

(ii) *For μ-a.e. x there exist conditional measures μ_x on T_p, normalized with $\mu_x(\{0\}) = 1$ and such that*

$$\mu_x^N = \frac{1}{\mu_x(T_p^{(N)})} \mu_x \big|_{T_p^{(N)}}$$

(iii) *If $y = x + t$ with $t \in T_p$, then*

$$\mu_y \sim t_* \mu_x$$

i.e. the two measures are proportional (possibly with vanishing proportionality constant!).

Proof. For (i), setting $X_N = \left\{ x \colon \mu_x^N(\{0\}) = 0 \right\}$ by Eqn. (8.3.6) gives

$$\int_{S^1} \mathbf{1}_{X_N} \, d\mu = \int_{S^1/M_p^N} \left(\sum_{y \in x + T_p^N} \mu_{[x]}^N(y) \cdot \delta_{\mu_{[x]}^N(y) = 0} \right) d(M_{p,*}^N \mu) = 0$$

so indeed for μ-a.e. x, $\mu_x^N(\{0\}) \neq 0$.

For (ii), consider the relation between the measures $\mu_{[x]}^N$ and $\mu_{[x]}^{N+1}$. The orbit $x + T_p^{(N+1)}$ is a union of p orbits $x + T_p^{(N+1)} + t_i$ where t_i are coset representatives coming from the exact sequence

$$0 \to T_p^{(N)} \to T_p^{(N+1)} \xrightarrow{M_p^N} T_p^{(1)} \to 0.$$

It follows that

$$\mu_{[x]}^{N+1} = \sum_i \alpha_i \cdot \mu_{[x+t_i]}^N \quad \text{with } \sum \alpha_i = 1$$

by applying a chain rule to conditional expectations. For μ-a.e. $x \in S^1$ and all $N \geq 0$ we have $\mu_x^N(\{0\}) \neq 0$ so define ν_x^N on T_p^N by

$$\nu_x^N = \frac{1}{\mu_x^N(\{0\})}\mu_x^N$$

so that from the above relation we have $\nu_x^{N+1}|_{T_p^N} = \nu_x^N$ and it is clear that the limit $\mu_x := \lim_N \nu_x^N$ is well-defined on T_p and has the required properties.

For (iii) recall that $\mu_{[x]}^N = \mu_{[y]}^N$ if $y = t + x$ with $t \in T_p^{(N)}$ and note that this holds for all $N \geq N_0 = N_0(x, y)$. The proportionality of $t_*\mu_x$ and μ_y then follows from their construction in the previous part. $\quad\square$

8.3.8. Remark. Although the above discussion did not use the S-arithmetic structure, §8.3.3 is included to emphasize the analogy with two commuting Anosov automorphisms of $\mathbb{T}^3$, or with the action of the diagonal group on $\mathbf{SL}_3\,\mathbb{Z}\backslash\mathbf{SL}_3\,\mathbb{R}$.

8.3.9. Definition (Recurrent measures). A probability measure μ on S^1 is T_p-*recurrent* if for μ-a.e. x the conditional measures μ_x constructed in Proposition 8.3.7 satisfy $\mu_x(T_p) = +\infty$. This is equivalent to $\mu_{[x]}^N(\{x\}) \to 0$ as $N \to \infty$ and x is fixed, by Proposition 8.3.7(ii).

See §8.3.13 for further interpretations of recurrence, and §8.3.14 for the relation with entropy.

8.3.10. Theorem (T_p-recurrent measures are M_2-rigid). *Suppose that μ is a probability measure on S^1 invariant under M_2, and T_p-recurrent for a prime $p \neq 2$. Then μ is Lebesgue measure.*

Proof. As in the proof of Theorem 8.2.1, it suffices to check that any non-zero Fourier coefficient of μ vanishes. Fix $v \in \mathbb{Z} \setminus 0$ and recall (§8.2.2) that from M_2-invariance we have $\hat{\mu}(v) = \hat{\mu}(2^k v), \forall k \geq 0$.

Recall that $e(x) := \exp(2\pi\sqrt{-1}x)$ and set

(8.3.11) $$g_m(x) := \frac{1}{m} \sum_{0 \leq k < m} e(2^k v \cdot x)$$

which satisfies $\int g_m d\mu = \hat{\mu}(v)$ by assumption. Lemma 8.3.12 below implies that

$$\int_{S^1} \frac{|g_m|^2}{\mu_x^N(\{0\})} d\mu(x) \leq \frac{p^N}{m}$$

with $m \geq \varepsilon p^N$ (by Exercise 9.5.2) and ε depending only on v and p.

