SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

Pointwise Ergodic Theory on Groups and Ratner's Theorems

av

Thomas Ohlson Timoudas

2012 - No 27

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

Pointwise Ergodic Theory on Groups and Ratner's Theorems

Thomas Ohlson Timoudas

Självständigt arbete i matematik 15 högskolepoäng, Grundnivå

Handledare: Henrik Shahgholian

2012

Contents

Acknowledgement

First of all, I wish to thank my adviser Henrik Shahgholian for having introduced me to this strange and remote area of mathematics. I would also like to extend my thanks to my examiner Rikard Bøgvad, for his valuable comments and corrections on an earlier draft of this paper.

I express my thanks to the members of my family, who have put up with me and kept me company through the years, especially so to my mother Lena, who pushed me through upper secondary school (senior high school) when I was about to quit prematurely. Had you not done so, I might have ended up just a drop-out IT-billionaire. I would most likely have been a lesser manifestation of myself, if at all, had none of you been there for me.

Lastly, I want to thank all my friends (you know who you are) for their support and the good times we have had together. One deserving special mention is Bashar Saleh, who has given me valuable comments on early drafts of this paper, not to mention having been a good friend throughout these past three years.

1 Introduction

Ergodic theory is the study of long-term behaviour of transformations. It can help us answer questions such as:

Is the set $\{\sin n \mid n \in \mathbb{N}\}$ dense in the interval $[-1,1]$?

The answer is yes, since $\sin(x)$ is continuous and $n \mod 2\pi$ is dense in $[0,2\pi)$. This question would not be possible to answer unless we had some knowledge of the distribution of $n \mod 2\pi$.

Let us try adding a twist to our example by asking:

Is the set $\{n \sin n \mid n \in \mathbb{N}\}$ dense on the real line $\mathbb{R}$?

I don't have a ready answer to this question, and it may very possibly be an open question. The problem is related to determining good bounds for how well π can be approximated by rational numbers $\frac{p}{q}$ (that is: how large does q have to be in order to make $|\frac{p}{q} - \pi| < \epsilon$?).

A very interesting example in diophantine approximation is the Littlewood conjecture, stating that

$$\liminf_{n\to\infty} n\|na\|\|nb\| = 0,$$

for every $a,b \in \mathbb{R}$, where $\|na\|$ is equal to the distance to the integer closest to na. This is still an open question, but it has been proved (by Einseidler, Katok, and Lindenstrauss, see [Taoa]) that the set of exceptions has Hausdorff dimension 0 (that is, the exceptions are basically isolated points).

An answered question of similar flavour is the famous Oppenheim conjecture, solved in 1989 by Margulis (see [Taob] for a more detailed discussion). It was conjectured in 1929 by Oppenheim, that:

Let Q be a real quadratic form[1] in $n \geq 3$ variables (he initially considered only ones in five or more variables), that is indefinite[2], non-degenerate[3], and not a scalar multiple of a quadratic form with rational coefficients. Then $Q(Z^n)$ is dense on the real line R. That is, to get any real value, it suffices to evaluate Q at integer points!

To prove this, Margulis settled a special case of Raghunathan's conjecture. Raghunathan's conjecture is a deep observation made in the 1960's by Raghunathan, which has been generalized in subsequent steps by many different mathematicians. In one of its many manifestations, Raghunathan's conjecture reads:

Let G be a connected Lie group and Γ a lattice in G ($\Gamma\backslash G$ has finite volume). If ϕ_t is a unipotent flow on $\Gamma\backslash G$, then the closure of every ϕ_t-orbit is homogeneous ($\overline{\phi_{\mathbb{R}}(x)} = xS$, for some closed subgroup $S \subseteq G$).

The conjecture was resolved in the early 1990's by Marina Ratner in a series of articles [Rat90b], [Rat90a] and [Rat91] (all together totalling more than 150 pages). It has since, by among others Ratner, Shah, and Margulis, been proved for more general Lie groups, and there are different versions of the conjecture. The one receiving focus in this thesis is Ratner's measure classification theorem:

THEOREM **1.1** ***Ratner's Measure Classification Theorem***
Let G be a Lie group, Γ a discrete subgroup in G and U a unipotent subgroup of G. Then every ergodic U-invariant probability measure on $\Gamma\backslash G$ is homogeneous.

The above theorem will be restated in section 4, and the proof for the case of $G = SL(2,\mathbb{R})$ will be given, following her article [Rat92]. The proof follows that in the article, but has been expounded on, in an attempt to make her ideas more transparent and easy to follow. The reader interested in the more technical details may consult her article.

[1] A quadratic form is a polynomial in which (possibly) mixed monomials occur, and the total degrees of each monomial is exactly 2

[2] Q attains both positive and negative values

[3] Q can not be written as a quadratic form in fewer variables

2 Motivation

Consider the circle group $S^1 = \{z \in \mathbb{C} : |z| = 1\}$ as a differentiable manifold (S^1 with complex multiplication and the usual analytical structure), making it into a Lie group.

There is an obvious smooth epimorphism[4] from $\mathbb{R}$ onto S^1 given by $\mathbb{R} \to S^1 : r \mapsto e^{2\pi i r}$. It is obvious that the map is surjective and that its kernel is $\mathbb{Z}$. This means that we get an induced smooth isomorphism

$$\mathbb{R}/\mathbb{Z} \to S^1 : r \mod 1 \mapsto e^{2\pi i r}.$$

In formal language, we say that $\mathbb{R}$ is a universal covering group of S^1. This allows us to identify $\mathbb{R}/\mathbb{Z}$ with S^1 under the isomorphism given above.

This construction generalizes to identifying the n-torus $\mathbb{T}^n$ (which is given as a direct product $(S^1)^n$ of n circles) with the quotient $\mathbb{R}^n/\mathbb{Z}^n \cong (\mathbb{R}/\mathbb{Z})^n$, by the smooth isomorphism

$$\mathbb{R}^n/\mathbb{Z}^n \to \mathbb{T}^n : (r_1 \mod 1, \cdots, r_n \mod 1) \mapsto (e^{2\pi i r_1}, \cdots, e^{2\pi i r_n}).$$

The point of this discussion is that we can forget the group structure of $\mathbb{T}^n$ and consider it only as a differentiable manifold. Of course we can still identify it with the differentiable manifold $(\mathbb{R}/\mathbb{Z})^n$. We may then consider the automorphism group $\mathrm{Diff}(\mathbb{T}^n)$ of diffeomorphisms of the n-torus. We shall write $\mathbb{T}^n$ for the n-torus considered only as a differentiable manifold, and $(\mathbb{R}/\mathbb{Z})^n$ for the n-torus with the group structure defined above.

Now, there is an interesting homomorphism from the group $(\mathbb{R}/\mathbb{Z})^n$ to the group $\mathrm{Diff}(\mathbb{T}^n)$ given by the map $\mathrm{Rot} : r \mod 1 \mapsto T_r$, where

$$T_r : \mathbb{T}^n \to \mathbb{T}^n : (z_1, \cdots, z_n) \mapsto (z_1 e^{2\pi i r_1}, \cdots, z_n e^{2\pi i r_n}).$$

The name was chosen to highlight the fact that T_r are just rotations of the points on the n-torus, by the angle $(r \mod 1)2\pi$. The map Rot maps elements from $(\mathbb{R}/\mathbb{Z})^n$ to $\mathrm{Diff}(\mathbb{T}^n)$ in a smooth fashion, since angles close to each other are mapped in a smooth way to rotations close to each other.

Perhaps more spectacular is the homomorphism given below, from $(\mathbb{R}/\mathbb{Z})^n$ into the group of smooth homomorphisms from $\mathbb{R}$ to $\mathrm{Diff}(\mathbb{T}^n)$ (try not to trip up on the notation, the meaning of this will become clear in the following discussion)[5]. Define the homomorphism

$$\mathrm{RectLin} : (\mathbb{R}/\mathbb{Z})^n \to \mathrm{Hom}(\mathbb{R}, \mathrm{Diff}(\mathbb{T}^n)) : r \mod 1 \mapsto \mathrm{RectLin}(r)$$

where $\mathrm{RectLin}(r)$ corresponds to the smooth flow on $\mathbb{T}^n$ defined by

$$\phi[r]_t(z_1, \cdots, z_n) = (z_1 e^{2\pi i t r_1}, \cdots, z_n e^{2\pi i t r_n}).$$

The r in brackets is only present to signify the dependence of the flow on the chosen r (which is the speed of rotation). These flows are called rectilinear (hence the suggestive name RectLin).

EXAMPLE **2.1** ***Rotation of the circle***
For the circle group $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, there is an obvious measure on this group induced by the Lebesgue measure on $\mathbb{R}$ ($\mu([a,b]) = |b-a|$). Let α be some real number and define the (measurable) transformation $T_\alpha : \mathbb{T} \to \mathbb{T}$ by

$$T_\alpha(z) = z e^{2\pi i \alpha}.$$

It is clear that $T_\alpha(z) = z T_\alpha(1)$ and so we only need to focus on the orbit of the point 1, that is the set $\{T^n(1) : n = 1, 2, 3, ...\}$. T_α is actually the element $\mathrm{Rot}(\alpha \mod 1)$ defined in the preamble of this section.

The long-term behaviour of the transformation T_α depends on whether α is rational or irrational.

[4] Surjective homomorphism.

[5] The group of smooth homomorphisms from $\mathbb{R}$ to $\mathrm{Diff}(\mathbb{T}^n)$ is just the group of smooth flows on $\mathbb{T}^n$ under pointwise addition.

When α is rational, the orbit of the transformation is finite; that is, we visit a finite number of points on the circle, before returning to the point from which we started.

More interesting behaviour appears when α is irrational, since then the orbit of the transformation is dense in the circle, or in other words: almost every point on the circle is visited at least once. The qualifier *almost* is used in the sense that the points that are not visited are negligible when integrating.

Of course, the notion of visiting a point becomes blurred in the limit, but the orbit being dense means that for any given point, the orbit visits points that come closer and closer to the given point as time increases.

Rational α

When α is rational, we can express it as $\alpha = p/q$ for p, q relatively prime integers. Now it is obvious that $T_\alpha^q(1) = (e^{2\pi i p/q})^q = e^{2\pi i q p/q} = e^{2\pi i p} = e^0 = 1$ so that $T_\alpha^q = Id_{\mathbb{T}}$ and hence T_α is periodic (with smallest period q). The orbit contains exactly q points, since p and q are relatively prime ($kp/q = n \in \mathbb{Z} \Rightarrow kp = qn \Rightarrow k = mq$).

We can construct a set L that is invariant under rotations by this α (that is, rotating the set by α gives us back the same set - $T_\alpha(L) = L$), by dividing the circle into small, evenly spaced out, segments and taking their union. Notice that rotating the segment $[e^{2\pi i \frac{k}{q}}, e^{2\pi i (\frac{k}{q} + \frac{1}{2q})}]$ by $\alpha = \frac{p}{q}$ yields the segment

$$[e^{2\pi i (\frac{k}{q} + \frac{p}{q})}, e^{2\pi i (\frac{k}{q} + \frac{1}{2q} + \frac{p}{q})}] = [e^{2\pi i \frac{k+p}{q}}, e^{2\pi i (\frac{k+p}{q} + \frac{1}{2q})}]$$

By taking the union $L = \bigcup_{k=0}^{q-1} [e^{2\pi i \frac{k}{q}}, e^{2\pi i (\frac{k}{q} + \frac{1}{2q})}]$ (we need q of these segments since the smallest period of T_α is q), we see that the set is invariant under T_α. The set L has measure $\mu(L) = q\frac{1}{2q} = \frac{1}{2}$ (the intervals are disjoint) but is invariant under T_α and so μ is not ergodic with respect to T_α (see 3.13 for a definition of ergodicity).

