Contents

Acknowledgment

I am grateful to Jinpeng An, Lihuang Ding, Jianru Duan, Yuxiang Jiao, Yuming Wei, Chengyang Wu, Hongrui Yuan for corrections, comments.

CHAPTER 0

Overview

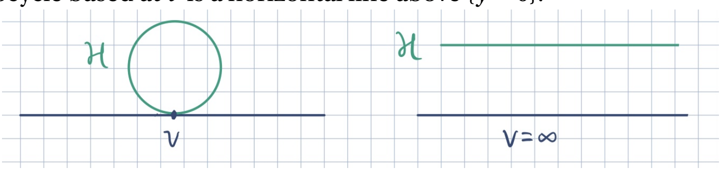
Back to the Top.

1. Motivations/Applications

A few sample results that turn out to be related to homogeneous dynamics are listed.

1.1. Horocycles on constant negative curvature surfaces. Equip $\mathbb{H}^2 := \{x + iy \in \mathbb{C}, y > 0\}$ with the metric $\frac{\mathrm{d}x^2 + \mathrm{d}y^2}{y^2}$. Let $\Gamma \leq \mathrm{Isom}(\mathbb{H}^2)$ be a discrete (torsion free) subgroup such that $\mathbb{H}^2/\Gamma$ is compact (such a subgroup is called a uniform lattice). Then $\mathbb{H}^2/\Gamma$ is a compact surface of constant negative curvature. Conversely, every surface with constant negative curvature arises this way. Let $\pi : \mathbb{H}^2 \to \mathbb{H}^2/\Gamma = M$ be the quotient map.

Consider a horocycle $\mathscr{H}$ in $\mathbb{H}^2$. Explicitly, for each $v \in \{x + iy, y = 0\}$, a horocycle based at v is a circle (with respect to the Euclidean metric) in $\mathbb{H}^2$ tangent to $\{y = 0\}$ at v. For $v = \infty$, a horocycle based at v is a horizontal line above $\{y = 0\}$.



Now we take the image of $\mathscr{H}$ under the projection π.

THEOREM 1.1 ([**Hed36**]). *For every $\mathscr{H}$, $\pi(\mathscr{H})$ is dense in M.*

If $M = \mathbb{H}^2/\Gamma$ ($\Gamma \leq \mathrm{Isom}(\mathbb{H}^2)$ still discrete) is just of finite volume, then

THEOREM 1.2.
1. *$\pi(\mathscr{H})$ is either closed or dense in M.*
2. *Let $\pi(\mathscr{H}_i)$ be a sequence of closed horocycles, then as the length goes to infinity, $\pi(\mathscr{H}_i)$ becomes dense in M.*

REMARK 1.3. *Assume $M = \mathbb{H}^2/\Gamma$ has finite volume. Then there exists closed $\pi(\mathscr{H})$ iff M is non-compact.*

By comparison, the image under π of a geodesic is very different. The image could be closed, dense, or in between. And closed geodesics do not necessarily equidistribute towards the volume measure (though on average they do equidistribute).

1.2. Isometric immersion of hyperbolic spaces. Let $\mathbb{H}^3$ be the three dimensional hyperbolic space $\{(x+iy, z) \in \mathbb{C}\times\mathbb{R}, z > 0\}$ equipped with the metric $\frac{1}{z^2}(\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2)$. Let $\Gamma \leq \mathbb{H}^3$ be a discrete (torsion free) subgroup, such that $\mathbb{H}^3$ is compact (finite volume suffices). Consider an isometric embedding $\iota : \mathbb{H}^2 \to \mathbb{H}^3$. The image of ι can be explicitly described. There are two cases:

1. given a circle on $\{z = 0\}$, then there exists a unique half-sphere in $\mathbb{H}^3$ whose boundary is this given circle;

2. given a line on $\{z = 0\}$, then there exists a unique half-plane in $\mathbb{H}^3$ whose boundary is this given line.

Then $\iota(\mathbb{H}^2)$ is either a half-sphere or a half-plane described above. Similarly, we consider the image of $\iota(\mathbb{H}^2)$ under $\pi : \mathbb{H}^3 \to \mathbb{H}^3/\Gamma =: M$,

THEOREM 1.4. 1. *$\pi(\iota(\mathbb{H}^2))$ is either closed or dense in M;*
2. *Given an infinite sequence of distinct closed $\pi(\iota_i(\mathbb{H}^2))$, then $\lim_i \pi(\iota_i(\mathbb{H}^2))$ is dense in M.*

REMARK 1.5. *That the volume of $\pi(\iota_i(\mathbb{H}^2))$ would go to infinity is automatic.*

1.3. Oppenheim conjecture/Margulis theorem. Consider a non-degenerate real quadratic form in three (larger than 3 also ok) variables, viewed as a function $Q : \mathbb{R}^3 \to \mathbb{R}$. Assume it is indefinite. Note that if Q is a quadratic form with rational coefficients or proportional to such a form, then $Q(\mathbb{Z}^3)$ is discrete in $\mathbb{R}$.

THEOREM 1.6. *If Q is NOT proportional to a quadratic form with rational coefficients, then $Q(\mathbb{Z}^3)$ is dense in $\mathbb{R}$.*

REMARK 1.7. *It is also true replacing $\mathbb{Z}^3$ by primitive vectors. The proof is a bit harder.*

REMARK 1.8. *This is false if Q has two variables.*

The above theorem admits a quantitative version in certain cases. Let Q be a (nondegenerate) quadratic form in 4 variables of signature $(3,1)$ (what follows does not apply to signature $(2,2)$, $(1,2)$). Assume Q is irrational as above.

THEOREM 1.9. *There exists $\lambda_Q > 0$ such that for every $a < b \in \mathbb{R}$,*

$$\#\left\{x \in \mathbb{Z}^4 \,\middle|\, Q(x) \in (a,b),\ \|x\| \le T\right\} \sim \mathrm{Vol}\left\{x \in \mathbb{R}^4 \,\middle|\, Q(x) \in (a,b),\ \|x\| \le T\right\} \sim \lambda_Q (b-a) T^2$$

1.4. Littlewood conjecture. Let $\alpha \in \mathbb{R}$ (assume everything is irrational just in case of some trivialities). By pigeon-hole principle(?), one can show that

$$\inf_{(m\neq 0,n)\in\mathbb{Z}^2} |m|\cdot|m\alpha+n| \le 1.$$

On the other hand there exists α ("badly approximable numbers") such that

$$\inf_{(m\neq 0,n)\in\mathbb{Z}^2} |m|\cdot|m\alpha+n| > 0.$$

The Littlewood conjecture is

CONJECTURE 1.10. *For every pair $(\alpha,\beta) \in \mathbb{R}^2$ irrational,*

$$\inf_{(m\neq 0,n_1,n_2)\in\mathbb{Z}^3} |m|\cdot|m\alpha+n_1|\cdot|m\beta+n_2| = 0.$$

To make it look closer to the Oppenheim conjecture, you may write $l_\alpha(x,y,z) := \alpha x + y$, $l_\beta(x,y) := \beta x + z$ and $\varphi(x,y,z) := x$. Let $L(x,y,z) := \varphi\cdot l_\alpha \cdot l_\beta$. Then the conjecture asserts that when L is “irrational”, then $\inf_{(x,y,z)\in\mathbb{Z}^3} |L(x,y,z)|$ is dense at 0. By comparison, the Oppenheim conjecture is equivalent to $Q(\mathbb{Z}^3)$ being dense at 0.

Our current knowledge is

THEOREM 1.11. *The set*

$$\left\{(\alpha,\beta)\in\mathbb{R}^2,\ \textit{that fails this conjecture}\right\}$$

has Hausdorff dimension 0.

REMARK 1.12. *For every $\delta > 0$, there exists (α, β) such that*

$$\liminf_{n\in\mathbb{Z}} n^{1+\delta} \|n\alpha\| \|n\beta\| > 0.$$

According to [Gal62], this was done in [Spe42]. So the exponent on n is the best one can hope for. On the other hand, maybe one can improve n by $\log n$ (see [Gal62] for some restrictions though).

1.5. Quantum unique ergodicity. Let (M, d) be a closed hyperbolic surface of constant negative curvature. Let Δ be the Laplacian operator $-y^2(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$.

Fact 1. Eigenvalues of Δ are non-negative and discrete in $\mathbb{R}$, say enumerated as

$$0 = \lambda_1 < \lambda_2 < \ldots$$

Fact 2. For each λ_i, the eigenspace E_{λ_i} consists of smooth functions and has finite dimension;

Fact 3. Different eigenspaces are mutually orthogonal and $L^2(M)$ is spanned by them.

(see e.g. Thm 3.2.1, Jost, Riemannian Geometry; Thm 4.43, Gallot, Hulin, Lafontaine.)

Now take $f_i \in E_{\lambda_i}$. We are interested in the limiting behavior of the sequence of measures $\{|f_i|^2 \mathrm{Vol}\}$, normalized to be probability measures.

A theorem (quantum ergodicity) of Snirelman says that there exists a density one subsequence n_i such that $\lim |f_{n_i}|^2 \mathrm{Vol} = \mathrm{Vol}$ (suitably normalized) in the weak* topology (this theorem holds for more general compact Riemannian manifold, as long as the geodesic flow is ergodic, a property that holds for every negatively curved surface).

CONJECTURE 1.13 (Quantum unique ergodicity). $\lim |f_n|^2 \mathrm{Vol} = \mathrm{Vol}$ *holds without passing to any subsequence.*

This is still open. Progress is made when the fundamental group is a "congruence subgroup" where there is an additional supply of operators, called Hecke operators, that commute with the Laplacian.

THEOREM 1.14. *Assume $\{f_i\}$ is a sequence of Hecke-Laplacian eigenfunctions. Then*

$$\lim |f_n|^2 \mathrm{Vol} = \mathrm{Vol}$$

in the weak topology.*

In the non-compact congruence case, this also holds for Hecke-Maass forms whose proof requires one more step to guarantee non-divergence.

2. Measure rigidity

2.1. Unipotent flows. Consider $\mathrm{SL}_2(\mathbb{R})$ and a discrete subgroup Γ. Equip $\mathrm{SL}_2(\mathbb{R})$ with a right invariant Riemannian metric. Then the volume measure m_X on $\mathrm{SL}_2(\mathbb{R})/\Gamma$ is left invariant under $\mathrm{SL}_2(\mathbb{R})$. We normalize it to be a probability measure.

Consider the subgroup

$$\mathrm{U} := \left\{ u_s := \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}, \ s \in \mathbb{R} \right\}$$

THEOREM 2.1. *Assume $\mathrm{SL}_2(\mathbb{R})/\Gamma$ is compact. Then m_X is the unique U-invariant probability measure.*

This would immediately imply the denseness result above.

THEOREM 2.2. *Assume $\mathrm{SL}_2(\mathbb{R})/\Gamma$ has finite volume. Then each U-invariant probability measure is a convex combination (possible in the form of an integral) of the following*

1. m_X;

2. *the* U*-invariant measure supported on a closed (and compact) orbit of* U.

The implication to orbit closure requires an analysis on this convex combination.

In general, Ratner's measure classification on **ergodic** invariant measures for Ad-unipotent flows roughly reads as follows.

THEOREM 2.3 (Measure rigidity theorem). *Assume the following*

- *a connected Lie group (2nd countable)* G *together with a discrete subgroup* Γ*;*
- *a one-parameter* ***Ad-unipotent*** *subgroup* $\mathrm{U} = \{\mathbf{u}_s,\ s \in \mathbb{R}\}$ *of* G.

Then every U*-invariant ergodic probability measure* μ *on* G/Γ *is* ***homogeneous****.*

Ad-unipotent means that the image of U under the Adjoint representation in $GL(\mathfrak{g})$ consist of unipotent matrices.

For a measure μ on G/Γ define the closed subgroup of G by

$$H := G_\mu := \{g \in G,\ g_*\mu = \mu\}.$$

We say that a probability measure μ is homogeneous if there exists $x \in X = G/\Gamma$ such that $\mu(Hx) = 1$.

REMARK 2.4. *When* G *is a semisimple closed subgroup of* SL_n*, Ad-unipotent is the same as being unipotent in* SL_n*.*

REMARK 2.5. *Various "connected" assumptions may be dropped with similar conclusions. E.g. one may consider* $\mathbf{u}_{s\in\mathbb{Z}}$*.*

REMARK 2.6. *Let* H*,* x *be as in the theorem and the definition above. Then* Hx *is closed in* G/Γ*. This is proved in* [Rag72, Sec.1.13] *assuming* G/Γ *admits a finite* G*-invariant measure (i.e.,* Γ *is a lattice in* G*), but the proof carries through without this assumption.*

REMARK 2.7. *Let* H*,* x *be as in the theorem and the definition above. Then by modifying* x*, one can show that* $Hx = H^\circ x$ *and* U *is contained in* H°*.*

THEOREM 2.8 (Equidistribution and topological rigidity I). *Further assume that* Γ *is a lattice in* G*. Then for every* x*, there exists* $\mathrm{U} \le H \le G$ *closed connected subgroup such that*

1. Hx *is closed and supports an* H*-invariant probability measure* μ_H*;*
2. *for every bounded continuous function* $f : G/\Gamma \to \mathbb{R}$*,*

$$\lim_{T\to\infty}\int_0^T f(u_t \cdot x)\,\mathrm{dt} \text{ exists and is equal to } \int f(x)\mu_H(x).$$

3. $\mathrm{U} \curvearrowright \mu_H$ *is ergodic;*
4. $\mathrm{U}\cdot x$ *is dense in* $H \cdot x$*.*

The logic of Ratner is

$$\text{Measure rigidity} \implies \text{Equidistribution} \implies \text{Topological rigidity}.$$

Nevertheless, there is a different (potential) approach by deducing the topological rigidity bypassing ergodic theory.