Apply now Cauchy–Schwarz to find

$$\left| \int_{S^1} g_m \, d\mu \right|^2 \leq \left(\int_{S^1} \frac{|g_m|^2}{\mu_x^N(\{0\})} \, d\mu \right) \cdot \left(\int_{S^1} \mu_x^N(\{0\}) \, d\mu(x) \right)$$

$$\leq \frac{p^N}{m} \cdot \left(\int_{S^1} \mu_x^N(\{0\}) \, d\mu(x) \right)$$

$$\leq \frac{1}{\varepsilon} \left(\int_{S^1} \mu_x^N(\{0\}) \, d\mu(x) \right) \xrightarrow{N \to \infty} 0$$

where the last convergence to 0 is from the T_p-recurrence of μ (see Definition 8.3.9) and dominated convergence. $\qquad\square$

8.3.12. Lemma (Estimating trigonometric sums using conditional measures)**.** *Suppose that v, m are such that the orbit $\{2^k v\}$ has at least m elements in $(\mathbb{Z}/p^N\mathbb{Z})^\times$. With g_m defined in Eqn. (8.3.11), we have:*

(i) *For any $x \in S^1$:*

$$\sum_{t \in T_p^{(N)}} |g_m(x+t)|^2 \leq \frac{p^N}{m}$$

(ii) *For any probability measure ν on S^1:*

$$\int_{S^1} \left(\sum_{t \in T_p^{(N)}} |g_m(x+t)|^2 \right) d\nu(x) \leq \frac{p^N}{m}$$

(iii) *For any probability measure μ on S^1 and with conditional measures μ_x^N as in Proposition 8.3.7, we have the identity*

$$\int_{S^1} \frac{|g_m|^2}{\mu_x^N(\{0\})} d\mu(x) = \int_{S^1} \left(\sum_{t \in T_p^{(N)}} |g_m(x+t)|^2 \right) d(M_{p,*}^N \mu)(x)$$

The integral appearing on the left is well-defined since $\mu_x^N(\{0\}) \neq 0$ for μ-a.e. x.

(iv) *With notation as above:*

$$\int_{S^1} \frac{|g_m|^2}{\mu_x^N(\{0\})} \, d\mu(x) \le \frac{p^N}{m}.$$

Proof. Part (i) implies (ii) by integration, part (iii) follows from a direct application of Eqn. (8.3.6), and part (iv) follows from concatenating the previous bounds. To establish (i), compute:

$$\sum_{t \in T_p^{(N)}} |g_m(x+t)|^2 = \frac{1}{m^2} \sum_{t \in T_p^{(N)}} \left(\sum_{0 \le k < m} e(2^k v(x+t)) \right) \cdot \left(\sum_{0 \le l < m} e(-2^l v(x+t)) \right)$$

$$= \frac{1}{m^2} \sum_{t \in T_p^{(N)}} \sum_{0 \le k,l < m} e\left((2^k - 2^l)vx \right) \cdot e\left((2^k - 2^l)v \cdot t \right)$$

$$= \frac{1}{m^2} \sum_{0 \le k,l < m} e\left((2^k - 2^l)vx \right) \cdot \sum_{t \in T_p^{(N)}} e\left((2^k - 2^l)v \cdot t \right)$$

$$= \frac{1}{m^2} \sum_{0 \le k,l < m} e\left((2^k - 2^l)vx \right) \cdot \delta_{k=l} \cdot p^N$$

$$\le \frac{p^N}{m}$$

where we used that $\sum_{t \in T_p^{(N)}} e(At)$ is p^N or 0, according to whether $A \equiv 0 \mod p^N$ or not. The assumption that $\{2^k v\}$ has at least m elements was used to conclude the sum vanishes for $A = (2^k - 2^l)v$ and $k \ne l$. $\square$

8.3.13. Recurrence of leaves.

8.3.14. Recurrence and Entropy.

9. Exercises

9.1. General Dynamics

9.1.1. Exercise. Let X be a compact metric space and $T : X \to X$ an isometry such that there exists $x_0 \in X$ with a dense orbit under T.
   (i) Show that the orbit of any other $x \in X$ is also dense.
   (ii) Show that X has a natural structure of compact abelian group and T acts as a translation by an element of the abelian group.