Irrational α

Whenever α is irrational, $T_\alpha^n \neq T_\alpha^m$ for $n \neq m$, and T_α will be an ergodic action on the circle with respect to the Lebesgue measure.

To prove that the orbit of T_α is dense on the circle, we need to show that every open set on the circle is visited by the orbit, or that we can find points on the circle that lie arbitrarily close to any given point. Let $\epsilon > 0$ be given and consider the (non-periodic) orbit

$$\{T_\alpha^n(1) = e^{2\pi i n\alpha} : n = 0, 1, 2, ...\}.$$

Let $N = N(\epsilon) = \lceil 1/\epsilon \rceil$[6] and divide the circle into N connected parts of equal length. Since $\{T_\alpha^n(1) = e^{2\pi i n\alpha} : n = 0, 1, 2, ...\}$ has infinitely many points, we may use the pigeonhole principle[7] to conclude that there are positive integers n, m, both less than $N + 1$, such that

$$0 < \arg(T_\alpha^n(1) T_\alpha^{-m}(1)) = \arg(T_\alpha^{n-m}(1)) < \epsilon,$$

where $\arg(z)$ stands for the argument of the complex number z. The above inequality expresses the fact that we can find two points on the circle, $T_\alpha^n(1)$ and $T_\alpha^m(1)$, that lie arbitrarily close to each other.

Now we see that the there is a point in the orbit, $T_\alpha^{n-m}(1)$, that can be chosen as close to 0 as we want, so by rotating this point we can get as close to any point on the circle we want. This is expressed by the fact that $T_\alpha^{k(n-m)}(1) = e^{2\pi i k\delta}$, for some $0 < \delta < \epsilon$, and so any point $z \in \mathbb{T}$ is at most at a distance ϵ from a point in the orbit $\{T_\alpha^n(1) : n = 0, 1, 2, ...\} \supset \{T_\alpha^{k(n-m)}(1) : k = 0, 1, 2, ...\}$. Since ϵ can be chosen arbitrarily small, this concludes the proof that the orbit is dense.

[6] The smallest integer greater than or equal to $1/\epsilon$.

[7] The pigeonhole principle says that, if we divide a set into N parts, and choose $N + 1$ elements from the set, then at least 2 of the chosen elements will belong to the same part.

The next example concerns the distribution of a line on the torus.

Example **2.2** ***Rectilinear flow on the 2-dimensional torus***
The (2-dimensional) torus can be expressed as the quotient space $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$. Given a vector $v = (\alpha_1, \alpha_2) \in \mathbb{R}^2$, we will consider the flow starting from $z = (z_1, z_2)$

$$\phi_t((z_1, z_2)) = (z_1 e^{2\pi it\alpha_1}, z_2 e^{2\pi it\alpha_2}) = (z_1, z_2)\phi_t((1,1)).$$

Since $\phi_t(z) = z\phi_t(1)$ we may, without loss of generality, restrict our discussion to the flow starting from the point $1 = (1,1)$.

There are three possible cases to consider, and they are as follows:

1. **(The trivial case)** $\alpha_1 = \alpha_2 = 0$**.**

   The closure of the orbit of the point z is $z\overline{\{\phi_t(1) : t \geq 0\}} = z\overline{\{1\}} = \{z\}$ (a single point or $\mathbb{T}^0$).

2. **At least one of** α_1 **and** α_2 **is not equal to zero, and there exist integers** m **and** n**, not both equal to zero, such that** $n\alpha_1 + m\alpha_2 = 0$**.**

   Without loss of generality, we may assume that $\alpha_1 \neq 0$.

   Let $n\alpha_1 + m\alpha_2 = 0$ for some integers m and n, not both equal to zero, then we have that $n = -m\frac{\alpha_2}{\alpha_1}$. There is a unique choice of a such integers m and n having the property that $\gcd(m,n) = 1$, ensuring that m is non-zero. Having chosen m as such, it is clear that the (smallest) period of the orbit will be $\frac{m}{\alpha_1}$:

   $$\phi_{\frac{m}{\alpha_1}}(1) = (e^{2\pi i\frac{m}{\alpha_1}\alpha_1}, e^{2\pi i\frac{m}{\alpha_1}\alpha_2}) = (e^{2\pi im}, e^{-2\pi in}) = (1,1) = \phi_0(1).$$

   This gives us a way to construct a diffeomorphism from $\mathbb{T}^1$ to $\overline{\phi_{[0,\frac{m}{\alpha_1})}} = \phi_{[0,\frac{m}{\alpha_1})}$ (and since the orbit is periodic, this means that the whole orbit is the 1-torus). Define

   $$\pi : \phi_{[0,\frac{m}{\alpha_1})} \to \mathbb{T}^1 : (e^{2\pi it\alpha_1}, e^{2\pi it\alpha_2}) \mapsto e^{2\pi i\frac{\alpha_1}{m}t},$$

   which is a diffeomorphism. We have thus proved that the closure of the orbit of the point z, $\phi_{\mathbb{R}}(z)$, is the 1-torus.

3. $n\alpha_1 + m\alpha_2 = 0$ **implies that** $n = m = 0$ **(implying that both** α_1 **and** α_2 **are non-zero, and that they are linearly independent over** $\mathbb{Q}$**).** We see that at least one of α_1 and α_2 is irrational (otherwise choose $m = p_2q_1, n = p_1q_2$, where $\alpha_i = \frac{p_i}{q_i}$). Looking back at the example of rotations of the circle, we see that the closure of the orbit $\phi_{\mathbb{R}}(1)$ contains every point $(1,a)$, where $a \in T^1$. Since the orbit is just a straight line, every point will lie in the closure of $\phi_{\mathbb{R}}(1)$.

More generally, for any $(\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$, the orbits of the flow

$$\phi_t((z_1, ..., z_n)) = (z_1 e^{2\pi it\alpha_1}, ..., z_n e^{2\pi it\alpha_n})$$

are subtori $\mathbb{T}^d$ of $\mathbb{T}^n$ for some $d \in \{0, 1, ..., n\}$, where d actually is the rank of $\alpha_1, ..., \alpha_n$ over $\mathbb{Q}$ (the dimension of their span over $\mathbb{Q}$).

One of the points of this thesis is exploring the generalization of the above result to unipotent flows on homogeneous spaces (these notions will be explained in section 4 and in the appendix on Lie groups).

3 Mathematical preliminaries

All measure spaces in this and the following sections are assumed to be Borel probability spaces, unless otherwise stated.

3.1 Ergodic theory

Two very good (and perhaps the only) introductions to ergodic theory on homogeneous spaces are the books [BM00] and [EW11]. Everything covered in this section can be found in any of these books.

The aim of this section is not to give a thorough treatment of ergodic theory, but to provide the ergodic theory needed in the proof of Ratner's theorem. Anyone interested in the proof of the general case should also read the account given by Ratner herself in [Rat90b].

Definition **3.1** ***Measure preserving transformation***
A map $T:(X,\mathscr{A},\mu)\to(Y,\mathscr{B},\nu)$ is called a measure-preserving transformation (m.p.t. for short) if it is measurable and $\mu(T^{-1}B)=\nu(B)$ for every $B\in\mathscr{B}$. If T^{-1} is an a.e. defined measurable map, then T is called an invertible measure preserving transformation. It follows that T^{-1} is also measure-preserving, since $\nu(TA)=\mu(T^{-1}TA)=\mu(A)$ for every $A\in\mathscr{A}$.

Remark. For a measure preserving transformations $T:X\to X$, the definition becomes that $\mu(T^{-1}A)=\mu(A)$, and we will instead say that μ is T-invariant (invariant with respect to T). This basically means that volume is invariant under the transformation (moving a set does not change its volume).

Proposition **3.2**
A probability measure μ on X is T-invariant if and only if for every $f\in L^1$

$$\int_X f\,d\mu=\int_X f\circ T\,d\mu. \tag{1}$$

Proof. Suppose that (1) holds for every $f\in L^1$. Since μ is finite, the characteristic functions are in L^1. We then have for any measurable A, that

$$\mu(A)=\int\chi_A\,d\mu=\int\chi_{T^{-1}A}\,d\mu=\int\chi_A\circ T\,d\mu=\mu(T^{-1}A).$$

To show the converse, suppose that μ is T-invariant, then for any characteristic function χ_A:

$$\int\chi_A\,d\mu=\mu(A)=\mu(T^{-1}A)=\int\chi_{T^{-1}A}\,d\mu=\int\chi_A\circ T\,d\mu.$$

By linearity (1) holds for all simple functions. We just need to show that it holds for every $f\in L^1$ by approximating them with simple functions. To any positive $f\in L^1$ there is associated a sequence of simple functions $\{f_n\}$ as the one in A.10. This may be chosen monotone increasing by A.11.

Let such a monotone increasing sequence be given, then $\{f_n\circ T\}$ is also a monotone increasing sequence of simple functions that is a.e. pointwise convergent to $f\circ T$. Lebesgue's monotone convergence gives us that

$$\int f\circ T\,d\mu=\lim_{n\to\infty}\int f_n\circ T\,d\mu=\lim_{n\to\infty}\int f_n\,d\mu=\int f\,d\mu.$$

Since any measurable function f can be decomposed into the sum f^+-f^-, where both f^+ and f^- are positive measurable functions, the claim follows. □

Definition **3.3** ***Associated operator***
Let T be a measure-preserving transformation on $(X,\mathscr{A},\mu)$. Then we have an operator $U_T:L^2_\mu\to L^2_\mu$ defined by

$$U_Tf=f\circ T.$$

By 3.2, $\text{image}(U_T) \subseteq L^2_\mu$. The operator is an isometry since

$$\langle U_T f, U_T g\rangle = \int f(Tx)\overline{g(Tx)}\,d\mu = \int f(x)\overline{g(x)}\,d\mu = \langle f, g\rangle\,.$$

If T is an invertible measure-preserving transformation U_T is a unitary operator and called the associated unitary operator of T. It is unitary, since if $f \in L^2(\mu)$, then $f \circ T^{-1} \in L^2(\mu)$, and $U_T(f \circ T^{-1}) = f$, so U_T is surjective.

Theorem **3.4** ***Mean ergodic theorem***
Let T be measure-preserving and $I = \{f \in L^2_\mu : U_T f = f\}$. Then I is closed and as $N \to \infty$

$$\frac{1}{N}\sum_{n=0}^{N-1} U_T^n f \xrightarrow[L^2_\mu]{} P_T f,$$

where P_T is the orthogonal projection onto I.