The topological rigidity is the original Raghunathan's conjecture.

THEOREM 2.9 (Topological rigidity II). *Let* G, Γ *be as in the last theorem. Let* $L \le G$ *be a Lie subgroup generated by one-parameter Ad-unipotent subgroups. Then for every* $x \in G/\Gamma$*, there exists* $L \le H \le G$ *and* $V \le L$ *some one-parameter Ad-unipotent subgroup (of* G*) such that*

1. Hx *is closed and supports an* H*-invariant probability measure;*
2. $\overline{Lx} = \overline{Vx} = Hx$*;*
3. $V \curvearrowright \mu_H$ *is ergodic.*

2.2. Higher rank diagonalizable action. Fact: Let $a_t := \begin{bmatrix} e^t & \\ & e^{-t} \end{bmatrix}$. Then the $\{a_t\}_{t\in\mathbb{R}}$ action on $\mathrm{SL}_2(\mathbb{R})/\Gamma$ admits many invariant probability measures/closed sets and they are not easy to classify. The conjecture is that the situation would become better in higher rank.

Let

$$A := \left\{ \begin{bmatrix} e^{t_1} & & \\ & e^{t_2} & \\ & & e^{t_3} \end{bmatrix} \middle| t_i \in \mathbb{R}, \sum t_i = 0 \right\} \cong \mathbb{R}^2.$$

Consider the $A \curvearrowright \mathrm{SL}_3(\mathbb{R}) / \mathrm{SL}_3(\mathbb{Z})$.

CONJECTURE 2.10.
- *Every ergodic invariant probability measure is homogeneous;*
- *Every bounded (in the unbounded case, statements need to be modified) orbit of A is homogeneous.*

Of course one can propose similar (but necessarily more complicated) conjectures for other (semisimple) Lie groups G and other A's.

THEOREM 2.11. *Let $G := \mathrm{SL}_3(\mathbb{R})$, $\Gamma = \mathrm{SL}_3(\mathbb{Z})$ and A same as above. Let μ be an A-invariant ergodic probability measure on G/Γ. Assume for some $a \in A$, $h_\mu(a) > 0$. Then μ is the G-invariant probability measure on G/Γ.*

The topological implication is that

THEOREM 2.12. *The Hausdorff dimension of*

$$\{x \in G/\Gamma,\ Ax \text{ is bounded}\}$$

is 2.

Note that the union of compact A-orbits is a countable union, hence also has Hausdorff dimension 2.

A theorem of slightly different flavor, related to the AQUE theorem above, is

THEOREM 2.13. *Let $G := \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$, $\Gamma \le G$ irreducible (e.g. $\Gamma = \mathrm{SL}_2(\mathbb{Z}[\sqrt{2}])$). Let $H := \{e\} \times \mathrm{SL}_2(\mathbb{R})$. And*

$$A := \{(a_t, \mathrm{id}),\ t \in \mathbb{R}\}$$

Let μ be an A-invariant probability measure such that
- *$h(a,\nu) > 0$ for every ergodic component ν of μ;*
- *μ is H-recurrent (some assumption weaker than H-invariant),*

then μ is the G-invariant probability measure.

REMARK 2.14. *Same conclusion holds replacing the 2nd factor $\mathrm{SL}_2(\mathbb{R})$ by $\mathrm{SL}_2(\mathbb{Q}_p)$. This p-adic version is what is required for the AQUE theorem.*

REMARK 2.15. *This theorem is not easily reduced to the ergodic case due to the recurrence condition.*

REMARK 2.16. *Whether one can eliminate the entropy assumption remains an open problem.*

3. Further reading

Here are some general references.

[**BM00**] is a nice introduction to homogeneous dynamics including a proof of Oppenheim conjecture in the last chapter.

Einsiedler and Ward have a (ongoing) book project on homogeneous dynamics available on the authors' homepages.

What we plan to cover in this course (and almost everything I write here) can be found in the monograph [EEE+10].

CHAPTER 1

Denseness of horocycles

Back to the Top.

1. Summary

In [**Hed36**], the author considers curves in the hyperbolic disk with constant "geodesic curvature" (let cur denote this number), measuring how far a curve is away from being a geodesic. Under this constancy condition, curves are divided into 4 types:

1. $\mathrm{cur} = 0$;
2. $\mathrm{cur} \in (0,1)$;
3. $\mathrm{cur} = 1$;
4. $\mathrm{cur} > 1$.

Geodesics belong to type 1. Type 2 are those equidistant to some geodesic. Type 3 are horocycles. Type 4 are Euclidean circles in the interior of the disk model. The paper is about their "transitivity modulo Γ", namely their image in certain hyperbolic surfaces (orbifolds).

This chapter is about horocycles, that is, curves of type 3 assuming the relevant surface is compact. Actually our discussion applies to the unit-tangent bundle, not just to the surface itself. Type 4 are compact. It is claimed that type 2 behave like type 1. One can also show that as the curvature tends to 1, type 2 and type 4 asymptotically behave like type 3.

Firstly, let us introduce some notations

- $\mathrm{U} := \left\{ \mathbf{u}_s := \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle| s \in \mathbb{R} \right\}$, $\mathrm{A} := \left\{ \mathbf{a}_t := \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix} \middle| t \in \mathbb{R} \right\}$;
- Let Γ be a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$ and let $\mathrm{X} := \mathrm{SL}_2(\mathbb{R})/\Gamma$, equipped with the quotient topology;
- for $g \in \mathrm{SL}_2(\mathbb{R})$, write $[g]_\Gamma$ for its image in X.

One can identify X (at least when $\{\pm 1\} \subset \Gamma$ and $\Gamma/\pm 1$ is torsion free) with the unit tangent bundle of some hyperbolic surface. Moreover, orbits of A are geodesics and orbits of U are horocycles.

THEOREM 1.1. *Assume in addition that Γ is cocompact in $\mathrm{SL}_2(\mathbb{R})$, namely, X is compact. Then the action of* U *on* X *is* ***minimal****, that is to say, for every $x \in \mathrm{X}$, $\mathrm{U}.x$ is dense in* X.

In a dual formulation, this says that for every nonzero vector $v \in \mathbb{R}^2$, $\Gamma.v$ is dense in $\mathbb{R}^2$. This fails for $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, but $\mathrm{SL}_2(\mathbb{Z})$ is not cocompact.

REMARK 1.2. *The proof below applies equally well to the discrete case $\{\mathbf{u}_s\}_{s\in\mathbb{Z}}$ with the same conclusion. Namely, for every $x \in \mathrm{X}$, $\{\mathbf{u}_s.x\}_{s\in\mathbb{Z}, s\geq 0}$ is dense in* X. *However, whether $\{u_{s^2}.x\}_{s\in\mathbb{Z}}$ is dense in* X *seems unknown (see* [**Zhe21**] *and references therein).*

This chapter roughly corresponds to [**BM00**, Chapter IV, Section 2].

2. Injectivity radius

We fix some right invariant metric $d(\cdot,\cdot)$ on $\mathrm{SL}_2(\mathbb{R})$, compatible with the topology. We will not be bothered about the explicit form of the metric. So just take its existence as a fact. Assuming this, define the quotient metric on $\mathrm{SL}_2(\mathbb{R})/\Gamma$ by

$$d([g]_\Gamma,[h]_\Gamma) := \inf_{\gamma\in\Gamma} d(g\gamma,h) = \inf_{\gamma_1,\gamma_2\in\Gamma} d(g\gamma_1,h\gamma_2).$$

Fix such a metric, we can define the injectivity radius at a point $x \in \mathrm{X}$ by

$$\mathrm{InjRad}(x) := \inf\left\{\delta>0 \,\middle|\, g\mapsto g.x \text{ is injective on } d(g,\mathrm{id})<\delta\right\}.$$

Since Γ is discrete, $\mathrm{InjRad}(x)>0$ for all $x\in\mathrm{X}$. Also note that InjRad is continuous. Hence if Γ is cocompact, there exists (and we fix such an) $r_{\mathrm{X}}>0$ such that $\mathrm{InjRad}(x)\geq r_{\mathrm{X}}$ for all $x\in\mathrm{X}$.

Also, one can check that for $d(g_i,\mathrm{id})<\frac{r_x}{4}$ for $i=1,2$, we have $d(g_1.x,g_2.x)=d(g_1,g_2)$.

LEMMA 2.1. *Assume* Γ *is cocompact. Then there are no compact* U*-orbits in* X*, that is, for every* $s\neq 0$ *and* $x\in\mathrm{X}$, $\mathbf{u}_s.x\neq x$. *As every unipotent matrix in* $\mathrm{SL}_2(\mathbb{R})$ *is conjugate to an element of* U*, this implies that* Γ *contains no (non-identity) unipotent matrices.*

PROOF. Assume otherwise, then we can find $g_0\in\mathrm{SL}_2(\mathbb{R})$ such that

$$s_0 := \inf\left\{s>0 \,\middle|\, \mathbf{u}_s.g_0\Gamma = g_0\Gamma\right\} > 0.$$

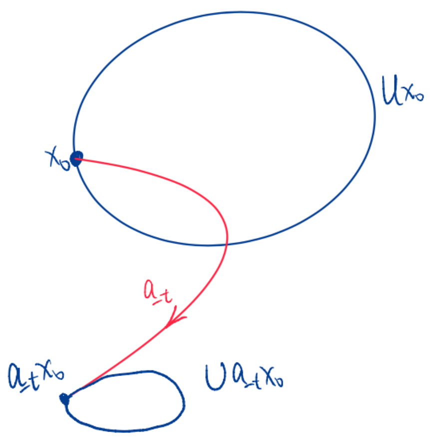
In the current case inf is actually achieved at s_0 and $\mathbf{u}_{s_0}.g_0\Gamma=g_0\Gamma$. Consider

$$\begin{aligned} &\mathbf{a}_{-t}\mathbf{u}_{s_0}g_0\Gamma = \mathbf{a}_{-t}g_0\Gamma \\ \Longrightarrow &\begin{bmatrix} 1 & e^{-2t}s_0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} e^{-t} & 0 \\ 0 & e^{t} \end{bmatrix} g_0\Gamma = \begin{bmatrix} e^{-t} & 0 \\ 0 & e^{t} \end{bmatrix} g_0\Gamma. \end{aligned}$$

As $t\to+\infty$, this implies the existence of compact orbit of U of arbitrarily small period, which is impossible due to the fact $r_{\mathrm{X}}>0$. More explicitly, for t large enough such that

$$d\left(\mathrm{id},\begin{bmatrix} 1 & e^{-2t}s_0 \\ 0 & 1 \end{bmatrix}\right) < r_{\mathrm{X}},$$

One has, by the definition of r_{X}, that $\mathbf{u}_{e^{-2t}s_0}=\mathrm{id}$, or in other words, $s_0=0$. Here is a picture



Existence of closed horocycles

$\Rightarrow$ Existence of short closed horocycles

□

COROLLARY 2.2. *Assume* Γ *is cocompact and take* $x \in \mathrm{X}$. *There exist* $t_n, s_n \to +\infty$ *with* $|t_n - s_n| \to \infty$ *such that* $d(x_n, y_n) \to 0$, *where* $x_n := \mathbf{u}_{t_n}.x$ *and* $y_n := \mathbf{u}_{s_n}.x$

PROOF. The map $t \mapsto \mathbf{u}_t.x$ from $\mathbb{R}_{\geq 0}$ to X is injective. Since X is compact, we may apply pigeon-hole principle. □

3. Additional invariance

Now we start to prove the theorem. The crucial notion here is

DEFINITION 3.1. *Let a (semi)group G act on a topological space W by homeomorphisms. A nonempty subset V of W is said to be G-**minimal** iff it is closed, G-stable and contains no proper non-empty closed G-stable subset.*

Let Y be a U-minimal set in the orbit closure $\overline{\mathrm{U}.x_0}$. The existence of Y is guaranteed by the compactness of X and Zorn's lemma. Our strategy is to find some $y \in Y$ and a larger group whose orbit based at y is contained in Y.