9.1.2. Exercise. Suppose that $T : (X, \mu) \to (X, \mu)$ is an ergodic transformation and $f \in L^1(\mu)$ is such that for μ-a.e. $x \in X$

$$\lim_{n \to +\infty} f(x) + f(Tx) + \cdots + f(T^{n-1}x) = +\infty.$$

Prove that $\int f d\mu > 0$.

9.1.3. Exercise (Vitali Covering Lemma). Suppose that $\Omega \subset \mathbb{R}^n$ is a bounded set, and for each $\omega \in \Omega$ there is given a radius r_ω. Show that one can select a countable set of points $\omega_i \in \Omega_i$ such that:
   (i) The balls $B(\omega_i, r_{\omega_i})$ are disjoint.
   (ii) The union of balls $\cup_i B(\omega_i, 5r_{\omega_i})$ covers Ω.

9.1.4. Exercise (Fekete lemma). Suppose that $a_n \in \mathbb{R}$, is a subadditive sequence, that is:

$$a_n + a_m \geq a_{n+m} \quad \forall n, m \in \mathbb{N}$$

Prove that $\lim_n \frac{1}{n} a_n$ exists, and moreover

$$\lim_{n \to \infty} \frac{1}{n} a_n = \inf_n \frac{1}{n} a_n.$$

9.2. Algebraic Groups and Measure Theory

9.2.1. Exercise (Weil Restriction of Scalars). Suppose that $k \subset l$ is a finite field extension and V is an affine algebraic variety over l. Show the existence of $\operatorname{Res}^l_k V$, characterized by:

$$\operatorname{Res}^l_k V(R) = V(R \otimes_k l) \quad \text{for any } k\text{-algebra } R.$$

9.2.2. Exercise (Quaternion Algebras). Show that for any field k of characteristic not 2, the quaternion algebra $\mathbb{H}_{a^2, b}$ is isomorphic to the algebra of 2×2 matrices over k (see Example 3.4.7 for definitions). Show that in characteristic 2, the quaternion algebra is never isomorphic to a matrix algebra.

9.2.3. Exercise (Tensor products). Compute $\mathbb{C} \otimes_\mathbb{R} \mathbb{C}$ as an algebra. Is it a field?

9.2.4. Exercise (Multiplicative and additive groups)**.** Show that there are no algebraic group homomorphisms $\mathbb{G}_m \to \mathbb{G}_a$ or $\mathbb{G}_a \to \mathbb{G}_m$. Give a counterexample at the level of k-points, i.e. give a homomorphism, for some k, between $\mathbb{G}_m(k)$ and $\mathbb{G}_a(k)$ (or vice-versa).

9.2.5. Exercise (Constructible sets)**.** Show that constructible sets (Definition 3.3.1) are the same as the finite unions of locally closed sets.

9.2.6. Exercise. [Normal Subgroup Theorem for Rank 1 groups]

  (i) Let Γ be a finitely generated free group. Show that any finitely-generated normal subgroup must be finite-index.
  (ii) Do the same for Γ the fundamental group of a compact surface.

9.2.7. Exercise (Modular functions)**.** Let G be a locally compact group and μ a left-invariant Haar measure.

  (i) Show that there exists a function $\Delta : G \to \mathbb{R}_{>0}$ such that if R_g denotes right multiplication by $g \in G$ then

$$(R_g)_* \mu = \Delta(g) \cdot \mu$$

  (ii) Suppose that $\Gamma \subset G$ is a discrete subgroup such that G/Γ has a finite left G-invariant measure. Show that $\Delta(G) = 1$.
  (iii) Compute Δ for the group of upper-triangular matrices in $\mathbf{SL}_2 \mathbb{R}$.

9.2.8. Exercise (Failure of continuity for the operator norm)**.** Show that the unitary representation $\mathbb{R} \to \mathcal{U}(L^2(\mathbb{R}))$, where $\mathbb{R}$ acts by translation on functions, is *not continuous* for the operator norm on $\mathcal{U}(L^2(\mathbb{R}))$.