Proof. Let $B = \{U_T g - g : g \in L^2_\mu\}$, then the orthogonal complement to B is I; that is $B^\perp = I$. It can be seen that I is closed since if $f_n(x) \in I$ converges pointwise everywhere to $f(x)$, then $U_T f(x)$ also converges pointwise everywhere to $U_T f$. Since $f_n = U_T f_n$, it follows that $U_T f = f$. Of course if $f \in I$, then

$$\langle f, U_T g - g\rangle = \langle U_T f, U_T g\rangle - \langle f, g\rangle = 0$$

for every $g \in L^2_\mu$, since U_T is an isometry. Suppose instead that

$$\langle f, U_T g - g\rangle = 0$$

for every $g \in L^2_\mu$, that is

$$\langle U_T g, f\rangle = \langle g, f\rangle$$

for every $g \in L^2_\mu$. Since the adjoint operator U_T^* of U_T is the unique operator such that for for every $g \in L^2_\mu$, and any $f \in L^2(\mu)$

$$\langle U_T g, f\rangle = \langle g, U_T^* f\rangle\,,$$

it follows that $U_T^* f = f$, since

$$0 = \langle g, f\rangle - \langle g, U_T^* f\rangle + \langle f, g\rangle - \langle U_T^* f, g\rangle = \langle f - U_T^* f, f - U_T^* f\rangle$$

implies $f = U_T^* f$. That f is indeed in I can be shown by noting that

$$\begin{aligned}
\|U_T f - f\|_2 &= \langle U_T f - f, U_T f - f\rangle = \\
&= \|U_T f\|_2^2 - \langle U_T f, f\rangle - \langle f, U_T f\rangle + \|f\|_2^2 = \\
&= 2\|f\|_2^2 - \langle f, U_T^* f\rangle - \langle U_T^* f, f\rangle = \\
&= 2\|f\|_2^2 - \langle f, f\rangle - \langle f, f\rangle = \\
&= 0.
\end{aligned}$$

This means that we get the decomposition $L^2_\mu = I \oplus \overline{B}$, where $\overline{B}$ denotes the closure of B, so that every $L^2_\mu \ni f = P_T f + h$ for a unique $h \in \overline{B}$. It is clear that if $U_T g - g = h \in B$ and $f = P_T f + h$, then

$$\begin{aligned}
\|\frac{1}{N}\sum_{j=0}^{N-1} U_T^j f - P_T f\|_2 &= \|\frac{1}{N}\left(\sum_{j=0}^{N-1} P_T f + \sum_{j=0}^{N-1} U_T^j h\right) - P_T f\|_2 = \\
&= \|\frac{1}{N}\sum_{j=0}^{N-1} U_T^j (U_T g - g)\|_2 = \frac{1}{N}\|U_T^N g - g\|_2 \le \\
&\le \frac{1}{N}\left(\|U_T^N g\|_2 + \|g\|_2\right) = \frac{2}{N}\|g\|_2 \xrightarrow[N\to\infty]{} 0.
\end{aligned}$$

To show that the convergence holds even if $h \in \overline{B}$, consider some sequence $h_i = U_T g_i - g_i$ in B converging to h, then

$$\|\frac{1}{N}\sum_{j=0}^{N-1} U_T^j h\|_2 \leq \frac{1}{N}\left(\|\sum_{j=0}^{N-1} U_T^j (h-h_i)\|_2 + \|\sum_{j=0}^{N-1} U_T^j h_i\|_2\right) \leq$$
$$\leq \|h-h_i\|_2 + \frac{2}{N}\|g_i\|_2 \leq \epsilon + \frac{2}{N}\|g_i\|_2 \xrightarrow[N\to\infty]{} \epsilon$$

where the last inequality holds for any $\epsilon > 0$ and sufficiently large i (depending on ϵ) since h_i converge to h in L^2_μ. Since ϵ is arbitrary, we get the desired conclusion. $\square$

Remark. The above theorem gives us a (non-unique) decomposition of any $f \in L^1$ into components

$$f = P_T f + U_T g - g + h,$$

where $U_T g - g$ is an element in $\{U_T g - g\}$ and h is the difference between $U_T g - g$ and the component of f in the closure of $\{U_T g - g\}$. Given any $\delta > 0$, we may choose this decomposition in such a way as to ensure $\|h\|_2 < \delta$.

Definition **3.5** ***Positive linear operator***
A linear operator $U : L^1 \to L^1$ is said to be positive if whenever $f \geq 0$, then $Uf \geq 0$.

Proposition **3.6** ***Maximal inequality***
Let $U : L^1 \to L^1$ be a positive linear operator with $\|U\| \leq 1$. For a real-valued function $f \in L^1$, we define inductively the functions

$$\begin{aligned} f_0 &= 0 \\ f_1 &= f \\ f_2 &= f + Uf \\ &\vdots \\ f_n &= f + Uf + \cdots + U^{n-1} f \end{aligned}$$

for $n \geq 1$, and $F_N = \max\{f_n \mid 0 \leq n \leq N\}$ (all of the functions are defined pointwise). Then

$$\int_{\{x \mid F_N(x) > 0\}} f \, d\mu \geq 0$$

for all $N \geq 1$.

Proof. See Proposition 2.26 in [EW11]. $\square$

Theorem **3.7** ***Maximal ergodic theorem***
Consider a measure-preserving transformation T and a real-valued function $g \in L^1$. Define

$$E_\alpha = \left\{ x \in X \,\middle|\, \sup_{n\geq 1} \frac{1}{n}\sum_{i=0}^{n-1} g(T^i x) > \alpha \right\}$$

for any $\alpha \in \mathbb{R}$. Then

$$\alpha\mu(E_\alpha) \leq \int_{E_\alpha} g \, d\mu \leq \|g\|_1.$$

Proof. We begin by noting that U_T is positive and $\|U_T\| = 1$ (U_T is isometric). This allows us to use 3.6.

For any $\alpha \in \mathbb{R}$ we define the function $f = g - \alpha$. We then see that

$$E_\alpha = \left\{ x \in X \,\middle|\, \sup_{n\geq 1} \frac{1}{n}\sum_{i=0}^{n-1} U_T^i g(x) > \alpha \right\} = \left\{ x \in X \,\middle|\, \sup_{n\geq 1} \frac{1}{n}\sum_{i=0}^{n-1} U_T^i f(x) > 0 \right\}$$

For any $n \geq 1$, define the sets

$$E_\alpha^n = \left\{ x \in X \,\middle|\, \sup_{1 \leq k \leq n} \frac{1}{k} \sum_{i=0}^{k-1} U_T^i f(x) > 0 \right\}$$

It is clear that $E_\alpha^n \subseteq E_\alpha^{n+1}$ and so $E_\alpha = \bigcup_{n=1}^{\infty} E_\alpha^n$. By the maximal inequality 3.6, we get the inequality

$$0 \leq \int_{E_\alpha} f \, d\mu = \int_{E_\alpha} g \, d\mu - \alpha\mu(E_\alpha)$$

which simplifies to $\alpha\mu(E_\alpha) \leq \int_{E_\alpha} g \, d\mu$. □

We will write S_N for the operator

$$S_N = \frac{1}{N} \sum_{n=0}^{N-1} U_T^n.$$

THEOREM **3.8** ***Birkhoff's Pointwise Ergodic Theorem***
Let μ be probability measure (on X), T a measure-preserving transformation, and $f \in L^1(X, \mu)$. Then there is a T-invariant function $\overline{f}$ in L^1 such that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f = \lim_{N \to \infty} S_N f = \overline{f}$$

for a.e. $x \in X$, where $\int_X \overline{f} \, d\mu = \int_X f \, d\mu$. If T is in addition ergodic then $\overline{f} = \int_X f \, d\mu$.

Proof. Since μ is positive by assumption (we are working exclusively with positive measures) we only need to show that for any $\epsilon > 0$

$$\mu(\{x \mid \limsup_{N \to \infty} |S_N f - P_T f| > \epsilon\}) = 0.$$

We note that since L^2 is dense in L^1 (Prop. A.15), for any $\delta > 0$ we can find a function $f_0 \in L^2$ such that $\|f - f_0\|_1 < \delta$. Since L^2 is even contained in L^1, $I = \{f \in L_\mu^2 : U_T f = f\}$ is a closed subspace of L^1. This means that P_T is also well-defined on L^1.

Let $\delta > 0$ be arbitrary, and remember the decomposition we gave in the remark following Theorem 3.4. This allows us to write $f_0 = P_T f_0 + (U_T g_0 - g_0) + h_0$ for some $\|h_0\|_2 < \delta$, and so by letting

$$h_1 = h_0 + (f - f_0) - (P_T f - P_T f_0) \in L^1$$

we get that

$$f = P_T f + (U_T g_0 - g_0) + h_1$$

where $\|h_1\|_1 \leq \|h_0\|_1 + \|f - f_0\|_1 + \|P_T f - P_T f_0\|_1 < 3\delta$.

Since L^∞ is dense in L^2 (Prop. A.15) we may find a $g \in L^\infty$ such that $\|g - g_0\|_2 < \delta$. Letting

$$h = h_1 - ((U_T g - g) - (U_T g_0 - g_0)) = h_1 + U_T(g_0 - g) + (g - g_0)$$

we get the decomposition $f = P_T f + (U_T g - g) + h$. Recalling that U_T is an isometry, we get the bound

$$\|h\|_1 \leq \|h_1\|_1 + 2\|g - g_0\|_1 \leq \|h_1\|_1 + 2\|g - g_0\|_2 < 5\delta,$$

Since $\delta > 0$ is arbitrary we may instead assume that $\|h\|_1 < \delta$. Since P_T is a projection operator ($P_T^k = P$, for $k \geq 1$), it follows that $S_N(P_T f) = P_T f$, and we get the inequality

$$\begin{aligned} |S_N f - P_T f| &= |S_N(P_T f) + S_N(U_T g - g) + S_N(h) - P_T f| = \\ &= |P_T f + S_N(U_T g - g) + S_N(h) - P_T f| \leq \\ &\leq |S_N(U_T g - g)| + |S_N(h)| \end{aligned}$$

As in the proof of the Mean Ergodic Theorem, $S_N(U_T g - g)$ telescopes, giving us $|S_N(U_T g - g)| = \frac{1}{N}|U_T^N g - g| \le \frac{2}{N}\|g\|_\infty$ except on a set of measure 0, since the maximum variance between $|U_T^N g(x) - g(x)|$ is at most equal to $2\|g\|_\infty$ for a.e. x. That is,

$$\begin{aligned}\mu(\{x \mid \limsup_{N\to\infty}|S_N f - P_T f| > \epsilon\}) &\le \mu(\{x \mid \limsup_{N\to\infty}|S_N(U_T g - g)| + |S_N(h)| > \epsilon\}) = \\ &= \mu(\{x \mid \limsup_{N\to\infty}|S_N(h)| > \epsilon\}) \le \\ &\le \mu(\{x \mid \sup_{N\ge 1}|S_N(h)| > \epsilon\}) \le \frac{\|h\|_1}{\epsilon} < \frac{\delta}{\epsilon}\end{aligned}$$

where the last inequality follows from the Maximal Ergodic Theorem, since

$$\epsilon\mu(\{x \mid \sup_{N\ge 1} S_N(|h|) > \epsilon\}) \le \|h\|_1.$$

Since ϵ is fixed, and δ can be chosen arbitrarily small, this gives us convergence a.e. of $S_N f$ to $P_T f$. Now, since

$$\int_X S_N f\, d\mu = \int_X f\, d\mu,$$

we see that, by Lebesgue monotone convergence,

$$\int_X f\, d\mu = \int_X \lim_{N\to\infty} S_N f\, d\mu = \int_X P_T f\, d\mu.$$

If T is ergodic, then any T-invariant function is constant a.e. (see Proposition 2.14 in [EW11]), and so

$$\int_X P_T f\, d\mu = P_T f \cdot \mu(X) = P_T f.$$

□

Definition **3.9** ***Smooth flow***
A flow on a smooth manifold X is an action of $\mathbb{R}$ on X. We say that $\phi_t(x) = \Phi(t,x) : \mathbb{R}\times X \longrightarrow X$ is a flow on X if:

- $\phi_0(x) = x$, and
- $\phi_s \circ \phi_t(x) = \phi_{s+t}(x)$,

for every $s,t \in \mathbb{R}$ and every $x \in X$.