PROOF OF THEOREM 1.4. Keep notations as Coro.2.2 above. When n is large, we find some $A_n \in \mathrm{SL}_2(\mathbb{R})$ with $d(A_n, \mathrm{id}) \leq r_{\mathrm{X}}/4$ such that $y_n = A_n x_n$. Write

$$A_n = \begin{bmatrix} 1+a_n & b_n \\ c_n & 1+d_n \end{bmatrix} \quad \text{with } a_n, b_n, c_n, d_n \to 0.$$

Coro.2.2 ensures that A_n is not an upper triangular unipotent matrix.

The key calculation is:

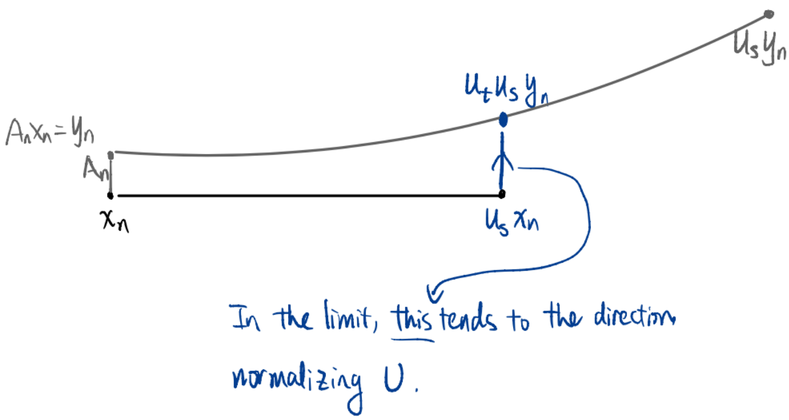
$$\begin{aligned} \mathbf{u}_s A_n \mathbf{u}_s^{-1} &= \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1+a_n & b_n \\ c_n & 1+d_n \end{bmatrix} \begin{bmatrix} 1 & -s \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1+a_n & b_n - s(1+a_n) \\ c_n & 1+d_n - sc_n \end{bmatrix} \\ &= \begin{bmatrix} 1+a_n+sc_n & b_n + s(d_n - a_n) - s^2 c_n \\ c_n & 1+d_n - sc_n \end{bmatrix}. \end{aligned} \tag{1}$$

Case I, $c_n = 0$ for infinitely many n.

This case is left to you as an exercise.

Case II. $c_n \neq 0$ for n large enough.

Equa.(1) above suggests that the upper right corner dominates when s is large (this is called "shearing phenomenon", we will return to this point later).



We will kill the upper right corner according to the following computation

$$\mathbf{u}_t(\mathbf{u}_s A_n \mathbf{u}_s^{-1}) = \begin{bmatrix} 1+a_n+(s+t)c_n & b_n + s(d_n-a_n) - s^2 c_n + t(1+d_n - sc_n) \\ c_n & 1+d_n - sc_n \end{bmatrix}. \tag{2}$$

Define $t = t(s)$ by imposing the following equality

$$\begin{aligned} & b_n + s(d_n-a_n) - s^2 c_n + t(1+d_n-sc_n) = 0 \\ \iff & t = -\frac{b_n + s(d_n-a_n) - s^2 c_n}{d_n - sc_n} = -\frac{b_n - sa_n - s}{1+d_n-sc_n} - s \end{aligned} \tag{3}$$

The range of s for which the $t(s)$ is ill-defined will be excluded from the discussion (see $s = s_{n,\delta}$ below, where one has $1+d_n-sc_n = 1\pm\delta$ with δ small). With this choice of $t = t(s)$,

$$\mathbf{u}_t(\mathbf{u}_s A_n \mathbf{u}_s^{-1}) = \begin{bmatrix} (1+d_n-sc_n)^{-1} & 0 \\ c_n & 1+d_n-sc_n \end{bmatrix}. \tag{4}$$

Now for $\delta > 0$ (we will let $\delta \to 0$ in a moment), choose $s = s_{n,\delta} \geq 0$ such that either $1+d_n-sc_n = 1+\delta$ or $1-\delta$, depending on the signature of c_n. So

$$s_{n,\delta} = \frac{d_n - \delta}{c_n} \text{ or } \frac{d_n+\delta}{c_n},$$

whichever is positive.

Define

$$y'_{n,\delta} := \mathbf{u}_{t(s)}\mathbf{u}_s.y_n, \quad x'_{n,\delta} := \mathbf{u}_s.x_n, \quad \text{where } s = s_{n,\delta}.$$

Then by definition

$$\begin{aligned} y'_{n,\delta} &= \mathbf{u}_{t(s)}\mathbf{u}_s A_n \mathbf{u}_s^{-1}\mathbf{u}_s.x_n = \mathbf{u}_{t(s)}\mathbf{u}_s A_n \mathbf{u}_s^{-1}.x'_{n,\delta} \\ &= \begin{bmatrix} (1\pm\delta)^{-1} & 0 \\ 0 & (1\pm\delta) \end{bmatrix}.x'_{n,\delta} \end{aligned} \tag{5}$$

Fix δ, let n vary. By passing to a subsequence n_k, assume that $y'_{n,\delta}$ and $x'_{n,\delta}$ converge to $y_{\infty,\delta}$ and $x_{\infty,\delta}$ respectively. Hence

$$y'_{\infty,\delta} = \begin{bmatrix} (1+\delta)^{-1} & 0 \\ 0 & (1+\delta) \end{bmatrix} \cdot x'_{\infty,\delta} \text{ or } \begin{bmatrix} (1-\delta)^{-1} & 0 \\ 0 & (1-\delta) \end{bmatrix} \cdot x'_{\infty,\delta}.$$

Without loss of generality, assume that the first case happens for infinitely many $\delta > 0$ converging to 0. It looks like we are not making any progress except that the "transverse difference" is now in the direction of the diagonal, which normalizes U. So it is time to invoke the following general fact, which is why we introduced the notion of minimal set.

LEMMA 3.2. *Let* $\Gamma \curvearrowright Z$ *by homeomorphisms.* Γ *is a semi-group and* Z *a topological space. Assume that* V *is a* Γ*-minimal set and* W *is a* Γ*-invariant closed set. If* $\phi \in Homeo(X)$ *normalizes (the image of)* Γ *and there exist* $v_0 \in V$ *and* $w_0 \in W$ *with* $\phi(v_0) = w_0$*. Then* $\phi(V)$ *is contained in* W*.*

PROOF OF THE LEMMA.

$$\phi(V) = \phi(\overline{\Gamma.v_0}) = \overline{\phi(\Gamma.v_0)} = \overline{\Gamma.w_0} \subset W.$$

$\square$

From the lemma (applied to $V = W = Y$, $Z =$ X), we see that for a set of δ converging to 0 and for every $y \in Y$,

$$\begin{bmatrix} (1+\delta)^{-1} & 0 \\ 0 & (1+\delta) \end{bmatrix}.Y \subset Y.$$

Since $\{g \in \mathrm{SL}_2(\mathbb{R}) \mid gY \subset Y\}$ is a closed sub-semigroup,

$$\{\mathbf{a}_t \mid t \geq 0\}.Y \subset Y.$$

is contained in Y for every y. By definition $\{\mathbf{u}_s \mid s \geq 0\}.Y \subset Y$. By minimality, actually $\{\mathbf{u}_s \mid s \geq 0\}.Y = Y$.

Fix some $y \in Y$, and take a limit point y' of $\mathbf{a}_t y$ as $t \to +\infty$. Then the orbit of the full group $\mathrm{A}.y' \subset Y \implies \mathrm{AU}.y' \subset Y$. Thus we are done modulo Lem.4.1.

□

4. A duality argument

Let $B^+ := \{\mathbf{a}_t\mathbf{u}_s\}_{t,s\in\mathbb{R}} = \mathrm{A}\cdot\mathrm{U}$ and $B := \{(\pm 1)\mathbf{a}_t\mathbf{u}_s\}_{t,s\in\mathbb{R}}$. B^+ is the identity component of B and $B = B^+ \sqcup (-1)B^+$ where we have abbreviated the matrix $\begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$ as " -1".

LEMMA 4.1. *The action of B^+ on X is minimal.*

REMARK 4.2. *The lemma also holds when only assuming Γ to be discrete and of finite covolume (referred to as a **lattice**). Actually, the lemma follows iff the limit set for Γ is the full boundary, which includes some infinitely generated, infinite covolume examples.*

PROOF. We are going to show that the B-action is minimal first and then explain why this is sufficient.

An equivalent formulation is that the Γ-action on $\mathrm{SL}_2(\mathbb{R})/B$ is minimal. To prove this, we will take a geometric point of view.

Recall that $\mathrm{SL}_2(\mathbb{R})$ acts on the upper half space $\mathscr{H}^2 := \{z = x + iy \mid x \in \mathbb{R},\, y > 0\}$ by

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot z := \frac{az+b}{cz+d}.$$

This action preserves the Riemannian metric (referred to as the hyperbolic metric)

$$(\mathrm{dx}^2 + \mathrm{dy}^2)/y^2.$$

Geodesics under the hyperbolic metrics are (Euclidean) circles perpendicular to the x-axis together with all the vertical lines.

Another important point is that as the y-coordinate approaches 0, the (hyperbolic) distance between two points of (Euclidean) distance $\asymp 1$ actually goes to ∞. The $\mathrm{SL}_2(\mathbb{R})$-action extends continuously to the "boundary" defined by

$$\partial\mathscr{H}^2 := \{(x,y),\, y = 0\} \sqcup \{\infty\}.$$

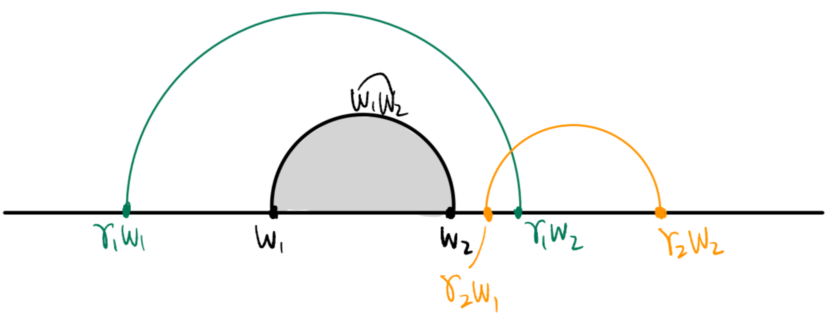
where the topology near ∞ is defined as the "one-point compactification". Thus topologically the boundary is a circle. The action at ∞ is given as follows

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \infty = \frac{a\infty + b}{c\infty + d} = \begin{cases} a/c, & \text{if } c \neq 0 \\ \infty, & \text{otherwise} \end{cases}.$$

Why care? Note that the stabilizer of ∞ is exactly B and the action is transitive on $\partial\mathscr{H}^2$ (Exercise: convince yourself that this gives a topological homeomorphism $\mathrm{SL}_2(\mathbb{R})/B \cong \mathscr{H}^2$) Thus it suffices to show that the action of Γ on $\partial\mathscr{H}^2$ is minimal.

CLAIM 4.1. *For every $z \in \mathscr{H}^2$, the orbit closure $\overline{\Gamma\cdot z} \supset \partial\mathscr{H}^2$.*

Assuming the claim, let W be a closed Γ-invariant set on $\partial\mathscr{H}^2 \cong S^1$. Thus its complement consists of disjoint union of open intervals (labelled as I_i's). Take such an interval I_0 with endpoints w_1, w_2. We argue that $\Gamma \cdot \widehat{w_1 w_2}$ (the unique geodesic connecting w_1 and w_2) never contains I_0 in its closure, which contradicts against the above claim. Indeed, Γ translates of $\widehat{w_1 w_2}$ are just geodesics with endpoints outside the region between $\widehat{w_1 w_2}$ and I_0.



Γ-translates of $\widehat{W_1 W_2}$ never enters the grey region.

Hence we are done.