9.2.9. Exercise (Coset Spaces)**.** Suppose that $A, B \subset G$ are two subgroups (closed if G is a topological group). Show that there is a natural identification

$$G \backslash (G/A \times G/B) \cong A \backslash G / B$$

The space $G/A \times G/B$ parametrizes pairs of conjugacy classes of A, B, and taking the quotient by G is considering them up to isomorphism. Often, G/A and G/B have a geometric meaning.

Compute the spaces in question in the following examples:

  (i) Set $A = B = S_{n-1} \subset S_n = G$ where S_k is the permutation group on k elements.
  (ii) Set $A = B = \mathbf{SO}_2 \mathbb{R} \subset \mathbf{SL}_2 \mathbb{R} = G$.
  (iii) Set $A = B = U^+ \subset \mathbf{SL}_2 \mathbb{R} = G$.
  (iv) Set $A = B = A^+ \subset \mathbf{SL}_2 \mathbb{R} = G$, where A^+ is the diagonal subgroup with positive entries.

Hint: Embed $G/B \hookrightarrow G/A \times G/B$ and observe that this is equivariant for the group embedding $A \hookrightarrow G$, then take quotients.

9.2.10. Exercise (Frobenius Lemma on lowest-dimensional representation). Show that aside from the trivial representation, any other irreducible representation of $\mathbf{SL}_2(\mathbb{F}_p)$ has dimension at least $\lfloor \frac{p-1}{2} \rfloor$.
   Hint: Consider a unipotent element in the matrix group and its eigenvalues in a unitary representation. To show that there are many distinct eigenvalues, take powers of the unipotent and consider also the conjugation action by a diagonal group element.

9.2.11. Exercise (Measurable group homomorphisms are continuous). Suppose that G_1, G_2 are locally compact, separable topological groups.

   (i) If $W \subset G_1$ is a set of positive Haar measure, prove that $W \cdot W^{-1}$ contains a neighborhood of the identity. *Hint: Show that for any two functions $f, g \in L^2(G, Haar)$, their convolution $f * g$ is continuous.*
   (ii) Suppose that $\rho : G_1 \to G_2$ is a measurable group homomorphism. Prove that ρ is continuous.
   (iii) Construct a discontinuous group homomorphism $\mathbb{R} \to \mathbb{R}$.

9.3. Entropy

9.3.1. Exercise (Controlling 3-cliques). Suppose that G is a graph with n edges. Show that the number of triangles (i.e. 3-cliques) is bounded above by $\frac{\sqrt{2}}{3} n^{3/2}$.
   Hint: Use Corollary 7.1.5 for an appropriately constructed set.

9.3.2. Exercise (Countably-generated σ-algebras).

   (i) Let $(X, \mathcal{B}, \mu)$ be a Borel space with no atoms and let T be an ergodic measure-preserving transformation of (X, μ). Show that the σ-algebra of T-invariant sets is not countably generated. *Hint: Divide the putative generators into "big" and "small", according to their measure. What is the measure of an orbit?*

9.4. Measure and Topological Rigidity

9.4.1. Exercise. Let $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ be a vector and consider on $\mathbb{T}^n := \mathbb{R}^n / \mathbb{Z}^n$ the translation:

$$T_v(x) = x + v.$$

Show that the orbit closure of any point is a subtorus, and that all orbit closures give isomorphic subtori that differ by a translation.

9.4.2. Exercise. Let $\Gamma \subset \mathbf{SL}_2(\mathbb{Z})$ be a finite index subgroup. Show that for any $x \in \mathbb{T}^2$, the orbit under Γ is either finite or dense. *Hint: Use unipotent elements.*

9.4.3. Exercise. Find a counterexample to the Oppenheim conjecture in $\mathbb{R}^2$ (see Theorem 1.2.1).

9.5. Other places

9.5.1. Exercise.

(i) Suppose that μ is a locally finite measure on $\mathbb{R}$, and there exists a sequence of numbers $\varepsilon_i \in \mathbb{R}$ with $|\varepsilon_i| \to 0$ such that μ is invariant under translation by ε_i. Show that μ is Haar (i.e Lebesgue) measure.

(ii) Fix some $a \in \mathbb{Q}_p^\times$ with $|a| \neq 1$ and suppose that a locally finite measure on $\mathbb{Q}_p$ is invariant by translation by all elements a^k with $k \in \mathbb{Z}$. Show that the measure is Haar.