If in addition the map $\Phi(t,x)$ is smooth w.r.t. both variables, we say that ϕ_t is a smooth flow. Since we will only consider smooth flows in this text, we will drop the qualifier "smooth" and simply say flow, unless otherwise stated.

Definition **3.10** ***Invariant measure***
A measure μ on a space X is said to be invariant under the action of the group G if $\mu(g^{-1}A) = \mu(A)$ for every measurable subset A of X and $g \in H$. We sometimes express this by saying that H is measure-preserving (with respect to the measure μ).

Definition **3.11** ***Unipotent flow***
Let $\{u^t\}_{t\in\mathbb{R}} = U$ be a unipotent one-parameter subgroup of a Lie group G. A flow defined on $\Gamma\backslash G$ by $\phi_t(\Gamma x) = \Gamma x u^t$, is called a unipotent flow.

Remark. We say that the measure μ on the space X is invariant under the flow ϕ_t if $\mu(\phi_t^{-1}(A)) = \mu(A)$, for every measurable subset A of X, and every $t \in \mathbb{R}$. We sometimes say that ϕ_t is measure-preserving instead of μ being ϕ_t-invariant.

Example **3.12** ***Examples of Smooth flows***
Define the one-parameter subgroups $u, a : \mathbb{R} \longrightarrow \mathrm{SL}(2, \mathbb{R})$ by

$$u^t = \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix},$$

$$a^t = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}.$$

It is clear that $u^s u^t = u^{s+t}$ and $a^s a^t = a^{s+t}$. Let Γ be a subgroup of G and define the flows η_t and γ_t on $\Gamma \backslash G$ by

$$\eta_t(\Gamma x) = \Gamma x u^t,$$
$$\gamma_t(\Gamma x) = \Gamma x a^t.$$

They are seen to be flows since

$$\eta_s(\eta_t(\Gamma x)) = \eta_s(\Gamma x u^t) = \Gamma x u^t u^s = \Gamma x u^{t+s} = \eta_{t+s}(\Gamma x)$$

(analogously for γ_t) and they act smoothly.

Remark. The two flows in the above example are so important that they have been named the horocycle flow (η_t) on $\Gamma \backslash G$ and the geodesic flow (γ_t) on $\Gamma \backslash G$.

Definition **3.13** ***Ergodic flow***
A measure-preserving flow ϕ_t on a probability space (X, μ) is said to be ergodic if, for each ϕ_t-invariant subset A of X ($\phi_t(A) = A$ for all $t \in \mathbb{R}$), we have either $\mu(A) = 0$ or $\mu(A) = 1$.

This means that there is only one distinct ϕ_t-orbit in X on which μ is non-zero. Why? Let $x \in X$ and let $\{\phi_t(x)\}_{t \in \mathbb{R}} = \phi_{\mathbb{R}}(x)$ be the orbit of x, then $\phi_{\mathbb{R}}(x)$ is ϕ_t-invariant:

Let $y \in \phi_{\mathbb{R}}(x)$, then $y = \phi_s(x)$ for some $s \in \mathbb{R}$, and we have that $\phi_t(y) = \phi_t(\phi_s(x)) = \phi_{t+s}(x) \in \phi_{\mathbb{R}}(x)$ for all $t \in \mathbb{R}$, and given a $t \in \mathbb{R}$ we have that $\phi_t(\phi_{s-t}(x)) = \phi_s(x) = y$, so that $\phi_t(\phi_{\mathbb{R}}(x)) = \phi_{\mathbb{R}}(x)$. Since orbits are either equal or disjoint, the conclusion follows.

Definition **3.14** ***Mixing***
A measure-preserving transformation $T : X \to X$ is called mixing w.r.t. a T-invariant probability measure μ if $\lim_{n \to \infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B)$ for all measurable sets A and B.

Remark. Note that it makes sense to say that a flow ϕ_t is mixing according to the above definition, since taking $T = \phi_1$, we see that for an integer n and a real number $0 \le r < 1$,

$$\mu(\phi_{-n+r}(A) \cap B) = \mu(T^{-n}(\phi_r(A)) \cap B) \xrightarrow[n \to \infty]{} \mu(\phi_r(A))\mu(B) = \mu(A)\mu(B).$$

A flow (or transformation) being mixing means that it distributes any measurable set randomly in the limit (asymptotically). The term "mixing" comes from the analogy to an ordinary mixer.

We will give an equivalent definition of mixing that is easier to check against. We say that a set of functions Φ is complete if the linear span of the set $L(\Phi)$ is dense in L^2.

Proposition **3.15**
A m.p.t. $T : X \to X$ is mixing if and only if for any given complete system of functions Φ in $L^2(X)$ and any $\phi, \psi \in \Phi$:

$$\langle U_T^n \phi, \psi \rangle = \int_X \phi(T^n x)\overline{\psi(x)}\, \mathrm{d}\mu \longrightarrow \int_X \phi(x)\, \mathrm{d}\mu \int_X \overline{\psi(x)}\, \mathrm{d}\mu = \langle \phi, 1 \rangle \langle 1, \psi \rangle \text{ as } n \to \infty \tag{2}$$

We note that both sides of (2) are linear w.r.t. ϕ and antilinear w.r.t. ψ. Hence, if (2) holds for $\phi,\psi \in \Phi$, then it holds for any $\phi,\psi \in L(\Phi)$. We proceed by showing that if it holds for $L(\Phi)$, then it holds for its closure L^2. This then allows us to prove the theorem by only proving it for the characteristic functions.

Proof. Suppose that (2) holds for a complete system of functions Φ (and hence for $L(\Phi)$, and let $f,g \in L^2$. Since $L(\Phi)$ is dense in L^2, for any $\epsilon > 0$, we can find functions $f',g' \in L(\Phi)$ such that

$$\|f - f'\|_2 < \epsilon, \|g - g'\|_2 < \epsilon.$$

Now

$$\begin{aligned}
|\langle U_T^n f, g\rangle - \langle f,1\rangle \langle 1,g\rangle| = & \\
&= |\left\langle U_T^n f, g-g'\right\rangle + \left\langle U_T^n (f-f'), g'\right\rangle + \left\langle U_T^n f', g'\right\rangle - \\
&- \langle f,1\rangle \left\langle 1, g-g'\right\rangle - \left\langle f-f',1\right\rangle \left\langle 1,g'\right\rangle - \left\langle f',1\right\rangle \left\langle 1,g'\right\rangle | \le \\
&\le \|f\|_2 \cdot \|g-g'\|_2 + \|f-f'\|_2 \cdot \|g'\|_2 + \\
&+ \|f\|_2 \cdot \|g-g'\|_2 \cdot \|1\|_2^2 + \|f-f'\|_2 \cdot \|g'\|_2 \cdot \|1\|_2^2 + |\left\langle U_T^n f', g'\right\rangle - \left\langle f',1\right\rangle \left\langle 1,g'\right\rangle | < \\
&< \epsilon \left(\|f\|_2 + \|g'\|_2 + \|f\|_2 + \|g'\|_2\right) + |\left\langle U_T^n f', g'\right\rangle - \left\langle f',1\right\rangle \left\langle 1,g'\right\rangle |.
\end{aligned}$$

Since $\epsilon > 0$ is arbitrary and $|\langle U_T^n f', g'\rangle - \langle f',1\rangle \langle 1,g'\rangle| = 0$ by assumption, the conclusion follows.

If a m.p.t. T is mixing, then clearly for any measurable sets A and B:

$$\int\limits_X \chi_A \circ T^n \chi_B \,\mathrm{d}\mu = \int\limits_X \chi_{T^{-n}A}\chi_B \,\mathrm{d}\mu = \mu(T^{-n}(A) \cap B) \xrightarrow[n\to\infty]{} \mu(A)\mu(B) = \int\limits_X \chi_A \,\mathrm{d}\mu \int\limits_X \chi_B \,\mathrm{d}\mu$$

Since the set of characteristic functions is complete, (2) must hold for all functions in L^2, and hence for any complete system of functions Φ. Clearly, by reversing the argument above, we see that (2) implies mixing. □

THEOREM **3.16** ***Birkhoff's Pointwise Ergodic Theorem for flows***
Let μ be probability measure (on X), ϕ_t an ergodic measure-preserving flow, and $f \in L^1(X,\mu)$. Then

$$\lim_{T\to\infty} \frac{1}{T} \int\limits_0^T f(\phi_t(x))\,\mathrm{d}t = \int\limits_X f \,\mathrm{d}\mu,$$

for a.e. $x \in X$.

Proof. We may assume that $f \ge 0$. For the general case, consider the composition of $f = f^+ - f^-$ in Prop. 3.2. Set

$$F(x) = \int\limits_0^1 f(\phi_t(x))\,dt,$$

then $F(x)$ is in L^1, by using Theorem 2.16.4 in [Fri82] (a version of Fubini's theorem asserting integrability where only one of the double integrals is absolutely integrable) and the fact that Prop. 3.2 implies

$$\int\limits_0^1 \left(\int_X f(\phi_t(x))\,d\mu\right) dt = \int\limits_0^1 \left(\int_X f(x)\,d\mu\right) dt = 1 \cdot \left(\int_X f(x)\,d\mu\right) < \infty.$$

Using the notation $[T]$ to mean the integer part of T, rewrite the limit as

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(\phi_t(x))\,dt = \lim_{T\to\infty} \frac{1}{[T]} \cdot \frac{[T]}{T} \left(\sum_{n=0}^{[T]-1} F(\phi_n(x)) \right) + \frac{1}{[T]} \int_{[T]}^T f(\phi_t(x))\,dt =$$

$$= \int_X F(x)\,d\mu,$$

where, in the last step, we have used the discrete version of Birkhoff's pointwise ergodic theorem, and the fact that

$$\lim_{T\to\infty} \int_{[T]}^T f(\phi_t(x))\,dt = 0$$

(which also follows from that same theorem). □

The following theorem tells us that we may study any invariant measure by considering only the ergodic invariant measures. This is done by breaking up the invariant measures into smaller parts that are ergodic as well.

Theorem **3.17** ***Ergodic Decomposition Theorem***
Let G be a Lie group acting smoothly on the space $X = \Gamma\backslash G$ and let μ be a G-invariant Borel measure on X. Then there is a measure space (Y, ν) and a partition of X into measurable G-invariant subsets $X_y, y \in Y$, and measures μ_y on X_y such that:

- For any measurable subset $A \subseteq X$, we have that $A \cap X_y$ is measurable w.r.t. μ_y for almost all $y \in Y$ (w.r.t. to the measure ν) and $\mu(A) = \int_Y \mu_y(A \cap X_y)\,\mathrm{d}\nu(y)$
- For almost all $y \in Y$, the action of G on X_y is ergodic w.r.t. the measure μ_y.

Proof. See the proof of Theorem 8.20 in [EW11, p. 271]. □

4 Ratner's Theorems

In a series of articles of 3 articles, [Rat90b], [Rat90a] and [Rat91] (together totalling more than 150 pages), Ratner proved a long-standing conjecture of Raghunathan regarding certain orbits on homogeneous spaces. For historical remarks, read our introductory section.

The theorems concern some arbitrary (real) Lie group G, any discrete subgroup Γ of G, and any unipotent subgroup U of G. The quotient manifold $\Gamma\backslash G = \{\Gamma g : g \in G\}$ is a homogeneous space, and the theorems tell us what the closure of xU, for some arbitrary $x \in G$, looks like when we map it to $\Gamma\backslash G$ by the natural projection, $\pi : G \to \Gamma\backslash G : g \mapsto \Gamma g$.