Proof of Claim 4.1. By co-compactness, we can find a bounded region $\mathscr{B} \subset \mathscr{H}^2$ (whose diameter under hyperbolic distance is denoted by $\mathrm{diam}(\mathscr{B})$) such that $\Gamma.\mathscr{B} = \mathscr{H}^2$. For every $z_{\neq\infty} \in \partial\mathscr{H}^2$ (the case $z = \infty$ is left to you) and a neighborhood $\mathscr{N}_{z,r_0}$ of radius r_0 (in the Euclidean metric) of z, we are going to show that some $\gamma\cdot\mathscr{B}$ is contained in $\mathscr{N}_{z,r_0}$. Indeed we can find $\gamma \in \Gamma$, $b \in \mathscr{B}$ such that $\gamma b.z \in \mathscr{N}_{z,r_0/2}$. When r_0 is sufficiently small one can show that for $z' \in \mathscr{H}^2$

$$d_{\mathrm{Hyperbolic}}(z', \gamma b.z) \leq \mathrm{diam}(\mathscr{B}) \implies d_{\mathrm{Euclidean}}(z', \gamma b.z) \leq r_0/2.$$

Applying this to $z' = \gamma.z$ finishes the proof. □

Finally, as promised, we explain how to get the minimality of B^+-action from that of B. So take $x_0 \in \mathrm{X}$ and we know that

$$\overline{B.x_0} = \overline{B^+.x_0 \cup B^+(-1).x_0} = \mathrm{X}.$$

As $\overline{B^+x_0} \cup \overline{B^+(-1)x_0}$ is B-invariant and closed, it is also equal to X. As X is connected (well, the group $\mathrm{SL}_2(\mathbb{R})$ is connected), their intersection $\overline{B^+x_0} \cap \overline{B^+(-1)x_0}$ is non-empty. But this again, is a B-invariant closed set, so has to be the full X. In particular $\overline{B^+x_0} = X$. And the proof completes. □

5. Exercises

Exercise 5.1. *Let $G := \mathrm{SL}_2(\mathbb{R})$ act continuously on a locally compact Hausdorff topological space X. Take $x \in X$, let $G_x := \{g \in g,\ g.x = x\}$. Assume $G.x$ is closed, or is open in its closure, show that the bijection induced from $g \mapsto g.x$*

$$G/G_x \to G.x$$

is a homeomorphism where G/G_x is equipped with the quotient topology and $G.x$ is equipped with the subspace topology.

Hint: Baire's Category theorem.

Exercise 5.2. *Assume Γ is cocompact in $\mathrm{SL}_2(\mathbb{R})$. Consider the standard action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{R}^2$. Show that for every $v_{\neq 0} \in \mathbb{R}^2$, $\Gamma.v$ is dense in $\mathbb{R}^2$.*

Exercise 5.3. *Show that $\mathrm{SL}_2(\mathbb{R})$ does not admit any bi-invariant Riemannian metric.*

5.1. A more geometric take on horocycles. I assume you have some familiarity with geometry on the upper half space in this section.

Notations:

- $\mathbb{H}^2 := \{(x, y) \in \mathbb{R}^2, y > 0\}$ equipped with the metric $\frac{\mathrm{dx}^2+\mathrm{dy}^2}{y^2}$ and the left action of $\mathrm{SL}_2(\mathbb{R})$ via fractional linear transformations;
- $T^1(\mathbb{H}^2)$ is the unit tangent bundle of $\mathbb{H}^2$;
- $\partial\mathbb{H}^2 := \{(x,0) \in \mathbb{R}^2, x \in \mathbb{R}\} \sqcup \{\infty\}$ be the boundary of $\mathbb{H}^2$; The topology on $\{(x, y), x \in \mathbb{R}, y \geq 0\}$ is the natural topology and the topology on $\overline{\mathbb{H}^2} := \mathbb{H}^2 \sqcup \partial\mathbb{H}^2$ is the one-point compactification topology. The action of $\mathrm{SL}_2(\mathbb{R})$ extends continuously to $\overline{\mathbb{H}^2}$;
- Let Γ_0 be a discrete subgroup of $\mathrm{SL}_2(\mathbb{Z})$ such that $\Gamma_0\backslash\mathbb{H}^2$ is a closed surface of genus $g \geq 2$;
- Let $\Gamma_0' := [\Gamma_0, \Gamma_0]$, recall that Γ_0' is a normal subgroup of Γ_0 and $\Gamma_0/\Gamma_0' \cong \mathbb{Z}^{2g}$;
- For $x \in \mathbb{H}^2$ and a discrete subgroup Γ of $\mathrm{SL}_2(\mathbb{R})$, define the limit set $\mathrm{Limit}_x(\Gamma) := \overline{\Gamma.x} \setminus \Gamma.x$ in $\overline{\mathbb{H}^2}$.

EXERCISE 5.4. $\mathrm{Limit}_x(\Gamma) \subset \partial\mathbb{H}^2$ *for every discrete subgroup* Γ *of* $\mathrm{SL}_2(\mathbb{R})$ *and every* $x \in \mathbb{H}^2$.

EXERCISE 5.5. *For every* $x, y \in \mathbb{H}^2$ *and discrete subgroup* Γ *of* $\mathrm{SL}_2(\mathbb{R})$, $\mathrm{Limit}_x(\Gamma) = \mathrm{Limit}_y(\Gamma)$.

Thus the limit set is independent of the choice of base point and we henceforth denote it by $\mathrm{Limit}(\Gamma)$.

EXERCISE 5.6. *Let* Γ *be a discrete subgroup of* $\mathrm{SL}_2(\mathbb{R})$. *Show that* $\mathrm{Limit}(\Gamma)$ *is a* Γ*-minimal set.*

(A Γ-set is said to be Γ-minimal iff either it is empty or for every x in this set, $\Gamma.x$ is dense in this set. Actually $\mathrm{Limit}(\Gamma)$, if infinite, is the unique nonempty Γ-minimal set)

Recall that for every geodesic Y (or closed convex subset) on $\mathbb{H}^2$ and every $x \in \mathbb{H}^2$, there is a unique point, denoted as $\pi_Y(x)$, in Y such that

$$\mathrm{dist}(x, Y) = \mathrm{dist}(x, \pi_Y(x)).$$

For every $x \in T^1(\mathbb{H}^2)$, let $x^+ := \lim_{t\to+\infty} g_t.x$ and $x^- := \lim_{t\to-\infty} g_t.x$ where g_t denotes the geodesic flow. Let $\widehat{x^-x^+}$ be the unique geodesic in $T^1\mathbb{H}^2$ connecting x^- and x^+. By abuse of notation we also let $\widehat{x^-x^+}$ denote its projection to $\mathbb{H}^2$. Fix some point $o \in \mathbb{H}^2$ (say, take $o = (0,1)$), and $x \in T^1\mathbb{H}^2$, let $t = t_o(x)$ be the unique real number such that

$$x = g_t.\pi_{\widehat{x^-x^+}}(o).$$

(a priori, $\pi_{\widehat{x^-x^+}}(o)$ is just an element in $\mathbb{H}^2$ but we identify it with the unique element on $\widehat{x^-x^+} \subset T^1\mathbb{H}^2$ whose projection to $\mathbb{H}^2$ is $\pi_{\widehat{x^-x^+}}(o)$)

EXERCISE 5.7. *The map* $\Phi_o : T^1\mathbb{H}^2 \to (\partial\mathbb{H}^2 \times \partial\mathbb{H}^2 \setminus \Delta\partial\mathbb{H}^2) \times \mathbb{R}$ *defined by*

$$x \mapsto \Phi_o(x) := (x^-, x^+, t_o(x))$$

is a homeomorphism.

This is the so-called Hopf coordinate.

EXERCISE 5.8. *Check that* $\Phi_o(g_t.x) = (x^-, x^+, t_o(x) + t)$.

EXERCISE 5.9. *Check that for* $\gamma \in \mathrm{SL}_2(\mathbb{R})$, $\Phi_o(\gamma.x) = (\gamma.x^-, \gamma.x^+, *)$ *for some real number* $*$.

Thus the orbits of Γ on $T^1\mathbb{H}^2/\{g_t\}_{t\in\mathbb{R}}$ corresponds to the orbits of Γ on $\partial\mathbb{H}^2 \times \partial\mathbb{H}^2 \setminus \Delta\partial\mathbb{H}^2$.

EXERCISE 5.10. *Let* Γ *be a discrete subgroup of* $\mathrm{SL}_2(\mathbb{R})$. *Using the fact that* g_t*-action on* $\Gamma\backslash T^1\mathbb{H}^2$ *is not minimal, show that the action of* Γ *on* $\partial\mathbb{H}^2 \times \partial\mathbb{H}^2 \setminus \Delta\partial\mathbb{H}^2$ *is not minimal.*

This action is still quite chaotic, at least when Γ is a lattice, but if we take one step further, it becomes totally discontinuous.

Let FATΔ be the "fat diagonal" in $\partial\mathbb{H}^2 \times \partial\mathbb{H}^2 \times \partial\mathbb{H}^2$, i.e.

$$\mathrm{FAT}\Delta := \left\{(x_1, x_2, x_3) \in (\partial\mathbb{H}^2)^3,\ x_i = x_j, \exists i \neq j\right\}.$$

EXERCISE 5.11. *Let Γ be a subgroup of $\mathrm{SL}_2(\mathbb{R})$. Show that the diagonal Γ-action on $(\partial\mathbb{H}^2)^3 \setminus$ FATΔ is conjugate to the Γ-action on $\mathbb{H}^2$.*

Now turn to the special Γ_0, Γ_0' we defined. Recall in Lec 2 we have shown that $\mathrm{Limit}(\Gamma_0)$ is the full $\partial\mathbb{H}^2$. Show that also

EXERCISE 5.12. $\mathrm{Limit}(\Gamma_0') = \partial\mathbb{H}^2$.

(Hint: use Exer 5.5 and the fact that Γ_0' is a normal subgroup)

EXERCISE 5.13. *Use this and the "thin" property of hyperbolic space to show that closed geodesics are dense in $\Gamma_0' \backslash T^1\mathbb{H}^2$.*

(In Lec.3 we established denseness of closed geodesics on $\mathrm{SL}_2(\mathbb{R}) / \mathrm{SL}_2(\mathbb{Z})$ by constructing an explicit one and then considering commensurable lattices)

For a point v on $\partial\mathbb{H}^2$ and $x \in \mathbb{H}^2$, let $\mathcal{H}_v(x)$ be the unique horocycle – the unique Euclidean circle tangent to $\partial\mathbb{H}^2$ at $v(\neq \infty)$ and passing through x (when $v = \infty$, $\mathcal{H}_v(x)$ is a horizontal line passing through x). We shall think of $\mathcal{H}_v(x)$ as a subset of $T^1\mathbb{H}^2$ by equipping every point $\mathcal{H}_v(x)$ with the unique unit tangent vector that is orthogonal to $\mathcal{H}_v(x)$ and pointing towards v.

In Lec.2 we have shown that the projection of every horocycle is dense in $\Gamma_0 \backslash T^1\mathbb{H}^2$. Here is a more geometric approach following Hedlund's paper.

EXERCISE 5.14. *Show that for every nonempty open interval $I \subset \partial\mathbb{H}^2$ and $x \in \mathbb{H}^2$, the set*

$$\bigcup_{v \in I} \mathcal{H}_v(x)$$

is dense in $\Gamma_0' \backslash T^1\mathbb{H}^2$.

(Hint: use Exer.5.13)

EXERCISE 5.15. *Let $v \in \partial\mathbb{H}^2$, show that if there exists $x \in \mathbb{H}^2$ such that $\mathcal{H}_v(x)$ is dense in $\Gamma_0' \backslash T^1\mathbb{H}^2$, then $\mathcal{H}_v(y)$ is dense in $\Gamma_0' \backslash T^1\mathbb{H}^2$ for every $y \in \mathbb{H}^2$.*

EXERCISE 5.16. *The set of v such that $\mathcal{H}_v(x)$ is dense in $\Gamma_0' \backslash T^1\mathbb{H}^2$ is dense in $\partial\mathbb{H}^2$.*

Let $\mathcal{D}$ be a Dirichlet fundamental domain for Γ_0'. Accept the fact that if Γ_0' were finitely generated, then $\mathcal{D}$ would have only finitely many sides.

EXERCISE 5.17. *Show that Γ_0' is not finitely generated.*

EXERCISE 5.18. *Let $v \in \partial\mathbb{H}^2 \cap \overline{\mathcal{D}}$, then $\mathcal{H}_v(x)$ is not dense in $\Gamma_0' \backslash T^1\mathbb{H}^2$.*

(Hint: without loss of generality assume $v = \infty$, argue that, fixing a base point o, there is an upper bound for the y-coordinate of $\gamma.o$ as γ varies in Γ_0'.)

Since $\mathcal{H}_v(x)$ is not compact in $\Gamma_0' \backslash T^1\mathbb{H}^2$, we have demonstrated an orbit of the horocycle flow that is neither dense nor compact.

EXERCISE 5.19. *Take some $y \in T^1\mathbb{H}^2$ such that $\{g_t.y\}$ is compact in $\Gamma_0 \backslash T^1\mathbb{H}^2$. Show that $\mathcal{H}_{y^+}(x)$ is dense in $\Gamma_0 \backslash T^1\mathbb{H}^2$.*

(Hint: approximate some dense horocycle in $T^1\mathbb{H}^2$)

EXERCISE 5.20. *Let* $v \in \partial\mathbb{H}^2$ *and fix some* $x \in \mathbb{H}^2$*. Suppose the Euclidean radius of* $\gamma.\mathscr{H}_v(x)$ *can be arbitrarily large as* γ *varies in* Γ_0*. Then* $\mathscr{H}_v(x)$ *is dense in* $\Gamma_0\backslash T^1\mathbb{H}^2$*.*

(When the horocycle is based at infinity, by saying the Euclidean radius is large, we mean that the horocycle could be very low) (Hint: show that you can approximate every periodic geodesic)

EXERCISE 5.21. *Show that indeed, since* $\Gamma_0\backslash T^1\mathbb{H}^2$ *is compact, that the Euclidean radius of* $\gamma.\mathscr{H}_v(x)$ *can be arbitrarily large as* γ *varies in* Γ_0 *for every pair* $v \in \partial\mathbb{H}^2$ *and* $x \in \mathbb{H}^2$*.*

(Hint: use the fact that the some (well, in the current case, every) geodesic stemming from v is bounded in $\Gamma_0\backslash T^1\mathbb{H}^2$)

CHAPTER 2

A dynamical reformulation of Oppenheim conjecture

Back to the Top.