(iii) Is either of the above statements true, with $\mathbb{R}$ or $\mathbb{Q}_p$ replaced by $\mathbb{F}_q((t))$?

9.5.2. Exercise (Structure of multiplicative groups)**.**

(i) Show that there exists $\varepsilon > 0$ such that if $3^m \equiv 1 \mod 2^n$ and $m \neq 0$ then $|m| \geq \varepsilon \cdot 2^n$.

    Hint: Write it as $(1+2)^m = 1 + k \cdot 2^n$ and "expand 2-adically". It is more convenient to work in $\mathbb{Z}_2$, take logarithms, and consider the closure of the group generated by a single element.

(ii) More generally, let p, q be distinct primes and $v \in \mathbb{Z} \setminus 0$. Show that there exists $\varepsilon > 0$ such that the sequence $\{q^k \cdot v\} \subset \mathbb{Z}/p^N\mathbb{Z}$ has at least εp^N elements.

    Hint: See [Ber16, 9.18] if you get stuck. Set $x = 1 + ap^l$ and note that $x^{N-l} \equiv 1(\mod p^N)$, hence the order of x in $(\mathbb{Z}/p^N\mathbb{Z})^\times$ must divide p^{N-l}. It suffices to check that the order isn't p^{N-l-1}.

9.5.3. Exercise (Ambiguity of convergence)**.** For $\mathbb{Q}$ let $|-|$ denote the usual archimedean norm and $|-|_p$ the p-adic norm.

(i) Find a sequence $x_i \in \mathbb{Q}$ and prime p such that

$$|x_i| \to 0 \quad \text{but} \quad |x_i|_p \to \infty.$$

(ii) Find another sequence $x_i \in \mathbb{Q}$ and prime p such that

$$|x_i| \to 0 \quad \text{but} \quad |x_i - 1|_p \to 0.$$

Appendix A. Transfer operator

References

[BdlHV08] Bekka B., de la Harpe P., & Valette A. — *Kazhdan's property (T)*, vol. 11 of *New Mathematical Monographs*. Cambridge University Press, Cambridge — 2008. https://doi.org/10.1017/CBO9780511542749.

[Ben08] Benoist Y. — "Reseaux des groupes de lie". *Course at Orsay* (2007-2008) . www.math.u-psud.fr/~benoist/.

[Ber16] Bergeron N. — *The spectrum of hyperbolic surfaces.* Universitext. Springer, Cham; EDP Sciences, Les Ulis — 2016. https://doi.org/10.1007/978-3-319-27666-3. Appendix C by Valentin Blomer and Farrell Brumley, Translated from the 2011 French original by Brumley [2857626].

[BFLM11] Bourgain J., Furman A., Lindenstrauss E., & Mozes S. — "Stationary measures and equidistribution for orbits of nonabelian semigroups on the torus". *J. Amer. Math. Soc.* **24** no. 1, (2011) 231–280. https://doi.org/10.1090/S0894-0347-2010-00674-1.

[BQ11] Benoist Y. & Quint J.-F. — "Mesures stationnaires et fermés invariants des espaces homogènes". *Ann. of Math. (2)* **174** no. 2, (2011) 1111–1162. https://doi.org/10.4007/annals.2011.174.2.8.

[CK12] Climenhaga V. & Katok A. — "Measure theory through dynamical eyes". *ArXiv e-prints* (Aug., 2012) — arXiv:1208.4550 [math.DS].

[EL10] Einsiedler M. & Lindenstrauss E. — "Diagonal actions on locally homogeneous spaces". in *Homogeneous flows, moduli spaces and arithmetic* — vol. 10 of *Clay Math. Proc.*, pp. 155–241. Amer. Math. Soc., Providence, RI — 2010.

[ELW] Einsiedler M., Lindenstrauss E., & Ward T. — *Entropy in Ergodic Theory and Homogeneous Dynamics.* in progress. http://www.personal.leeds.ac.uk/~mattbw/entropy/welcome.html.

[EM04] Eskin A. & Margulis G. — "Recurrence properties of random walks on finite volume homogeneous manifolds". in *Random walks and geometry* — pp. 431–444. Walter de Gruyter, Berlin — 2004.

[Esk10] Eskin A. — "Unipotent flows and applications". in *Homogeneous flows, moduli spaces and arithmetic* — vol. 10 of *Clay Math. Proc.*, pp. 71–129. Amer. Math. Soc., Providence, RI — 2010.