Whenever U is a unipotent one-parameter subgroup of G, we will represent U by the family $\{u^t\}_{t\in\mathbb{R}}$, with the corresponding flow on $\Gamma\backslash G$ given by $\phi_t(\Gamma x) = \Gamma x u^t$.

We begin by making an important definition.

Definition **4.1** ***Homogeneous measure***
A Borel probability measure μ on $\Gamma\backslash G$ is called a homogeneous measure if there exists an $x \in \Gamma\backslash G$ and a closed subgroup $H \subseteq G$ such that xH is homogeneous (w.r.t. H) and μ is an H-invariant Borel probability measure supported on xH.

The main theorem is:

Theorem **4.2** ***Ratner's Measure Classification Theorem***
Let G be a Lie group, Γ a discrete subgroup in G and U a unipotent subgroup of G. Then every ergodic U-invariant probability measure on $\Gamma\backslash G$ is homogeneous.

The following two theorems can be obtained as corollaries from the first one (see for instance [Mor05]).

THEOREM **4.3** ***Ratner's Orbit Closure Theorem***
Let G be a Lie group and Γ a lattice in G ($\Gamma\backslash G$ has finite volume). If ϕ_t is a unipotent flow on $\Gamma\backslash G$, then the closure of every ϕ_t-orbit is homogeneous.

This means that if $U = \{u^t\}$ is the underlying unipotent subgroup of ϕ_t, then the closure $\overline{xU} = xS$ (and has finite volume), and $U \subseteq S$, for some connected, closed subgroup S of G. This means that the ϕ_t-orbit of $[x]$ is dense in (or everywhere present in) [xS].

THEOREM **4.4** ***Ratner's Equidistribution Theorem***
Let G be a Lie group and Γ a lattice in G such that $\Gamma\backslash G$ has finite volume. If ϕ_t is a unipotent flow on $\Gamma\backslash G$, then the orbit of ϕ_t is equidistributed in the set $[xS]$.

We will sketch the proof of Ratner's Measure Classification Theorem for the case of $G = \mathrm{SL}(2,\mathbb{R})$. The technical details can be found in her article [Rat92] or Starkov's book [Sta00]. The proof is very technical and a lot of details are implicit, so the aim of this section is to add some of these omitted details and explain the ideas behind the proof.

Since any unipotent flow on $G = \mathrm{SL}(2,\mathbb{R})$ is conjugate to the horocyclic flow U_t, there are two possible ways to prove the theorem for this special case. One way is to use properties of the horocyclic flow, but that would involve techniques that can not be generalized to general unipotent flows. The proof given here is essentially an outline of some of the ideas used in her proof for unipotent flows, without the burdening technicalities that are necessary to handle a general Lie group G.

4.1 Proof of Ratner's measure classification theorem for $G = \mathrm{SL}(2,\mathbb{R})$

Let the following subgroups of $G = \mathrm{SL}(2,\mathbb{R})$ be given

$$U = \{U_t = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} : t \in \mathbb{R}\},$$

$$A = \{A_t = \begin{bmatrix} e^t & 0 \\ 0 & e^{-1} \end{bmatrix} : t \in \mathbb{R}\} \text{ and}$$

$$H = \{H_t = \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix} : t \in \mathbb{R}\}.$$

Since matrices can be written in upper-triangular form, every unipotent element of $\mathrm{SL}(2,\mathbb{R})$ is conjugate to U_t for some t. These subgroups satisfy the following important commutation relations

$$U_s A_\tau = A_\tau U_{se^{-2t}}$$
$$H_s A_\tau = A_\tau H_{se^{2t}}$$

for every $s, t \in \mathbb{R}$.

Now we form the subgroups $W = AH$ and $B = AU$ and their respective neighbourhoods

$$W(\delta) = \{A_\tau H_b : |\tau| \leq \delta, |b| \leq \delta\} \text{ and}$$

$$B(\delta) = \{A_\tau U_s : |\tau| \leq \delta, |s| \leq \delta\}.$$

Let $y \in xW(\delta)$, that is $y = xA_\tau H_b$ for some $|\tau| < \delta, |b| < \delta$. If $\delta > 0$ is sufficiently small, then for any $y = xA_\tau H_b \in xW(\delta)$ and any $0 \leq s \leq 1$, there is a unique(!) function $\alpha(y,s)$ that is strictly increasing in s, continuous in (y,s) and satisfying the condition $\alpha(y,0) = 0$, such that $yU_{\alpha(y,s)} \in xU_s W(10\delta)$. Solving some equations, we see that it is given by

$$\alpha(y,s) = \frac{s}{e^{2\tau} - sb}$$

and that $yU_{\alpha(y,s)} = xU_sA_{\tau(y,s)}H_{b(y,s)}$, where

$$\tau(y,s) = \ln(e^{\tau} - sbe^{2\tau})$$

$$b(y,s) = b(1 - bse^{-2\tau})$$

are the respective coordinates in the directions of A and H. As can be seen, these functions are defined only for $sb < e^{2\tau}$. This tells us that the divergence between points close to eachother in some W-leaf will be very slow in the W-direction, but proportional to time in the U-direction. In particular, it says that if $b = 0$, then the two orbits will diverge only in the U-direction.

If however $b \neq 0$, then there will be a point in time $s_\theta > 0$, such that there is a $y \in xW(\delta)$ where $yU_{\mathbb{R}}$ does not intersect $xU_{\mathbb{R}}W$. The important part here is that this ciritcal point in time may be made arbitrarily large by making δ smaller.

The so-called R-property in [Rat90b] captures this behaviour more exactly, and plays a crucial role in the proof of Ratner's measure-classification theorem.

Lemma **4.5** ***R-property of the horocycle flow***
There exist constants $0 < \eta < 1$ and $C > 1$ such that if for some $t > 1$

$$|\tau(y,t)| = \theta \text{ and } |\tau(y,s)| \leq \theta \text{ for } 0 \leq s \leq t,$$

where $y \in xW(\delta)$, $0 < \delta < \theta/10$, then

$$\theta/2 \leq |\tau(y,s)| \leq \theta, |b(y,s)| \leq C\theta/s$$

for all $s \in [(1-\eta)t, t]$.

Let λ be the Lebesgue measure on $\mathbb{R}$ and $\phi_y(s) = \alpha(y,s)$. The following lemma tells us that s and $\alpha(y,s)$ will be close to each other.

Lemma **4.6**
For any $\epsilon > 0$, there is a $\theta = \theta(\epsilon) > 0$ such that, given any $y \in xW(\theta)$ and any Borel set $C \subseteq [0, s_\theta]$

$$\left| \frac{\lambda(C)}{\lambda(\phi_y(C))} - 1 \right| < \epsilon.$$

Let μ be an ergodic U-invariant Borel probability measure on $X = \Gamma\backslash G$ and let $\Lambda = \Lambda(\mu) = \{g \in G :$ the action of g on X preserves $\mu\}$. It is clear from the definition of μ that $U \subseteq \Lambda(\mu)$.

Proposition **4.7**
The set $\Lambda = \Lambda(\mu)$ as defined above is a closed subgroup of G.

Proof. Suppose that $\{g_n\} \subseteq \Lambda$ is a sequence converging to g. According to Lusin's theorem, the space of continuous functions with compact support is dense in L^p, and so it suffices to prove that

$$\int f(gx)\, d\mu = \int f(x)\, d\mu$$

for all bounded continuous functions with compact support, f. This follows easily from the fact that $f(g_nx) \to f(x)$ for a.e. x, and Lebesgue's monotone convergence theorem:

$$\int f(x)\, d\mu = \lim_{n\to\infty} \int f(g_nx)\, d\mu = \int f(gx)\, d\mu.$$

□

The proof will basically be split into two cases: the one were μ is also A-invariant and the one were μ is not A-invariant.

LEMMA **4.8**
If $A \not\subset \Lambda$, there is a $Y \subset X$ such that $\mu(Y) = 1$ and $Y \cap Yq = \emptyset$ for every $q \in B - \Lambda$.

Proof. We begin by showing that for any $q \in B - \Lambda$, there is a set $X_q \subseteq X$, $\mu(X_q) = 1$, and $\epsilon(q) > 0$ such that

$$X_q \cap X_q g = \emptyset$$

for every $g \in qB_{\epsilon(q)}(e) = B_{\epsilon(q)}(q)$, where $B_\epsilon(x)$ is the ϵ-ball in B around the point x, that is the set of all points in xB at a distance less than ϵ from x.

We then cover $B - \Lambda$ by a countable number of such neighbourhoods $B_{\epsilon(q_i)}(q_i)$, for $q_i \in B - \Lambda, i = 1, 2, 3, \ldots$, such that

$$B - \Lambda \subseteq \bigcup_{i=1}^{\infty} B_{\epsilon(q_i)}(q_i).$$

Letting $Y = \bigcap_{i=1}^{\infty} X_{q_i}$ for the X_{q_i} corresponding to q_i as defined above, we see that

$$\mu(Y) = 1 \text{ and } Y \cap Yg = \emptyset \text{ for every } g \in B - \Lambda$$

To find our set X_q for some arbitrary $q \in B - \Lambda$, we begin by noting that the measure

$$\mu_q(E) = \mu(Eq) \text{ for each Borel set } E \subseteq X$$

is different from μ (since q doesn't preserve it by assumption) but that the action of U on (X, μ_q) ergodic. This implies that μ and μ_q are mutually singular - there is a set E_q with $\mu(E_q) = 1$ and $\mu(E_q q) = \mu_q(E_q) = 0$. By taking $E'_q = E_q - E_q q$ we see that $\mu(E'_q) = 1$ and $E'_q \cap E'_q q = \emptyset$.

Now, we choose a compact set $K \subseteq E'_q$ such that K has almost full measure, that is $\mu(K) > 0.99$ (say). Since E'_q and $E'_q q$ are disjoint and K contains all its limit points, we see that there is a $\epsilon(q) > 0$ such that

$$d_X(K, Kq) \geq \epsilon(q).$$

That is, moving any point in K by q, moves it outside of K by some positive distance. Since U acts ergodically on (X, μ), by Birkhoff's pointwise ergodic theorem, there is a set $X_q \subseteq X$, $\mu(X_q) = 1$ such that every point in X_q spends most of its time inside of K (at least 99 percent of its time).

Now we prove that $X_q \cap X_q g = \emptyset$ for every $g \in B_{\epsilon(q)}(q)$. To prove this, suppose that $X_q \cap X_q g \neq \emptyset$, that is $x = yg$ for some $x, y \in X_q$.

But since x and y lie in X_q, we see that for sufficiently large $t > 0$, the orbits of the two points x and y will spend almost all their time in K. There is then some $0 \leq s \leq t$ such that $yU_s = z \in K$ and $yU_s g = zg \in K$. However, $zg = zqp$ for some $p \in B_{\epsilon(q)}(q)$, and so

$$d_X(K, KA_\tau) \leq d_X(zqp, zq) < \epsilon(q),$$

a contradiction. The conclusion is that Y satisfies the assumptions in the theorem. □

THEOREM **4.9**
If $A \not\subset \Lambda$, then there is an $x \in X$ such that μ is supported on the closed (periodic) orbit xU.

Proof. Since Λ is a closed subgroup, there will be a $0 < \theta < 0.1$ (the lsat inequality is just a technical assumption), such that $A_\tau \notin \Lambda$ for every $0 < |\tau| \leq \theta$. We may further assume that $\theta < \theta(0.1)$, where $\theta(0.1)$ is as in 4.6.