We recommend the last chapter of [**BM00**] for an elementary account of the proof of Oppenheim conjecture. See [**Mar97, LM14, BGHM10**] for history and more recent stories.

1. The statement

The goal of this and the next lecture is to prove a weak Oppenheim conjecture. In this lecture we will reduce the proof to a dynamical statement whose proof is delegated to the next lecture. A stronger form will be treated later with the help of non-divergence of unipotent flows.

THEOREM 1.1. *Let Q be a non-degenerate indefinite quadratic from with real coefficients in $N \geq 3$ variables. Assume that Q is not a scalar multiple of some quadratic form with rational coefficients. Then the closure of $Q(\mathbb{Z}^N \setminus \mathbf{0})$ contains 0.*

REMARK 1.2. *This theorem says nothing nontrivial to the quadratic form $Q_1 = xy - \sqrt{2}z^2$ since $Q_1(1,0,0) = 0$. However, it is nontrivial for $Q_2 = x^2 + y^2 - \sqrt{2}z^2$ since the value of Q_2 at integral points can never be 0 unless $(x,y,z) = (0,0,0)$.*

Later we will specialize to the case when $N = 3$, from which the general case would follow. Details are left to the reader.

REMARK 1.3. *Counter examples exist when $N = 2$. For instance consider the quadratic form $Q(x_1,x_2) := (x_1 - \sqrt{2}x_2)x_2$. Note that $\sqrt{2}$ is badly approximable which means that there exists $c > 0$ such that $\{\sqrt{2}m\}|m| \geq c$ for all non-zero integer m where $\{\cdot\}$ stands for the distance to the nearest integer. We will sketch a dynamical explanation below.*

2. The space of lattices

For a quadratic form Q in N variables, define for $k = \mathbb{R}, \mathbb{Q}, \mathbb{Z}$,

$$\mathrm{SO}_Q(k) := \left\{ g \in \mathrm{SL}_N(k) \,\middle|\, Q \circ g = Q \right\}. \tag{6}$$

The definition makes sense for Q irrational. It might happen that $\mathrm{SO}_Q(\mathbb{Z})$ is trivial. If M_Q is the symmetric matrix representing of Q, i.e. $Q(v) = v^{tr} M_Q v$ (v written as a column vector), then

$$\mathrm{SO}_Q(k) := \left\{ g \in \mathrm{SL}_N(k) \,\middle|\, g^{tr} M_Q g = M_Q \right\}. \tag{7}$$

One can compute the Lie algebra of $\mathrm{SO}_Q(\mathbb{R})$ as

$$\mathfrak{so}_Q = \left\{ X \in \mathfrak{sl}_n(\mathbb{R}) \,\middle|\, M_Q X + X^{tr} M_Q = 0 \right\}.$$

Where does it act on?

DEFINITION 2.1. *A subgroup Λ of $\mathbb{R}^N$ is said to be a **(unimodular) lattice** if Λ is discrete and cocompact in $\mathbb{R}^N$ (with $\mathrm{Vol}(\mathbb{R}^N/\Lambda) = 1$).*

Here Vol is taken with respect to the standard Euclidean metric on $\mathbb{R}^N$.

EXAMPLE 2.2. *$\mathbb{Z}^N$ is a unimodular lattice in $\mathbb{R}^N$.*

EXAMPLE 2.3. *$\mathbb{Z}[\sqrt{2}]$ may be viewed as a lattice in $\mathbb{R}^4$ by the geometric embedding, i.e.*

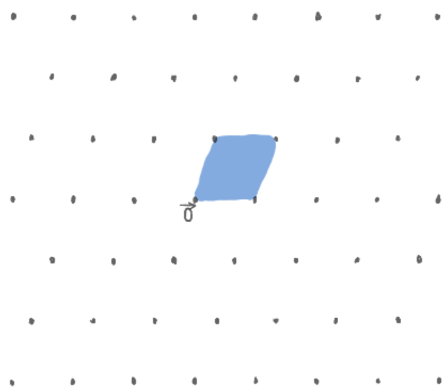
$$\Lambda := \{(x, y) \mid x, y \in \mathbb{Z}[\sqrt{2}],\ y = \sigma(x)\}$$

where σ is the nontrivial element in $\mathrm{Gal}(\mathbb{Q}(\sqrt{2})/\mathbb{Q})$.

EXAMPLE 2.4. *$\mathbb{Z}[\sqrt{-2}]$ may be viewed as a lattice in $\mathbb{R}^2$ by identifying it with $\mathbb{C}$, explicitly,*

$$\Lambda = \{(x, \sqrt{2}y) \mid x, y \in \mathbb{Z}\}.$$

EXAMPLE 2.5. *You can get a unimodular lattice starting from a lattice by multiplying a scalar.*



a lattice together with a fundamental domain

Explicitly, for every discrete subgroup Λ of $\mathbb{R}^N$, one can find $v_1, ..., v_n$ in $\mathbb{R}^N$ that are $\mathbb{R}$-linearly independent and $\Lambda = \mathbb{Z}v_1 \oplus \mathbb{Z}v_2 \oplus ... \oplus \mathbb{Z}v_n$. Such a set $\{v_1, ..., v_n\}$ will be called a **basis** of Λ. And n is called the **rank** of Λ. Λ is a lattice iff $n = N$. Conversely, given n vectors $v_1, ..., v_n$ that are $\mathbb{R}$-linearly independent, the subgroup $\mathbb{Z}v_1 + ... + \mathbb{Z}v_n$ is a discrete subgroup of $\mathbb{R}^N$.

Assume a discrete subgroup Λ has rank N with basis $(v_1, ..., v_N)$. Then $\mathrm{Vol}(\mathbb{R}^N/\Lambda) = |\det(v_1, .., v_N)| = \|v_1 \wedge ... \wedge v_N\|$. This is because

$$\{a_1 v_1 + ... + a_N v_N \mid a_i \in [0, 1)\}$$

forms a strict fundamental domain for $\mathbb{R}^N/\Lambda$, namely, it is in bijection with $\mathbb{R}^N/\Lambda$ under the quotient map. Also, let us recall that

$$\mathrm{Vol}(\mathbb{R}^N/\Lambda) = \|v_1\| \cdot \mathrm{dist}(v_2, \mathbb{R}v_1) \cdot \mathrm{dist}(v_3, \mathbb{R}v_1 + \mathbb{R}v_2) \cdot ... \cdot \mathrm{dist}(v_N, \mathbb{R}v_1 + ... + \mathbb{R}v_{N-1}).$$

Thus Λ is a unimodular lattice iff $\det(v_1, ..., v_N) = \pm 1$.

It is useful to be familiar with quotient constructions in Euclidean spaces. More precisely, an **Euclidean space** is a finite-dimensional $\mathbb{R}$-linear space together with a non-degenerate positive definite quadratic form (or an "inner product", if you prefer). The "standard" $\mathbb{R}^N$ is nothing but the vector space $\mathbb{R}^N$ together with the form $Q_{std}(x_1, ..., x_N) := x_1^2 + ... + x_N^2$. Once an Euclidean space is given, we can talk about distance, volume...

If W is an $\mathbb{R}$-subspace of $\mathbb{R}^N$, then we think of W as an Euclidean space by restricting the quadratic form to W. Since Q_{std} is positive definite, its restriction to every subspace is also positive definite. Also, the quotient $\mathbb{R}^N/W$ is equipped with a natural Euclidean structure by identifying it with the orthogonal complement of W in $\mathbb{R}^N$. Alternatively, you can define the quotient metric on $\mathbb{R}^N/W$ and then argue that it comes from a quadratic form. These two methods give the same Euclidean structure on $\mathbb{R}^N/W$.

DEFINITION 2.6. *Let* X_N *be the set of unimodular lattices in* $\mathbb{R}^N$ *equipped with the Chabauty topology.*

Alternatively, one may think of X_N as the set of all lattices of $\mathbb{R}^N$ up to $\mathbb{R}^*$-action.

A detailed treatment of Chabauty topology may be found in [**BP92**, Chapter E, Section 1]. For us, it suffices to know that under the Chabauty topology, a sequence $(\Lambda_n) \subset X_N$ converges to $\Lambda \in X_N$ iff one can find a basis $v_1^n, ..., v_N^n$ of Λ_n such that as $n \to \infty$, $\left(v_i^n\right)_n$ converges to some $v_i^\infty \in \mathbb{R}^N$ for every $i = 1, ..., N$ and $\Lambda = \oplus_i \mathbb{Z} v_i^\infty$.

Note that for a sequence $(\Lambda_n) \subset X_N$, if there are bases $\left(v_1^n, ..., v_N^n\right)$ with $\left(v_i^n\right)_n$ converging to some $v_i^\infty \in \mathbb{R}^N$ for every i, then $\left(v_1^\infty, ..., v_N^\infty\right)$ are automatically $\mathbb{R}$-linearly independent and they span a lattice Λ with covolume $\mathrm{Vol}(\mathbb{R}^N/\Lambda) = 1$.

The space X_N admits a natural action of $SL_N(\mathbb{R})$ and

LEMMA 2.7. *The map* $g \mapsto g.\mathbb{Z}^N$ *from* $SL_N(\mathbb{R})$ *to* X_N *descends to a homeomorphism* $SL_N(\mathbb{R})/SL_N(\mathbb{Z}) \cong X_N$.

PROOF. $SL_N(\mathbb{Z})$ is equal to the stabilizer of $\mathbb{Z}^N$ in $SL_N(\mathbb{R})$, this proves the injectivity.

For every $\Lambda \in X_N$, find a basis $v_1, ..., v_N$. Replacing v_1 by $-v_1$ if necessary, assume $M := (v_1, ..., v_N)$ (v_i written as column vectors) has determinant 1. Then $M.\mathbb{Z}^N = \Lambda$. This proves the surjectivity.

We leave it to the reader to convince himself/herself that the map is open and continuous. □

DEFINITION 2.8. *For a discrete subgroup* $\Lambda \le \mathbb{R}^N$ *we define*

$$\mathrm{sys}(\Lambda) := \inf_{v \neq 0 \in \Lambda} \|v\| \tag{8}$$

where $\|\cdot\|$ *is the standard Euclidean norm.*

Clearly $\mathrm{sys}(\Lambda) > 0$.

You may interpret $\mathrm{sys}(\Lambda)$ as the length of the smallest geodesic in the quotient flat torus $\mathbb{R}^N/\Lambda$.

One can check that $\mathrm{sys} : X_N \to \mathbb{R}_{>0}$ is continuous.

The following is sometimes referred to as Mahler's criterion.

LEMMA 2.9.
1. *A set* $\mathcal{B} \subset X_N$ *does not have compact closure (we will simply write unbounded later) if for every* $\varepsilon > 0$ *there exists* $\Lambda \in \mathcal{B}$ *with* $\mathrm{sys}(\Lambda) \le \varepsilon$.
2. *For every* $\varepsilon > 0$, *the set*
$$\left\{\Lambda \mid \mathrm{sys}(\Lambda) \ge \varepsilon\right\}$$
is compact in X_N.

DEFINITION 2.10. *For a discrete subgroup* Λ *of* $\mathbb{R}^N$, *we let* $\|\Lambda\| := \mathrm{Vol}(V/\Lambda)$ *where* V *is the* $\mathbb{R}$*-linear span of* Λ. *For a lattice* Λ *of some Euclidean space* V, *we let* $\|\Lambda\|_V := \mathrm{Vol}(V/\Lambda)$.

As we have discussed, if $v_1, ..., v_n$ is a basis of Λ, then

$$\|\Lambda\| = \|v_1\| \cdot \mathrm{dist}(v_2, \mathbb{R}.v_1) \cdot ... \cdot \mathrm{dist}(v_n, \mathbb{R}v_1 + ... + \mathbb{R}v_{n-1}).$$

Let us also remark that $\mathrm{dist}(v_2, \mathbb{R}v_1) = \|v_2\|_{\mathbb{R}^N/\mathbb{R}v_1}$ and more generally

$$\mathrm{dist}(v_k, \mathbb{R}v_1 + ... + \mathbb{R}v_{k-1}) = \|v_k\|_{\mathbb{R}^N/(\mathbb{R}v_1 + ... + \mathbb{R}v_{k-1})}.$$

PROOF OF LEM. 2.9. 1. follows from the continuity of sys. Let us prove 2.