[EW11] Einsiedler M. & Ward T. — *Ergodic theory with a view towards number theory*, vol. 259 of *Graduate Texts in Mathematics*. Springer-Verlag London, Ltd., London — 2011. https://doi.org/10.1007/978-0-85729-021-2.

[EW15] ———, "Homogeneous dynamics: a study guide". in *Introduction to modern mathematics*, vol. 33 of *Adv. Lect. Math. (ALM)*, pp. 171–201. Int. Press, Somerville, MA, 2015.

[Fil] Filip S. — "Notes on the multiplicative ergodic theorem". *Ergodic Theory and Dynamical Systems* 1–37. https://doi.org/10.1017/etds.2017.68.

[Fur67]   FURSTENBERG H. — "Disjointness in ergodic theory, minimal sets, and
          a problem in Diophantine approximation". *Math. Systems Theory* **1**
          (1967) 1–49. https://doi.org/10.1007/BF01692494.

[Fur98]   FURSTENBERG H. — "Stiffness of group actions". in *Lie groups and
          ergodic theory (Mumbai, 1996)* — vol. 14 of *Tata Inst. Fund. Res. Stud.
          Math.*, pp. 105–117. Tata Inst. Fund. Res., Bombay — 1998.

[Ghy92]   GHYS E. — "Dynamique des flots unipotents sur les espaces homogènes".
          Astérisque no. 206, (1992) Exp. No. 747, 3, 93–136. Séminaire Bourbaki,
          Vol. 1991/92.

[Led84]   LEDRAPPIER F. — "Quelques propriétés des exposants caractéristiques".
          in *École d'été de probabilités de Saint-Flour, XII—1982* — vol. 1097 of
          Lecture Notes in Math., pp. 305–396. Springer, Berlin — 1984. https:
          //doi.org/10.1007/BFb0099434.

[Lyo88]   LYONS R. — "On measures simultaneously 2- and 3-invariant". *Is-
          rael J. Math.* **61** no. 2, (1988) 219–224. https://doi.org/10.1007/
          BF02766212.

[Mar91]   MARGULIS G. A. — *Discrete subgroups of semisimple Lie groups*, vol. 17
          of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in
          Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin — 1991.
          https://doi.org/10.1007/978-3-642-51445-6.

[McM17]   MCMULLEN C. — "Hyperbolic manifolds, discrete groups and ergodic
          theory". *Course Notes* (2017) . math.harvard.edu/~ctm.

[Mil17]   MILNE J. S. — *Algebraic Groups: The Theory of Group Schemes of
          Finite Type over a Field*. Cambridge Studies in Advanced Mathematics.
          Cambridge University Press — 2017.

[MT94]   MARGULIS G. A. & TOMANOV G. M. — "Invariant measures for
          actions of unipotent groups over local fields on homogeneous spaces".
          Invent. Math. **116** no. 1-3, (1994) 347–392. https://doi.org/10.1007/
          BF01231565.

[Rat91]   RATNER M. — "On Raghunathan's measure conjecture". *Ann. of Math.
          (2)* **134** no. 3, (1991) 545–607. https://doi.org/10.2307/2944357.

[Rud90]   RUDOLPH D. J. — "×2 and ×3 invariant measures and entropy".
          Ergodic Theory Dynam. Systems **10** no. 2, (1990) 395–406. https:
          //doi.org/10.1017/S0143385700005629.

[Sha48]   SHANNON C. E. — "A mathematical theory of communication". *Bell
          System Tech. J.* **27** (1948) 379–423, 623–656. https://doi.org/10.
          1002/j.1538-7305.1948.tb01338.x.

[WM15]   WITTE MORRIS D. — *Introduction to Arithmetic Groups*. Deductive
          Press — 2015. http://deductivepress.ca.

[Zim84]   ZIMMER R. J. — *Ergodic theory and semisimple groups*, vol. 81 of
          Monographs in Mathematics. Birkhäuser Verlag, Basel — 1984. https:
          //doi.org/10.1007/978-1-4684-9488-4.

DEPARTMENT OF MATHEMATICS
HARVARD UNIVERSITY
1 OXFORD ST, CAMBRIDGE, MA 02139
Email address: sfilip@math.harvard.edu