Let $Y \subset X$ ($\mu(Y) = 1$) be as in 4.8. As in the lemma, let K be a compact set of almost full measure, and $\delta > 0$ (depending on K) such that

$$d_X(K, KA_\tau) \geq \delta$$

for every sufficiently large $|\tau|$ (for technical reasons we choose the bounds $\frac{\theta}{2} \leq |\tau| \leq \theta$).

Since the action of U is ergodic, there is a set $F \subseteq X$ of positive measure ($\mu(F) > 0$), such that, after some sufficiently long time ($t_0 \geq 1$ in the article), the U-orbit of every $x \in F$ will have spent almost all its time in K (an amount of time approximately equal to the measure of K).

What we want to show now is that there is a small neighbourhood $N(x) \cap F$ of some point x in F of positive measure that looks like some small piece of the orbit xU, say $xU(\xi)$. This gives us that the small piece $xU(\xi)$ has positive measure, and so take a sufficiently large finite union

$$P = \{xU_s \mid -\xi \leq s \leq (2N+1)\xi + r\} = \left(\bigcup_{n=0}^{N} xU(\xi)U_{2n\xi}\right) \bigcup xU(r/2)U_{(2N+1)\xi+r/2}$$

of those pieces, ensuring that $\mu(P) = 1$. Now, since $PU_{\mathbb{R}}$ doesn't change the measure, the orbit must be periodic (with period $2(N+1)\xi + r$).

How do we show that there is a small neighbourhood $N(x) \cap F$ of some point x in F of positive measure that looks like some small piece of the orbit xU?

We do this by choosing ξ to be some extremely small quantity (several magnitudes smaller than our θ), and let $x, y \in F$ be such that $d_X(x,y) < \xi$. The thing to show now is that

$$y \in xB(\xi).$$

We do so by assuming that $y \notin xB(\xi)$, that is $y = xA_\tau H_b \in xW(\xi)$, where $b \neq 0$. Since $|\tau|$ is increasing, there will be some time t, such that

$$|\tau(y,t)| = \theta = \max\{|\tau(y,s)| \mid 0 \leq s \leq t\}.$$

The value θ will be attained since θ was chosen small. Since $x, y \in F$, they will both spend almost all their time in K, so there will be some point in time s very close to, but smaller than t (chosen sufficiently close to t to enable us to use the R-property, which can be made possible by choosing the measure of K to be sufficiently large), such that the U-orbit of x will be in K at time s, and that of y will be in K at time $\alpha(y,s)$; that is

$$xU_s \in K \text{ and } yU_{\alpha(y,s)} = xU_s A_{\tau(y,s)} H_{b(y,s)} \in K.$$

This of course means that

$$xU_s A_{\tau(y,s)} \in KA_{\tau(y,s)},$$

so that the distance

$$d_X(K, KA_{\tau(y,s)}) \leq |b(y,s)|.$$

By the R-property, we must have that for some constant $C \geq 1$, that

$$\frac{\theta}{2} \leq |\tau(y,s)| \leq \theta, \qquad |b(y,s)| \leq C\theta/t \leq 0.1\delta.$$

The last inequality follows because ξ was chosen sufficiently small to ensure that bound (see her article for technical details). Now we see that

$$d_X(K, KA_{\tau(y,s)}) \leq 0.1\delta,$$

in contradiction to our initial assumption that δ be chosen to satisfy

$$d_X(K, KA_{\tau(y,s)}) \geq \delta.$$

We conclude that if $x, y \in F$ are such that $d_X(x,y) < \xi$, then $y \in xB(\xi)$.

Since G can be covered by countably many neighbourhoods $O_\epsilon(x)$, there will be some $x \in F \cap Y$ (we take the intersection with Y to simplify our last step), such that

$$\mu(O_\epsilon(x) \cap F) > 0$$

for every $\epsilon > 0$. This of course implies that, since for sufficiently small $\epsilon > 0$, $O_\epsilon(x) \cap F = xB(\epsilon) \cap F$, that $\mu(xB(\xi)) > 0$. We must then have that

$$\mu(xB(\xi) \cap Y) > 0$$

and

$$xB(\xi) \cap Y = xU(\xi)$$

since Y and YA_τ are disjoint for small $|\tau| \le \xi$. We conclude that

$$0 < \mu(xB(\xi) \cap Y) = \mu(xU(\xi) \cap Y) = \mu(xU(\xi)),$$

implying that we can cover the U-orbit of x, xU with a finite number of translated small pieces $xU(\xi)$, so xU is periodic (and closed by our discussion).

□

PROPOSITION **4.10**
If $A \subset \Lambda$, then A is mixing.

Proof. See [Rat92, p. 26] or [Sta00, p. 107]. □

THEOREM **4.11**
Suppose that $A \subset \Lambda$. Then Γ is a lattice and μ is G-invariant, meaning that $\nu < \infty$ and $\mu = \nu/\nu(X)$.

Proof. Let f be a continuous function on X with compact support. Since the action of A on (X,μ) is ergodic, there is a set $C_f \subseteq X$, consisting of points $y \in X$ for which

$$S_{f,n}(y) = \frac{1}{n}\sum_{i=0}^{n-1} f(yA_{-i}) \to f_\mu = \int_X f\, d\mu,$$

that is of full μ-measure.

Since H-orbits are the contracting horocycles for geodesics in the negative direction, we see that for any $z \in yH$, $d_X(yA_{-n}, zA_{-n}) \to 0$ as $n \to \infty$. Since f is uniformly continuous and $S_{f,n}(y) \to f_\mu$, it follows that $S_{f,n}(z) \to f_\mu$, and so $C_f H = C_f$.

Now we need to prove that C_f is of full ν-measure. To do so we consider first the neighbourhood

$$O_\delta(x) = xB(\delta/2)H(\delta) \cap xH(\delta/2)B(\delta)$$

for some sufficiently small $\delta > 0$, and the decomposition of μ on this neighbourhood into conditional measures $\mu_y(E) = \mu(yB(\delta/2) \cap E)$ on the leaves $yB(\delta/2)$, $y \in xH(\delta)$. Since μ is B-invariant, so will almost every μ_y be, hence almost every μ_y is the Lebesgue measure on $yB(\delta/2)$.

Since C_f is of full μ-measure, and $C_f \cap xB(\delta/2) = C_f h \cap xB(\delta/2)h = C_f \cap xB(\delta/2)h$ for every $h \in H$ it follows that $C_f \cap O_\delta(x)$ has the same Lebesgue measure as $O_\delta(x)$. Since ν is "the" Lebesgue measure up to a constant, C_f must be of full ν-measure.

Now, we further assume that f is non-negative and $f_\mu > 0$. By Fatou's lemma

$$f_\mu \nu(X) = f_\mu \nu(C_f) = \int_{C_f} f_\mu\, d\nu \le \lim_{n\to\infty} \int_{C_f} S_{f,n}\, d\nu = \int_{C_f} f\, d\nu = \int_X f\, d\nu < \infty$$

where we have used that ν is A-invariant in order to evaluate the limit to $\int_X f\, d\nu$.

We now wish to prove that $\mu = \nu/\nu(X)$. In order to do so we turn to Lebesgue's Dominated Convergence Theorem to get that

$$f_\nu = \int_X f\, d\nu = \int_{C_f} f\, d\nu = \int_{C_f} S_{f,n}\, d\nu \to \int_{C_f} f_\mu\, d\nu = f_\mu \nu(X)$$

for every uniformly continuous function f with compact support, showing that $\mu = \nu/\nu(X)$. □

5 Applications

5.1 The Oppenheim Conjecture

THEOREM **5.1** ***Oppenheim's Conjecture***
Let B be a real non-degenerate indefinite quadratic form in $n \geq 3$ variables. Suppose that B is not a multiple of a form with rational coefficients. Then $B(\mathbb{Z}^n)$ is dense in $\mathbb{R}$.

The Oppenheim Conjecture was made by Oppenheim in 1929 for the case of $n \geq 5$, and subsequently extended to the case $n \geq 3$ by Davenport in 1946.

In the 1970's, Raghunathan realized that the case $n = 3$ can be stated as a problem in homogeneous dynamics on the space $\mathrm{SL}(3,\mathbb{R})/SL(3,\mathbb{Z})$. The other cases follow from $n = 3$.

We will sketch how one may use Ratner's orbit closure theorem to prove the Oppenheim conjecture. For the details and the full proof, which uses the theory of algebraic groups, consult [Mor05].

Sketch of the proof of Oppenheim's Conjecture. Let $G = \mathrm{SL}(n,\mathbb{R})$, $\Gamma = \mathrm{SL}(n,\mathbb{Z})$, and the stabilizer of Q,

$$H = SO(Q) = \{h \in \mathrm{SL}(n,\mathbb{R}) \mid Q(vh) = Q(v) \text{ for all } v \in \mathbb{R}^n\}$$

be given. The subgroup Γ fixes the lattice $\mathbb{Z}^n$, that is $\Gamma\mathbb{Z}^n = \mathbb{Z}^n$. The only two closed subgroups of G containing H, is H itself and G. Applying Ratner's orbit closure theorem on $z = \Gamma e \in \Gamma\backslash G$, we see that either $\overline{zH} = zH$, or $\overline{zH} = zG$. The first case $\overline{zH} = z$ implies that Q is a multiple of a rational form, so the second case $\overline{zH} = zG$ must be true. We then see that ΓH is dense in G, and so

$$\begin{aligned}\overline{Q(\mathbb{Z}^n)} &= \overline{Q(\mathbb{Z}^n\Gamma)} \\ &= \overline{Q(\mathbb{Z}^n\Gamma H)} \\ &= Q(\mathbb{R}^n\backslash 0) = \mathbb{R}.\end{aligned}$$

In the first step, Γ fixes Z^n. In the second, Q is invariant under H. In the third, we have used that the image of $v \in \mathbb{R}^n\backslash 0$ under G, that is vG, is $\mathbb{R}^n$. □

A Measure theory and functional analysis

Most of this is standard material and is covered, for example, in [Rud], [Fri82] and [AB06].

Definition **A.1** ***Topological space***
A topological space is a pair $(X,\mathscr{T})$ where X is a set and $\mathscr{T}$ a collection of subsets of X satisfying the following axioms:

A1 $X \in \mathscr{T}$ and $\emptyset \in \mathscr{T}$.

A2 Given an arbitrary collection of $\{X_\alpha \in \mathscr{T}\}$, the union $\bigcup X_\alpha$ is in $\mathscr{T}$.

A3 Given a finite collection of $\{X_\alpha \in \mathscr{T}\}$, the intersection $\bigcap X_\alpha$ is in $\mathscr{T}$.

The elements of $\mathscr{T}$ are called open sets, and we usually refer to X as a topological space when the topology $\mathscr{T}$ is implied. A subset whose complement is an open set, is called a closed set.

Definition **A.2** ***σ algebra***
A σ-algebra on a set X is a collection $\mathscr{A}$ such that:

A1 $X \in \mathscr{A}$.

A2 If $A \in \mathscr{A}$, then $A^c \in \mathscr{A}$ (A^c is the complement of A relative to X).

A3 If $A = \bigcup_k^\infty A_\alpha$, and $A_k \in \mathscr{A}$ for all $k = 1,2,3,\ldots$, then $A \in \mathscr{A}$.

Definition **A.3** ***Measure***
A (positive) measure μ on a σ-algebra $(X,\mathscr{A})$, is a map $\mu : \mathscr{A} \to [0,\infty]$, such that:

A1 $\mu(\emptyset) = 0$.