Fix some $\varepsilon > 0$ and take $\Lambda \in X_N$ satisfying $\mathrm{sys}(\Lambda) \ge \varepsilon$. It suffices to construct a basis of Λ with bounded distance to the origin.

Consider the projection $p:\mathbb{R}^N\to\mathbb{R}^N/\Lambda$. As $\mathrm{Vol}(\mathbb{R}^N/\Lambda)=1$, p restricted to the subset $[-1,1]^N$ is not injective. This shows that there exists $v_{\neq 0}\in\Lambda$, $\|v\|\le C_1(N)$ for some positive constant $C_1(N)$ depending only on N. In particular, if we choose $v_1\in\Lambda$ such that

$$\|v_1\|=\mathrm{sys}(\Lambda),$$

then $\|v_1\|\le C_1(N)$. Note that v_1 is primitive in the sense that v_1 is not an integral multiple of any vector in Λ other than $\pm v_1$.

Let π_1 be the projection from $\mathbb{R}^N$ to $V_1:=\mathbb{R}^N/\mathbb{R}v_1$. Since $\Lambda_1:=\pi_1(\Lambda)$ has rank $N-1$ and spans V_1, we have that Λ_1 is discrete and actually a lattice in V_1.

Note that

$$\begin{aligned}1=\|\Lambda\|&=\|v_1\|\cdot\mathrm{dist}(v_2,\mathbb{R}v_1)\cdot\ldots\cdot\mathrm{dist}(v_N,\mathbb{R}v_1+\ldots+\mathbb{R}v_{N-1})\\&=\|v_1\|\cdot\|\pi_1(v_2)\|_{V_1}\cdot\mathrm{dist}(\pi_1(v_3),\mathbb{R}\pi_1(v_2))\cdot\ldots\cdot\mathrm{dist}(\pi_1(v_N),\mathbb{R}\pi_1(v_2)+\ldots+\mathbb{R}\pi_1(v_{N-1}))\\&=\|v_1\|\cdot\|\Lambda_1\|_{V_1}\ge\varepsilon\cdot\|\Lambda_1\|_{V_1}\\&\implies\|\Lambda_1\|_{V_1}\le\varepsilon^{-1}=:C_2(\varepsilon).\end{aligned}$$

Now choose $v_2\in\Lambda\setminus\mathbb{R}v_1$ such that

$$\|\pi_1(v_2)\|=\mathrm{sys}_{V_1}(\Lambda_1).$$

A similar argument as above shows that $\|\pi_1(v_2)\|<C_3(N,\varepsilon)$. By modifying v_2 by some integral multiple of v_1, we assume that $\|v_2\|<C_3(N,\varepsilon)=C_3$ with a possibly larger C_3.

Next we want to argue that $\mathrm{sys}_{V_1}(\Lambda_1)>c_1(N,\varepsilon)$ for some constant $c_1(N,\varepsilon)>0$ (we will soon see that can take $c_1=0.4\varepsilon$) depending only on N,ε.

Say we have a nonzero vector in Λ_1 of length smaller than λ. Then its lift $v\in\Lambda$ has the property $0<\mathrm{dist}(v,\mathbb{R}v_1)<\lambda$. So if we write $v=x.v_1+w$ for some w orthogonal to v_1 then $\|w\|\le\lambda$. Let n_x be the nearest integer to x, then $v':=(x-n_x)v_1+w\in\Lambda$ has norm

$$\left\|v'\right\|\le|x-n_x|\,\|v_1\|+\lambda\le 0.5\,\|v_1\|+\lambda$$

So if we had chosen $\lambda=0.4\varepsilon\le 0.4\,\mathrm{sys}(\Lambda)$, then $\left\|v'\right\|\le 0.9\,\mathrm{sys}(\Lambda)$, which is a contradiction. Thus every non-zero vector in Λ_1 has length greater than 0.4ε.

Let π_2 be the natural projection $\mathbb{R}^N\to\mathbb{R}^N/(\mathbb{R}v_1+\mathbb{R}v_2)=:V_2$. By abuse of notation, also denote the natural projection $V_1\to V_2$ by π_2.

With similar arguments, $\Lambda_2:=\pi_2(\Lambda_1)$ is a lattice in V_2 and

$$\|\Lambda_1\|_{V_1}=\|\pi_1(v_2)\|\cdot\|\Lambda_2\|_{V_2}\implies\|\Lambda_2\|_{V_2}\le c_1^{-1}\cdot C_2=:C_4(N,\varepsilon)=:C_4.$$

With similar arguments, $\mathrm{sys}_{V_2}(\Lambda_2)>c_2(N,\varepsilon)$. So we can find v_3, ... up to v_N with bounded norms. And one can check that each step you get a primitive subgroup of Λ and $\{v_1,\ldots,v_N\}$ forms a basis of Λ. So we are done. □

DEFINITION 2.11. *A* ***primitive subgroup*** *of Λ is a subgroup Δ such that the $\mathbb{Q}$-span (or equivalently, the $\mathbb{R}$-span) of Δ intersecting with Λ gives back Δ.*

The $\mathbb{Z}$-span of two primitive subgroups may not be primitive. e.g., consider $(1,1),(1,-1)$ in $\mathbb{Z}^2$, each of which is primitive, but they span a index 2 subgroup of $\mathbb{Z}^2$, hence not primitive.

3. Values of a quadratic form and orbits of its symmetric group

Now comes the equivalent formulation of weak Oppenheim. For a rational quadratic form Q, this would imply that $\mathrm{SO}_Q(\mathbb{Z})$ is not cocompact in $\mathrm{SO}_Q(\mathbb{R})$ if $Q(v)=0$ admits a solution in $v_{\neq 0}\in\mathbb{Z}^N$ (in which case we say Q is isotropic over $\mathbb{Q}$). When $N\ge 5$, a rational indefinite quadratic form is always isotropic over $\mathbb{Q}$ (see [**O'M63**, 63:19, 66:1])

LEMMA 3.1. *For a non-degenerate quadratic form Q in N variables with real coefficients, the following two are equivalent:*

1. *the closure of* $Q(\mathbb{Z}^N \setminus 0)$ *contains* 0;
2. *the orbit closure of* $\mathrm{SO}_Q(\mathbb{R})$ *based at the identity coset is unbounded in* $\mathrm{X_N}$. *In other words,* $\mathrm{SO}_Q(\mathbb{R}).\mathbb{Z}^N$ *contains non-zero vectors of arbitrarily small length.*

PROOF OF $2 \implies 1$. By assumption and Mahler's criterion, there exists $g_n \in \mathrm{SO}_Q(\mathbb{R})$ and $u_n(\neq 0) \in \mathbb{Z}^N$ such that $g_n \cdot u_n$ tends to $\mathbf{0}$. Hence

$$Q(u_n) = Q(g_n \cdot u_n) \to 0.$$

And we are done. □

REMARK 3.2. *For the proof of Thm.1.1 this direction is sufficient. However we feel that it is conceptually better to do the converse, too. Actually, this provides a different way of understanding why Thm.1.1 fails when $N = 2$ – it suffices to find a bounded, yet non-closed orbit of the diagonal group A on* $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$. *And one can do this by constructing two closed orbits of A and a third orbit Ay such that in the forward direction, Ay approximates one closed orbit and in the backward direction Ax approximates the other. This relies on the fact that closed A-orbits are dense (for instance, one can find one by explicit construction and then consider all lattices commensurable to it) and an argument with local coordinates in stable/unstable/flow direction.*

Why is this sufficient? Note that if Q is an indefinite rational quadratic form in two variable, then either Q is $\mathbb{Q}$-equivalent to $Q_0 = xy$ or $Q_1 = x^2 - by^2$ for some $b > 0$ and $\sqrt{b} \notin \mathbb{Q}$. In the former case, the orbit of $\mathrm{SO}_Q(\mathbb{R})$ *based at the identity coset diverges (that is, the orbit map is proper) and in the second case the orbit is compact, stabilizer of which comes from certain elements in* $\mathbb{Q}(\sqrt{b})$.

Now go back to the proof of $1 \implies 2$ of Lem.3.1. We need the following fact.

LEMMA 3.3. *For every* $r_{\neq 0} \in \mathbb{R}$, $\mathrm{SO}_Q(\mathbb{R})$ *acts transitively on the level set*

$$V_r := \left\{ v \in \mathbb{R}^N \,\middle|\, Q(v) = r \right\}.$$

And for $r = 0$, there are at most 2-orbits of $\mathrm{SO}_Q(\mathbb{R})$ *on* $V_0 \setminus \{\mathbf{0}\}$.

PROOF. By linear algebra, up to change of $\mathbb{R}$-coordinate (i.e. up to $\mathrm{GL}_N(\mathbb{R})$), we may and do assume that Q takes the form

$$Q(x_1,...,x_N) = (x_1^2 + ... + x_s^2) - (x_{s+1}^2 + ... + x_{s+t}^2) =: Q_1(x_1,...,x_s) - Q_2(x_{s+1},...,x_{s+t})$$

for some $s + t = N$ and $s, t \in \mathbb{Z}_{\geq 0}$. When one of s, t is equal to 0, the form is definite and we assume that we already know how to handle this case.

For $\boldsymbol{x} \in \mathbb{R}^N$, we write $v_x := (x_1,...,x_s)$ and $w_x := (x_{s+1},...,x_N)$.

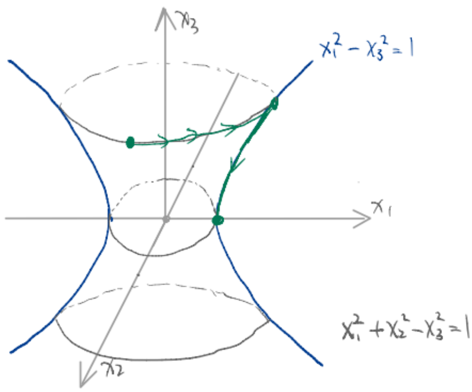
Now we fix r_0 and if V_{r_0} is empty there is nothing to prove. So assume otherwise and take $\boldsymbol{x}_0 \in V_{r_0}$. Let $r_1 := Q_1(v_{x_0})$ and $r_2 := Q_2(w_{x_0})$. Thus by transitivity in the (positive) definite case, we can find $k_i \in \mathrm{SO}_{Q_i}(\mathbb{R})$ (i=1,2) such that

$$k_1.v_{x_0} = (\sqrt{r_1}, 0, ..., 0)$$
$$k_2.w_{x_0} = (\sqrt{r_2}, 0, ..., 0).$$

Let $\mathrm{SO}_{(x_1^2 - x_{s+1}^2)}(\mathbb{R})$ be embedded in $\mathrm{SO}_Q(\mathbb{R})$ by leaving the rest of the coordinates unchanged. When $r = r_1 - r_2 \neq 0$, it acts on $x_1^2 - x_{s+1}^2 = r$ transitively. The level sets are not connected, but the group $\mathrm{SO}_{(xy)}(\mathbb{R})$ is also not! Both have 2 components.

As for the case $r = 0$, one can show that $\mathrm{SO}_{(x_1^2 - x_{s+1}^2)}(\mathbb{R})$ has two orbits on $x_1^2 - x_{s+1}^2 = 0$. □

Here is an illustration of the proof by pictures



REMARK 3.4. *Assume Q is indefinite non-degenerate. One can further show that when $N \geq 3$, $\mathrm{SO}_Q(\mathbb{R})$ acts on $V_0 \setminus \mathbf{0}$ transitively. When $N = 2$, $\mathrm{SO}_Q(\mathbb{R})$ acts on $V_0 \setminus \mathbf{0}$ with exact two orbits.*

PROOF OF $1 \implies 2$. By assumption for every $\varepsilon > 0$ there exists $u_{\varepsilon \neq 0} \in \mathbb{Z}^N$ such that $|Q(u_\varepsilon)| \leq \varepsilon$. On the other hand, By the (proof of) Lem.3.3, there exists non-zero $u_\varepsilon^1, u_\varepsilon^2 \in \mathrm{SO}_Q(\mathbb{R}).u_\varepsilon$ such that

$$\|u_\varepsilon^1\|, \|u_\varepsilon^2\| \leq \theta(Q,\varepsilon) = \theta$$

where θ tends to 0 (for a fix Q) as ε does so. Hence $\mathrm{sys}(g_\varepsilon \mathbb{Z}^N) \leq \theta$ and we see that $\mathrm{SO}_Q(\mathbb{R}) \cdot \mathbb{Z}^N$ is unbounded as $\varepsilon \to 0$ by Lem.2.9. □

Now we specialize to $N = 3$.