A2 $\mu(\bigcup_k^\infty A_k) = \sum_k^\infty \mu(A_k)$, if $\{A_k \in \mathscr{A} \mid k = 1,2,3,\ldots\}$ is a countable collection of pairwise disjoint sets.

Definition **A.4** ***Measurable space***
A measurable space is a triple $(X,\mathscr{A},\mu)$, where $\mathscr{A}$ is a σ-algebra of X and μ is a measure on $\mathscr{A}$.

Definition **A.5** ***Borel σ-algebra***
Let G be a topological space. A subset E of G that can be obtained by a combination of taking countable unions, countable intersections and relative complements [8] of the open sets of G is a Borel set. The Borel σ-algebra is the collection of all Borel sets in G (they form a σ-algebra).

Definition **A.6** ***Measurable map***
A map $f : (X,\mathscr{A},\mu) \to (Y,\mathscr{B},\nu)$ between two measurable spaces is called measurable if for every $B \in \mathscr{B}$, $T^{-1}B = \{x \in X : Tx \in B\} = A \in \mathscr{A}$.

Definition **A.7** ***Borel probability space***
A measurable space $(X,\mathscr{B},\mu)$ where X is a topological space, $\mathscr{B}$ its Borel σ-algebra and μ a probability measure ($\mu(X) = 1$) is called a Borel probability space.

Definition **A.8** ***Characteristic function***
Let A be a measurable set of X. The function

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

is called the characteristic function associated with A.

[8]The relative complement of a set B by a set A is the set $B \backslash A$

Definition A.9 ***Simple function***
Let $A_1, \cdots, A_n$ be measurable sets and $a_1, \cdots, a_n$ be real numbers such that a_i is non-zero if and only if $\mu(A_i) < \infty$. A function

$$f(x) = \sum_{i=1}^{n} a_i \chi_{A_i}(x)$$

is called a simple function.

Simple functions are measurable since for any open set $O \subseteq \mathbb{R}$, $f^{-1}(O)$ is a union of measurable sets in X.

Let $f(x) = \sum\limits_{i=1}^{n} a_i \chi_{A_i}(x)$ be a simple function. The sum $\sum\limits_{i=1}^{n} a_i \mu(A_i) = \int_X f\, d\mu$ is called the integral of f.

Definition A.10 ***Integral***
Let f be measurable function. If there is a sequence $\{f_n\}$ of simple functions that is Cauchy in the mean ($\int |f_n - f_m|\, d\mu \xrightarrow[m,n\to\infty]{} 0$) and converging a.e. to f, then we define the integral of f to be

$$\int_X f\, d\mu = \lim_{n\to\infty} \int_X f_n\, d\mu$$

keeping with the convention that the integral is ∞ if the right-hand side does not converge to a real number.

We may define the integral over any measurable set A by letting $\int_A f\, d\mu = \int_X f \cdot \chi_A\, d\mu$.

Proposition A.11 ***Theorem 1.17 in [Rud]***
Given any non-negative measurable function $f(x)$, there exists a monotone increasing sequence of simple functions $\{s_n\}$ such that

$$0 \le s_1(x) \le s_2(x) \le \cdots \le f(x),$$
$$\text{and } s_n(x) \to f(x) \text{ as } n \to \infty, \text{ for every } x.$$

Definition A.12 ***L^p space***
Let $1 \le p < \infty$ be given, then the measurable functions f such that

$$\|f\|_p = \left(\int_X |f|^p\, d\mu \right)^{1/p} < \infty,$$

form a Hilbert space under the equivalence relation f g if $f(x) = g(x)$ for a.e. x, denoted $L^p(\mu)$. The definition can be extended to $p = \infty$ by considering the measurable functions f such that the essential supremum[9] exists, that is

$$\|f\|_\infty < \infty.$$

Under the same equivalence relation, this is a Hilbert space denoted $L^\infty(\mu)$. The norm of a function $f \in L^p(\mu)$ is simply

$$\|f\|_p = \left(\int_X |f|^p\, d\mu \right)^{1/p},$$

if $1 \le p < \infty$, or the essential supremum

$$\|f\|_\infty$$

if $p = \infty$.

[9]The essential supremum $\|f\|_\infty$ of a measurable function f is the smallest constant C such that $|f| \le C$ for a.e. x. If such a constant does not exist we set $\|f\|_\infty = \infty$.

Proposition **A.13** ***Hölder's inequality***
If $\frac{1}{p}+\frac{1}{q}=1$, and $f \in L^p(\mu)$ and $g \in L^q(\mu)$, then $fg \in L^1(\mu)$ and

$$\int_X |fg|\,d\mu = \|fg\|_1 \leq \|f\|_p \cdot \|g\|_q.$$

Proof. See [Fri82] or any other book on functional analysis. □

The following proof is from [AB06].

Proposition **A.14**
Let μ be a finite measure, then for every $1 \leq p < q \leq \infty$, we have that

$$L^q(\mu) \subseteq L^p(\mu)$$

If moreover μ is a probability measure, then

$$\|f\|_p \leq \|f\|_q.$$

Proof. If f is a measurable, essentially bounded function, then

$$\|f\|_p = \left(\int_X |f(x)|^p\,d\mu\right)^{1/p} \leq \left(\int_X \|f\|_\infty^p\,d\mu\right)^{1/p} = \|f\|_\infty\,(\mu(X))^{1/p},$$

hence the theorem follows for $q=\infty$. Suppose then that $1 \leq p < q < \infty$, and set $r=\frac{q}{p}>1, s=\frac{q}{q-p}>1$. Simple calculations show that

$$\frac{1}{r}+\frac{1}{s}=1.$$

Now let $f \in L^q(\mu)$, so that $(|f|^p)^r = |f|^q \in L_1(\mu)$. Since

$$\left(\int_X (|f|^p)^r\,d\mu\right)^{1/r} = \left(\int_X (|f|^q\,d\mu\right)^{p/q} = (\|f\|_q)^p < \infty,$$

we see that $f \in L_r(\mu)$. Since μ is finite, $1 \in L_s(\mu)$, and so using Hölder's inequality, we get

$$\begin{aligned}(\|f\|_p)^p = \int_X |f|^p\,d\mu &\leq \left(\int_X (|f|^p)^r\,d\mu\right)^{1/r} \cdot \left(\int_X |1|^s\,d\mu\right)^{1/s} = \\ &= \left(\int_X |f|^q\,d\mu\right)^{p/q} \cdot \mu(X)^{1/s} = (\|f\|_q)^p \cdot \mu(X)^{1/s},\end{aligned}$$

or

$$\|f\|_p = \|f\|_q \cdot \mu(X)^{1/ps}.$$

Now both conclusions follow. □

Proposition **A.15**
Let μ be a finite measure, then for every $1 \leq p < q \leq \infty$, we have that $L^q(\mu)$ is dense in $L_p(\mu)$.

Proof. By the above theorem (A.14), we only need to prove that $L^\infty(\mu)$ is dense in $L^1(\mu)$. To accomplish this, suppose that $f \in L^\infty(\mu)$, and consider the measurable functions

$$f_n(x) = \begin{cases} n & \text{when } n \leq f(x), \\ f(x) & \text{when } -n \leq f(x) \leq n, \\ -n & \text{when } f(x) \leq -n. \end{cases}$$

They all belong to $L^\infty(\mu)$ and satisfy the assumptions in Lebesgue's Dominated convergence theorem, hence they converge to f in the L^1-norm. □

Let V be a vector space over some field $\mathbb{F}$ that is either $\mathbb{R}$ or $\mathbb{C}$. We will use the notation $L(\Phi) = \{a_1\phi_1 + \cdots + a_n\phi_n : n \in \mathbb{N}, a_k \in \mathbb{F}, \phi_k \in \Phi, 1 \leq k \leq n\}$ to denote the linear span of a subset Φ of V.

Definition **A.16** ***Complete system of functions***
A complete system of functions Φ in $L^p(X,\mu)$, $1 \leq p \leq \infty$, is just a set of functions $\Phi \subseteq L^p(X,\mu)$ such that $L(\Phi)$ is dense in L^p. This means that given a function $f \in L^p$ and an $\epsilon > 0$, there is a function $\phi \in L(\Phi)$ such that $\|f - \phi\|_p < \epsilon$.

The reason for our interest in complete systems of functions is because they greatly reduce the complexity of proving certain theorems for L^p, especially given the statement below, that says such theorems need only be proved for characteristic functions. This will prove very handy indeed!

Proposition **A.17**
Let μ be a Borel probability measure. Then the set of all characteristic functions χ_A of X, where A is a measurable subset of X, is complete in $L^2(\mu)$.

Proof. First, we note that χ_A is in $L^2(\mu)$ for any measurable set A, since it is integrable and $|\chi_A|^2 = \chi_A$. We note that the linear span of the characteristic functions are the simple functions, and they are, by definition of the integral, dense in $L^1(\mu)$. Since $L^2(\mu)$ is contained in $L^1(\mu)$ by A.14, the conclusion follows. □

B Lie groups and Lie algebras

Two references on the basics of Lie groups are [War83] and [SW73]. A more concrete approach, dealing with linear Lie groups (Lie groups that are also matrix groups), is given in [Ros02].

Definition **B.1** ***Lie group***
Let G be a smooth real manifold G, such that G is also a group for which multiplication $\mu : G \times G \longrightarrow G$ and inversion $g \mapsto g^{-1}$ are both smooth maps. We write gh instead of the more cumbersome $\mu(g,h)$.

Example **B.2** ***Examples of Lie groups***
There are several examples of Lie groups:

- The group $\mathbb{R}$ with ordinary multiplication. It is trivially a manifold and both the operation and its inverse are smooth, since $(x,y) \mapsto xy$ is a polynomial and $x \mapsto \frac{1}{x}$ is a rational function that is smooth on $\mathbb{R}\backslash\{0\}$.
- The group $\mathbb{R}^n$ with ordinary vector addition. It is trivially a manifold and both the operation and its inverse are smooth, since $(x,y) \mapsto x + y$ and $x \mapsto -x$ are both polynomials and hence smooth.
- The General Linear group $GL(n,\mathbb{R})$, the set of the invertible matrices ($\det \neq 0$) with matrix multiplication as its operation.

  We identify the matrices with vectors in $\mathbb{R}^{n^2}$ by assigning to each matrix the vector with the same elements (in a pre-determined order), and give it the topology induced by $\mathbb{R}^{n^2}$, making it into a real manifold.

  Since the operations of multiplication and inversion of matrices are basically polynomial operations, they are smooth. The group $GL(n,\mathbb{R})$ is a disconnected Lie group (since $\det : GL(n,\mathbb{R}) \to \mathbb{R}$ is continuous and attains both negative and positive values, but not 0).
- The Special Linear group $SL(n,\mathbb{R})$, the subgroup of $GL(n,\mathbb{R})$ consisting of those matrices with determinant equal to 1, is a connected Lie group.

Definition **B.3** ***Discrete subgroup***
A Lie subgroup Γ of a Lie group G is said to be discrete if it has the discrete topology relative to G.

A discrete subgroup Γ of G can be considered an "evenly spaced" subset of G such that the points of Γ don't lie too close to eachother in G.

Definition **B.4** ***Lattice***
A discrete subgroup Γ of a Lie group G is said to be a lattice if there is a measurable subset $\mathcal{F}$ of G such that

- $\Gamma\mathcal{F} = G$,
- $\gamma\mathcal{F} \cap \mathcal{F}$ has measure 0, for all $\gamma \in \Gamma \setminus \{e\}$, and
- $\mathcal{F}$ has finite volume.