In light of Lem.3.1, to prove Thm.1.1, it is sufficient to show that $\mathrm{SO}_Q(\mathbb{R}) \cdot \mathbb{Z}^3$ is unbounded. Find $g_0 \in \mathrm{SL}_3(\mathbb{R})$ such that $Q \circ g_0^{-1}$ is a scalar multiple of $Q_0 = 2x_1x_3 - x_2^2$. Then

$$\mathrm{SO}_{Q_0} = g_0 \, \mathrm{SO}_Q \, g_0^{-1}.$$

So sufficient to show that $\mathrm{SO}_{Q_0}(\mathbb{R}) g_0.\mathbb{Z}^3$ is unbounded in X_3, which will follow from

THEOREM 3.5. *Let $\Lambda \in \mathrm{X}_3$ be such that $\mathrm{SO}_{Q_0}(\mathbb{R}).\Lambda$ is bounded, then $\mathrm{SO}_{Q_0}(\mathbb{R}).\Lambda$ is closed, and hence compact.*

In some sense we cheated a little bit. Because we are going to use a trick on quadratic forms . And the true dynamical result we are going to prove is (to be proved in the next lecture):

THEOREM 3.6. *Let $\Lambda \in \mathrm{X}_3$ be such that $\mathrm{SO}_{Q_0}(\mathbb{R}).\Lambda$ is bounded, then either $\mathrm{SO}_{Q_0}(\mathbb{R}).\Lambda$ is closed and hence compact, or the closure of $\mathrm{SO}_{Q_0}(\mathbb{R}).\Lambda$ contains a $\{\mathbf{v}_s\}_{s\geq 0}$-orbit or a $\{\mathbf{v}_s\}_{s\leq 0}$-orbit. where*

$$\mathbf{v}_s := \exp\left(s \cdot \begin{bmatrix} 0 & 0 & 1 \\ & 0 & 0 \\ & & 0 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & s \\ & 1 & 0 \\ & & 1 \end{bmatrix}.$$

Note that $\{\mathbf{v}_s\}$ is not contained in $\mathrm{SO}_{Q_0}(\mathbb{R})$.

PROOF OF THM.3.5 ASSUMING THM.3.6. Say, we have a $\{\mathbf{v}_s\}_{s\geq 0}$-orbit (the other case is similar) based at Λ' for some $\Lambda' \in \overline{\mathrm{SO}_{Q_0}(\mathbb{R}).\Lambda}$. Write $\mathbf{x} = (x_1, x_2, x_3) \in \Lambda'$. Then

$$Q_0(\mathbf{v}_s.\mathbf{x}) = Q_0(x_1 + sx_3, x_2, x_3) = (2x_3^2)s + (2x_1x_3 - x_2^2).$$

First we can find some $\mathbf{x} \in \Lambda'$ such that $Q_0(\mathbf{x}) < 0$ and $x_3 \neq 0$ (I leave it to you to convince yourself that this is possible). Then there is some s (replace x_1 by $-x_1$ if necessary) with $Q_0(\mathbf{v}_s \cdot \mathbf{x}) = 0$. By Lem.3.1, this implies $\mathrm{SO}_{Q_0}(\mathbb{R})\mathbf{v}_s \cdot \Lambda \subset \overline{\mathrm{SO}_{Q_0}(\mathbb{R}) \cdot \Lambda}$ is unbounded. □

PROOF OF THM.1.1 ASSUMING THM.3.6. To prove Thm.1.1, by Lem.3.1, if $\mathrm{SO}_{Q_0}(\mathbb{R})g_0.\mathbb{Z}^3$ is unbounded in X_3 then we are done. Now we assume otherwise. If $\mathrm{SO}_{Q_0}(\mathbb{R})g_0.\mathbb{Z}^3$ is compact, or equivalently, $\mathrm{SO}_Q(\mathbb{R}).\mathbb{Z}^3$ is compact, then by Lem.3.7 below, Q is proportional to a rational quadratic form, contradiction. Thus we have a $\{\mathbf{v}_s\}_{s\geq 0}$ (the other case $s \leq 0$ is similar) orbit in the closure of $\mathrm{SO}_{Q_0}(\mathbb{R})g_0.\mathbb{Z}^3$. Repeat the argument above, we find $s \in \mathbb{R}$ such that $Q_0(\mathbf{v}_s.x) = 0$ for some $x \in g_0\mathbb{Z}^3$. But $\mathbf{v}_s.g_0\mathbb{Z}^3$ is in the closure of $\mathrm{SO}_{Q_0}(\mathbb{R})g_0.\mathbb{Z}^3$, implying that we can find $(v_n) \subset g_0\mathbb{Z}^3$, $(g_n) \subset \mathrm{SO}_{Q_0}(\mathbb{R})$ such that $g_n.v_n \to \mathbf{v}_s.x$. Hence

$$Q_0(v_n) = Q_0(g_n v_n) \to Q_0(\mathbf{v}_s.x) = 0.$$

Thus the closure of $Q(\mathbb{Z}^3) = Q_0(g_0.\mathbb{Z}^3)$ contains 0. $\square$

LEMMA 3.7. *For a non-degenerate quadratic form Q, if $\mathrm{SO}_Q(\mathbb{Z})$ is cocompact in $\mathrm{SO}_Q(\mathbb{R})$, then Q is a multiple of a rational quadratic form.*

Note that if Q is NOT a multiple of a rational quadratic form, then for some non-zero coefficients α, β of Q, one has $\alpha/\beta \notin \mathbb{Q}$. Hence there exists $\sigma \in \mathrm{Aut}(\mathbb{R}/\mathbb{Q})$ such that $\sigma(\alpha/\beta) \neq \alpha/\beta$, in particular, σQ is not proportional to Q.

So it suffices to complete

Step 1. for every $\sigma \in \mathrm{Aut}(\mathbb{R}/\mathbb{Q})$, show $\mathrm{SO}_Q(\mathbb{R})^\circ = \mathrm{SO}_{\sigma(Q)}(\mathbb{R})^\circ$;

Step 2. for every pair Q_1, Q_2 of non-degenerate quadratic forms of the same rank, show $\mathrm{SO}_{Q_1}(\mathbb{R})^\circ = \mathrm{SO}_{Q_2}(\mathbb{R})^\circ \implies Q_1 = \lambda Q_2$ for some $\lambda \in \mathbb{R}_{\neq 0}$.

STEP 1. First note that

$$\mathrm{SO}_{\sigma(Q)}(\mathbb{R}) = \sigma(\mathrm{SO}_Q(\mathbb{R})) \supset \mathrm{SO}_Q(\mathbb{Z}).$$

Consider the linear representation

$$\mathrm{SL}_3(\mathbb{R}) \curvearrowright \mathrm{Sym} := \{\mathbb{R} - \text{Symmetric matrices}\}, \quad g.M := gMg^{tr},$$

and the map (call it ϕ) $g \mapsto g.\sigma(Q)$ from $\mathrm{SO}_Q(\mathbb{R})$ to Sym. Then ϕ factors through

$$\mathrm{SO}_Q(\mathbb{R})/\mathrm{SO}_Q(\mathbb{Z}) \to \mathrm{Sym}$$

and hence has compact (and bounded) image. Now we need two facts

1. $\mathrm{SO}_Q(\mathbb{R})^\circ$ is generated (as closed subgroup, which follows by a Lie algebra calculation) by one-parameter unipotent flows $\{\mathbf{u}_t := \exp \mathbf{u}t\}_{t\in\mathbb{R}}$ ($\mathbf{u}$ is some nilpotent matrix in $\mathfrak{so}_Q(\mathbb{R})$);
2. For every unipotent flow $\{\mathbf{u}_t\}$ and $M \in \mathrm{Sym}$, either $\{\mathbf{u}_t \cdot M\}$ is unbounded or M is fixed by $\{\mathbf{u}_t\}$. (if you do not believe this, do some explicit calculation with upper triangular unipotent flows)

But we already saw that $\mathrm{SO}_Q(\mathbb{R}).\sigma(Q)$ is bounded, thus $\mathrm{SO}_Q(\mathbb{R})^\circ$ fixes $\sigma(Q)$. So $\mathrm{SO}_Q(\mathbb{R})^\circ$ is contained in $\mathrm{SO}_{\sigma(Q)}(\mathbb{R})$. But they are both Lie subgroups of $\mathrm{SL}_3(\mathbb{R})$ of the same dimension, so we must have

$$\mathrm{SO}_Q(\mathbb{R})^\circ = \mathrm{SO}_{\sigma(Q)}(\mathbb{R})^\circ.$$

$\square$

STEP 2. By conjugation we assume $Q_1 = Q_0 = 2x_1x_3 - 2x_2^2$. One can compute that $\mathfrak{so}_{Q_0}(\mathbb{R})$ contains (and is generated by)

$$\begin{bmatrix} 1 & & \\ & 0 & \\ & & -1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ & 0 & 1 \\ & & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & & \\ 1 & 0 & \\ 0 & 1 & 0 \end{bmatrix}$$

(note that they do not form an $\mathfrak{sl}_2$-triple, you should multiply the first and the second, but not the third, by 2) and hence $\mathrm{SO}_{Q_0}(\mathbb{R})$ contains

$$a_t := \begin{bmatrix} e^t & & \\ & 1 & \\ & & e^{-t} \end{bmatrix}, \quad u_s := \begin{bmatrix} 1 & s & s^2/2 \\ & 1 & s \\ & & 1 \end{bmatrix}, \quad u_s^- := \begin{bmatrix} 1 & & \\ s & 1 & \\ s^2/2 & s & 1 \end{bmatrix}. \tag{9}$$

Then a direct computation (at the level of Lie algebra is perhaps easier) shows that in order for $\mathfrak{so}_{Q_2}(\mathbb{R})$ to contain these elements, Q_2 must be a scalar multiple of Q_1 and we are done. □

4. Exercises

4.1. Non-commensurable lattices in $\mathrm{SL}_2(\mathbb{R})$, I. We apply ideas in Lec.4 to a different example. Our ultimate goal is to show that two cocompact lattices in $\mathrm{SL}_2(\mathbb{R})$ is either commensurable or their product is dense in $\mathrm{SL}_2(\mathbb{R})$, which will (hopefully) be achieved in the next set of exercises.

Notations:

- $\mathrm{G} := \mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})$, $\mathrm{H} := \Delta(\mathrm{SL}_2(\mathbb{R}))$ and Γ is a cocompact lattice in G;
- $\mathfrak{g} := \mathrm{Lie}(\mathrm{G})$ and $\mathfrak{h} := \mathrm{Lie}(\mathrm{H})$;
- $A := \left\{\left(\begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}, \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}\right), t \in \mathbb{R}\right\} = \{\Delta \mathbf{a}_t, t \in \mathbb{R}\}$;
- $U := \left\{\left(\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}\right), t \in \mathbb{R}\right\} = \{\Delta \mathbf{u}_t, t \in \mathbb{R}\}$;
- $V := \left\{\left(\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & -t \\ 0 & 1 \end{bmatrix}\right), t \in \mathbb{R}\right\} = \{\mathbf{v}_t, t \in \mathbb{R}\}$;
- $V^+ := \{\mathbf{v}_t, t \geq 0\}$, $V^- := \{\mathbf{v}_t, t \leq 0\}$;
- $W := AUV$, $W^+ := AUV^+$, $W^- := AUV^-$.

EXERCISE 4.1. *Show that W is a group and W^+, W^- are semigroups.*

EXERCISE 4.2. *Let*

$$\mathfrak{h}^\perp := \{(X, -X) \mid X \in \mathfrak{sl}_2(\mathbb{R})\} \subset \mathfrak{g}.$$

Show that $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{h}^\perp$ and this decomposition is preserved by $\mathrm{Ad}(\mathrm{H})$.

Now take $\Lambda_0 \in G/\Gamma$ such that $\mathrm{H}.\Lambda_0$ is not closed. Define $\mathrm{Y}_0 := \overline{\mathrm{H}.\Lambda_0}$ and

$$\mathscr{O} := \left\{y \in \mathrm{Y}_0 \,\middle|\, \mathrm{H}.y \text{ is open in } \mathrm{Y}_0\right\}.$$

EXERCISE 4.3. *Show that $\mathscr{O} \neq \mathrm{Y}_0$.*

Let Y_1 be a nonempty U-minimal set in $\mathrm{Y}_0 \setminus \mathscr{O}$.

EXERCISE 4.4. *Show that Y_1 is not a closed U-orbit.*

EXERCISE 4.5. *Assume Y_1 is not preserved by A. Show that Y_0 contains a W-orbit.*

(Hint: consider $\mathrm{Aut}(\mathrm{Y}_1)$.)

EXERCISE 4.6. *Assume Y_1 is preserved by A. Show that Y_0 contains a W^+-orbit or a W^--orbit.*

(Hint: consider $\mathrm{Map}(\mathrm{Y}_0, \mathrm{Y}_1)$.)

4.2. Totally geodesic hyperbolic planes in $\mathbb{H}^3$, I. We apply ideas in Lec.4 to yet another example. Our ultimate goal (hopefully achieved in the next set of exercises) is to show that the image of a totally geodesic immersion of a hyperbolic plane in a closed hyperbolic three manifold is either closed or dense.