A lattice Γ in a Lie group G is a discrete space that in a sense "covers" G; that is: Γ is a collection of evenly, but not too closely, spaced out points such that no parts of G are distanced too far away from the points of Γ.

Example **B.5**
An example of a lattice is $\mathbb{Z}$ in $\mathbb{R}$ with the corresponding fundamental domain $[0,1)$ (or $[0,1]$ depending on who you ask). In our examples we have used the fact that $\mathbb{Z}^n$ is a lattice in $\mathbb{R}^n$ with the corresponding fundamental domain $[0,1)^n$.

An example of a discrete subgroup that is not a lattice is $\mathbb{Z} \times 0$ in $\mathbb{R}^2$. It is discrete, but the fundamental domain is the vertical strip $[0,1) \times \mathbb{R}$, which does not have finite volume.

Definition **B.6** ***Homogeneous space***
Given a Lie group G and a smooth manifold M, we say that M is (G)-homogeneous if there is a transitive action of G, by diffeomorphisms on M; that is to say, for every $x, y \in M$, there is a $g \in G$, such that g acting on x produces y $(g(x) = y)$. Given a point $p \in M$, we may then identify M with the quotient manifold G/G_p, where $G_p = \{g \in G \mid g(p) = p\}$ is the stabilizer of p.

The point here is that G_p is a closed Lie subgroup of G, and given any closed subgroup Γ, we obtain a manifold G/Γ. In Ratner's theorems, Γ is discrete (and hence closed) - this is why we call our space $\Gamma\backslash G$ homogeneous!

A space X being homogeneous means that, from a differentiable point of view, neighbourhoods of different points of X look the same.

Remark. There is a more general notion of homogeneous space, where one requires that G be only a topological group.

Example **B.7** ***Homogeneous spaces*** 1. $\mathbb{R}^n$ is a homogeneous space in the canonical sense.

2. Given a Lie group G and a closed Lie subgroup D of G, G/D is a homogeneous space (given by the canonical action of G).
3. The hyperbolic plane is a homogeneous space and can be identified with $\mathrm{PSL}(2,\mathbb{R})/\mathrm{PSO}(2,\mathbb{R})$.
4. The unit tangent bundle of the hyperbolic plane is homogeneous and can be identified with $\mathrm{PSL}(2,\mathbb{R})$.

Actually, being of interest to our discussion, the last two examples will be expounded on below.

Example **B.8** ***The hyperbolic plane***
The group $G = \mathrm{SL}(2,\mathbb{R})$ can be considered a double cover of the unit tangent bundle of the hyperbolic plane. A convenient model of the hyperbolic plane is the upper half-plane $\mathbb{H} = \{x + iy \in \mathbb{C} \mid y > 0\}$ equipped with the hyperbolic metric induced by the Riemannian metric[10] on the tangent bundle[11]

[10] This is a smoothly varying family of inner products on the tangent planes. This just means that you give a notion of length to tangent vectors at any given point, and going to the tangent plane of a nearby point shouldn't change the notion of length too much.

[11] The tangent bundle is the disjoint union of the tangent planes $\mathrm{T_z}\mathbb{H}$ at the point $z \in \mathbb{H}$

$\mathrm{T}\mathbb{H} \cong \mathbb{H} \times \mathbb{C}$

$$\langle u, v\rangle_z = \frac{1}{y^2}(u, v),$$

where $u, v \in \mathrm{T_z}\mathbb{H}$ and (u, v) is the usual inner product on $\mathbb{C}$. This gives us the norm $\|w\|_z$ at the point z ($w \in \mathrm{T_z}\mathbb{H}$). The hyperbolic metric is then given by

$$\mathrm{d}(z_0, z_1) = \inf_c \int_0^1 \|c'(t)\|_{c(t)}\, dt = \int_0^1 \frac{\sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}}{y(t)}\, dt,$$

where the infimum is taken over all possible paths[12]. This metric induces the same topology on $\mathbb{H}$ as the one induced by the Euclidean norm on $\mathbb{C} \supseteq \mathbb{H}$.

The group $\mathrm{PSL}(2, \mathbb{R})$ acts transitively on $\mathbb{H}$ by Möbius transformations, with the stabilizer of i given by $G_i = PSO(2, \mathbb{R})$. Similarly, the action of g by $g(z) = (g(z), g'(z))$, gives us the identification of $\mathrm{T}^1\mathbb{H}$ with $\mathrm{PSL}(2, \mathbb{R})$.

If we identify $\mathrm{T}^1\mathbb{H}$ with $\mathrm{PSL}(2, \mathbb{R})$, the geodesic flow is given by the action of

$$a_t = \begin{bmatrix} e^t & 0 \\ 0 & e^{-1} \end{bmatrix},$$

and the horocycle flow by the action of

$$u_t = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}.$$

For an explanation of what the geodesic flow and the horocycle flow are, see [EW11] or [BM00].

The reason for the above discussion is that Ratner's ideas are generalizations of observations she made of the horocycle flow on $\mathrm{SL}(2, \mathbb{R})$.

Definition **B.9** ***Haar measure***
A (left-invariant) Haar measure is a measure μ on a the Borel σ-algebra of a Lie group G such that:

- $\mu(gE) = \mu(E)$ for any $g \in G$ and any Borel set E,
- $\mu(K)$ is finite for every compact set K,
- $\mu(E) = \inf\{\mu(U) : E \subseteq U, U \text{open}\}$.
- $\mu(E) = \sup\{\mu(K) : K \subseteq E, K \text{compact}\}$.

Remark. There is a more general notion of Haar measure, where one requires that G be only a topological group.

Proposition **B.10** ***Existence and uniqueness of Haar measure***
Every Lie group G has a unique (up to constant) Haar measure ν. That is, if $\tilde{\nu}$ is another Haar measure on G, then there is a constant C such that

$$\nu(E) = C\tilde{\nu}(E)$$

for every measurable set E.

Definition **B.11** ***Linear Lie group***
A (real) Linear Lie group G is a subgroup of the matrix group $GL(n, \mathbb{R})$ (the set of invertible matrices with n^2 entries from $\mathbb{R}$).

[12] a path c in $\mathbb{H}$ is a continuous piecewise differentiable curve $c : [0, 1] \to \mathbb{H}$, $c(t) = x(t) + iy(t)$ between the points z_0 and z_1

Definition **B.12** ***Lie Algebra of a Lie Group***
The Lie algebra $\mathrm{Lie}(G)$ of a Lie group G is the vector space of left-invariant vector fields with the Lie bracket of vector fields. There is a canonical identification between the tangent space at the identity and the Lie algebra.

Definition **B.13** ***The exponential map***
There is a unique map $\exp : \mathrm{Lie}(G) \to G$ such that $\exp(tX) = \phi^X(t)$, where ϕ^X is the left-invariant flow generated by X (the derivative of ϕ^X at the point 1 of G is X), and $\exp(tX)\exp(sX) = \exp((s+t)X)$, for $t \in \mathbb{R}$. This map is called the exponential map.

The exponential map between $\mathrm{Lie}(G)$ and G establishes a (local) diffeomorphic correspondence between a neighbourhood of 0 in $\mathrm{Lie}(G)$ and a neighbourhood of 1 in G, by assigning to an element $X \in \mathrm{Lie}(G)$ the element $\exp(X) = \phi^X(1) \in G$.

The exponential map actually assigns to every $X \in \mathrm{Lie}(G)$ a smooth one-parameter subgroup (a smooth flow) of G.

For linear Lie groups, the exponential map is given by taking the familiar expression (the matrix exponent):

$$\exp(X) = 1 + \frac{X}{1!} + \frac{X^2}{2!} + \frac{X^3}{3!} + \dots + \frac{X^n}{n!} + \dots$$

The above sum converges to an element in G for every $X \in \mathrm{Lie}(G)$.

Definition **B.14** ***The adjoint map***
Given a Lie group G and its associated Lie algebra $\mathrm{Lie}(G)$, we define, for any $g \in G$, the map $\mathrm{Ad}_g : \mathrm{Lie}(G) \to \mathrm{Lie}(G)$ by

$$\mathrm{Ad}_g(v) = \frac{\mathrm{d}}{\mathrm{d}t}\big(g^{-1}(\exp tv)g\big)\big|_{t=0}$$

Proposition **B.15**
The adjoint map Ad_g is a Lie algebra homomorphism for every $g \in G$.

Definition **B.16** ***Unipotent subgroup***
A Lie subgroup U of a Lie group G is said to be unipotent if for every $u \in U$, Ad_u is a unipotent automorphism of the Lie algebra of G, that is $(\mathrm{Ad}_u - \mathrm{id})^n = 0$ for some $n > 0$.

Remark. For a Linear Lie group G, being unipotent means that every matrix in G has all its eigenvalues equal to 1.

Example **B.17**
Upper-triangular matrices with ones (1) on the diagonal are unipotent. Unipotent elements in a Lie group G are generated by nilpotent elements in $\mathrm{Lie}(G)$.

Examples include:

$$\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}, t \in \mathbb{R},$$

and

$$\begin{bmatrix} 1 & s & t \\ 0 & 1 & u \\ 0 & 0 & 1 \end{bmatrix}, s, t, u \in \mathbb{R}.$$

List of Notation

References

[AB06] Charalambos Aliprantis and Kim Border. Springer-Verlag, 2006.

[BM00] M. Bachir Bekka and Matthias Mayer. *Ergodic Theory and Topological Dymanics of Group Actions on Homogeneous Spaces*. Cambridge University Press, 2000.

[EW11] Manfred Einsiedler and Thomas Ward. *Ergodic Theory with a view towards Number Theory*. Springer-Verlag, 2011.

[Fri82] Avner Friedman. *Foundations of Modern Analysis*. Dover, 1982.

[Mor05] Dave Witte Morris. Ratner's theorems on unipotent flows. The Online Final Version 1.1 (February 11, 2005), 2 2005.

[Rat90a] Marina Ratner. On measure rigidity of unipotent subgroups of semisimple groups. *Acta Math.*, 165:229–309, 1990.

[Rat90b] Marina Ratner. Strict measure rigidity for unipotent subgroups of solvable groups. *Invent. Math.*, 101:449–482, 1990.

[Rat91] Marina Ratner. On Raghunathan's measure conjecture. *The Annals of Mathematics*, 134:545–607, 1991.

[Rat92] Marina Ratner. Raghunathan's conjectures for sl(2,r). *Israel Journal of Mathematics*, 80:1–31, 1992.

[Ros02] Wulf Rossmann. *Lie Groups - An Introduction Through Linear Groups*. Oxford University Press, 2002.

[Rud] Walter Rudin. *Real and Complex analysis*. McGraw-Hill, 3 edition.

[Sta00] N. Starkov. *Introduction to Dynamical Systems*. Cambridge University Press, 2000.

[SW73] Arthur A. Sagle and Ralph E. Walde. *Introduction to Lie Groups and Lie Algebras*. Academic Press, 1973.

[Taoa] Terry Tao. http://terrytao.wordpress.com/2009/01/13/distinguished-lecture-series-i-gregory-margulis-homogeneous-dynamics-and-number-theory-i/#more-1464.

[Taob] Terry Tao. http://terrytao.wordpress.com/2007/09/29/ratners-theorems/.

[War83] Frank W. Warner. *Foundations of Differentiable Manifolds and Lie Groups*. Springer-Verlag, 1983.

Index