Notations:

- $\mathrm{G} := \mathrm{SL}_2(\mathbb{C})$, $\mathrm{H} := \mathrm{SL}_2(\mathbb{R})$ and Γ is a cocompact lattice in G;
- $\mathfrak{g} := \mathrm{Lie}(\mathrm{G})$ and $\mathfrak{h} := \mathrm{Lie}(\mathrm{H})$;
- $A := \left\{ \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix},\ t \in \mathbb{R} \right\} = \{\mathbf{a}_t,\ t \in \mathbb{R}\}$;
- $U := \left\{ \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix},\ t \in \mathbb{R} \right\} = \{\mathbf{u}_t,\ t \in \mathbb{R}\}$;
- $V := \left\{ \begin{bmatrix} 1 & it \\ 0 & 1 \end{bmatrix},\ t \in \mathbb{R} \right\} = \{\mathbf{v}_t,\ t \in \mathbb{R}\}$;
- $V^+ := \{\mathbf{v}_t,\ t \geq 0\}$, $V^- := \{\mathbf{v}_t,\ t \leq 0\}$;
- $W := AUV$, $W^+ := AUV^+$, $W^- := AUV^-$;

EXERCISE 4.7. *Let* $\mathfrak{h}^\perp := \{i \cdot X,\ X \in \mathfrak{sl}_2(\mathbb{R})\}$. *Show that* $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{h}^\perp$. *Moreover, this decomposition is preserved by the* $\mathrm{Ad}(\mathrm{H})$*-action.*

EXERCISE 4.8. *Let* $\mathrm{H}.\Lambda_0$ *be a non-closed* H*-orbit in* G/Γ. *Show that* $\mathrm{Y}_0 := \overline{\mathrm{H}.\Lambda_0}$ *contains a* W^+ *or a* W^-*-orbit.*

CHAPTER 3

Orbit closure of orthogonal groups in the space of lattices

Back to the Top.

Notations and assumptions.

- $Q_0(x_1, x_2, x_3) = 2x_1x_3 - x_2^2$;
- $\mathrm{H} := \mathrm{SO}_{Q_0}(\mathbb{R}) \le \mathrm{G} := \mathrm{SL}_3(\mathbb{R})$, $\mathrm{X}_3 \cong \mathrm{SL}_3(\mathbb{R}) / \mathrm{SL}_3(\mathbb{Z})$;
- $\mathrm{A} := \left\{ \mathbf{a}_t := \begin{bmatrix} e^t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-t} \end{bmatrix} \,\middle|\, t \in \mathbb{R} \right\}$, $\mathrm{U} := \left\{ \mathbf{u}_s := \begin{bmatrix} 1 & s & s^2 \\ 0 & 1 & s \\ 0 & 0 & 1 \end{bmatrix} \,\middle|\, s \in \mathbb{R} \right\}$;
- $\mathrm{V} := \left\{ \mathbf{v}_s := \begin{bmatrix} 1 & 0 & s \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \,\middle|\, s \in \mathbb{R} \right\}$;
- $u_0 := \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad v_0 := \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$;
- $\mathfrak{h}$ is the Lie algebra of H and $\mathfrak{h}^\perp$ denotes its orthogonal complement, see Sec.3.1 for explicit calculations;
- Fix some $x_0 \in \mathrm{X}_3$ with $\mathrm{H}.x_0$ being bounded and non-closed. Write Y_0 for the closure of $\mathrm{H}.x_0$.

See last lecture for the precise definition of $\mathfrak{h}^\perp$.

1. Overview

In this lecture we prove Theorem 3.6 from Chapter 2.

THEOREM 1.1. *Let $\Lambda \in \mathrm{X}_3$ be such that $\mathrm{SO}_{Q_0}(\mathbb{R}).\Lambda$ is bounded, then either $\mathrm{SO}_{Q_0}(\mathbb{R}).\Lambda$ is closed and hence compact, or the closure of $\mathrm{SO}_{Q_0}(\mathbb{R}).\Lambda$ contains a $\{\mathbf{v}_s\}_{s\ge 0}$-orbit or a $\{\mathbf{v}_s\}_{s\le 0}$-orbit.*

By comparison, the ultimate knowledge regarding this is:

THEOREM 1.2. *Every $\mathrm{SO}_{Q_0}(\mathbb{R})$-orbit in X_3 is either closed or dense.*

Outline of proof. Recall that from Chapter 1, we wish to obtain something nontrivial that preserve Y_0 (or a minimal subset of Y_0) in the direction of the normalizer. The normalizer of H is basically H itself. We restrict our attention to an one-parameter unipotent subgroup U of H, then we can start to apply the same argument as in Ch.1. Take a U-minimal subset Y_1 of Y_0. The possibility of having a closed U orbit is excluded. Then by arguments as in Ch.1, we would have some additional elements preserving Y_1 in the normalizer of U. Under the current situation, two possibilities for these additional invariants exist. They could be contained in A or V (or in between). So for our purpose we may assume that the U-minimal set is A-stable. And to treat this case, instead of considering those preserving Y_1, we consider the set of $g \in \mathrm{G}$ mapping Y_0 to Y_1. The lack of group structure here would cause us some difficulty, taken care of by A.

2. The proof

In this section we prove Thm.1.1.

Consider the following

$$\mathscr{O} := \{y \in \mathrm{Y}_0 \,|\, \mathrm{H}\cdot y \text{ is open in } \mathrm{Y}_0\}$$

Thus $\mathscr{O}$ is an H-invariant open (possibly empty) subset of Y_0, in other words, $\mathrm{Y}_0 \setminus \mathscr{O}$ is an H-invariant compact set.

LEMMA 2.1. *$\mathscr{O} \neq \mathrm{Y}_0$ unless $\mathrm{Y}_0 = \mathrm{H}\cdot x_0$.*

PROOF. Otherwise each H-orbit is open, and hence closed in Y_0. In particular $\mathrm{H}.x_0$ is closed. But Y_0 is not closed, so here is a contradiction. □

Eventually we would know $\mathscr{O}$ is empty, but this is what we can do at the moment.

Now take Y_1 to be a nonempty U-minimal set in $\mathrm{Y}\setminus\mathscr{O}$. There are three cases to consider

Case 1. Y_1 is a compact U-orbit;
Case 2. Y_1 does not fall in case 1 and Y_1 is A-invariant ;
Case 3. Y_1 does not fall in case 1 or 2.

Actually, case 1 is not an option since Y_1 is bounded and hence has a lower bound on injectivity radius. So we are left with case 2 and 3.

Before we proceed, let us note that given $x \in \mathrm{X}_3$, for y close enough to x, there exists unique small $h_y \in \mathfrak{h}$ and small $w_y \in \mathfrak{h}^{\perp}$ such that

$$y = \exp(h_y)\exp(w_y).x.$$

2.1. Case 2.

In this case, we are going to show that

$$\mathbf{v}_{s\geq 0} \text{ or } \mathbf{v}_{s\leq 0} \subset \{g \in \mathrm{H} \,|\, g\mathrm{Y}_1 \subset \mathrm{Y}_0\},$$

which implies the conclusion of Thm.1.1.

Since $\mathrm{Y}_1 \subset \mathrm{Y}_0 \setminus \mathscr{O}$ and by the definition of $\mathscr{O}$, for every $x \in \mathrm{Y}_1$, there exists $y_n \to x$ with $y_n \in \mathrm{Y}_0$ and

$$y_n = \exp(h_n)\exp(w_n)x$$

where $h_n \in \mathfrak{h}$, $w_n \in \mathfrak{h}^{\perp}$ both converging to 0 and $w_n \neq 0$. Replacing y_n by $\exp(-h_n)y_n$ we assume $h_n = 0$. The case when w_n belongs to Lie(V) for infinitely many n's is easier and we assume this is not the case.

LEMMA 2.2. *Assume w_n does not belong to* Lie(V) *for n large enough. For $\delta > 0$ small enough and n large enough, there exists $t_{n,\delta}$ such that*

1. $\|\mathrm{Ad}(\mathbf{u}_{t_{n,\delta}})\cdot w_n\| \in [\frac{\delta}{10^{10}}, 10^{10}\delta]$;
2. *every limit point of* $\left(\mathrm{Ad}(\mathbf{u}_{t_{n,\delta}})\cdot w_n\right)$ *lies in* Lie(V).

See Sec.3.4 for the proof. Now assume the lemma and choose $t_{n,\delta}$ as above. Define

$$x_{n,\delta} := \mathbf{u}_{t_{n,\delta}}.x_0, \quad y_{n,\delta} := \mathbf{u}_{t_{n,\delta}}.y_n,$$

then

$$y_{n,\delta} = \exp\left(\mathrm{Ad}(\mathbf{u}_{t_{n,\delta}})\cdot w_n\right).x_{n,\delta}.$$

By passing to a subsequence depending on δ, we assume $\lim x_{n,\delta} = x_{\infty,\delta} \in \mathrm{Y}_1$, $\lim y_{n,\delta} = y_{\infty,\delta} \in \mathrm{Y}_0$ and $\lim \mathrm{Ad}(\mathbf{u}_{t_{n,\delta}})\cdot w_n = s_\delta v_0$ for some $|s_\delta| \in [\frac{\delta}{10^{10}}, 10^{10}\delta]$. Hence

$$y_{\infty,\delta} = \exp(s_\delta v_0).x_{\infty,\delta} = \mathbf{v}_{s_\delta}.x_{\infty,\delta}$$

with $s_\delta \to 0$ as δ does so. Thus

$$\begin{aligned} &\mathrm{Y}_0 \supset \mathrm{U}\mathbf{v}_{s_\delta}.x_{\infty,\delta} = \mathbf{v}_{s_\delta}\mathrm{U}.x_{\infty,\delta} \\ \Longrightarrow\ &\mathrm{Y}_0 \supset \overline{\mathbf{v}_{s_\delta}\mathrm{U}.x_{\infty,\delta}} = \mathbf{v}_{s_\delta}\mathrm{Y}_1. \end{aligned}$$

The closed set

$$\left\{g \in \mathrm{H} \,\middle|\, g\mathrm{Y}_1 \subset \mathrm{Y}_0\right\}$$

is not necessarily a group. Hence we can not conclude the existence of a $\mathbf{v}_{\mathbb{R}}$ (or half of it) orbit inside Y_0 immediately. This is where the assumption that Y_1 is A-invariant steps in. Indeed,

$$\mathbf{v}_{e^{2t}s_\delta}\mathrm{Y}_1 = \mathbf{a}_t\mathbf{v}_{s_\delta}\mathbf{a}_t^{-1}\mathrm{Y}_1 = \mathbf{a}_t\mathbf{v}_{s_\delta}\mathrm{Y}_1 \subset \mathbf{a}_t\mathrm{Y}_0 = \mathrm{Y}_0, \quad \forall t \in \mathbb{R},$$

so depending on the sign of s_δ, $\left\{g \in \mathrm{H} \,\middle|\, g\mathrm{Y}_1 \subset \mathrm{Y}_0\right\}$ contains $\mathbf{v}_{s\geq 0}$ or $\mathbf{v}_{s\leq 0}$. We are done.

2.2. Case 3. In this case, we are going to show that

$$\mathrm{V} \subset \left\{g \in \mathrm{H} \,\middle|\, g\mathrm{Y}_1 = \mathrm{Y}_1\right\},$$

which implies the conclusion of Thm.1.1.

Take $x \in \mathrm{Y}_1$. Since $\mathrm{U}.x$ is not closed, we can find $y_n = \exp(h_n)\exp(w_n)x \in \mathrm{Y}_1$ with $h_n \in \mathfrak{h}$, $w_n \in \mathfrak{h}^\perp$, $h_n, w_n \to 0$ and $h_n + w_n \notin \mathrm{Lie}(\mathrm{U})$. We can no longer assume $h_n = 0$.

LEMMA 2.3. *For $\delta > 0$ small enough and n large enough, there exist $t_{n,\delta}$ and $s_{n,\delta}$ such that*

$$\mathbf{u}_{s_{n,\delta}} \cdot \mathbf{u}_{t_{n,\delta}} \exp(h_n)\exp(w_n)\mathbf{u}_{t_{n,\delta}}^{-1} = \exp(h_{n,\delta})\exp(w_{n,\delta}),$$

for some $h_{n,\delta} \in \mathfrak{h}$, $w_{n,\delta} \in \mathfrak{h}^\perp$ with

$$\max\left\{\left\|h_{n,\delta}\right\|, \left\|w_{n,\delta}\right\|\right\} \in \left[\frac{\delta}{10^{100}}, 10^{100}\delta\right]$$

and every limit point of $\left(h_{n,\delta} \oplus w_{n,\delta}\right)$ lies in $\mathrm{Lie}(\mathrm{A}) \oplus \mathrm{Lie}(\mathrm{V})$.

See Sec.3.8 for the proof. By passing to a subsequence, let

$$\begin{aligned} y_{\infty,\delta} &= \lim y'_n := \lim \mathbf{u}_{s_{n,\delta}}\mathbf{u}_{t_{n,\delta}}.y_n; \\ x_{\infty,\delta} &= \lim x'_n := \lim \mathbf{u}_{t_{n,\delta}}.x. \end{aligned}$$

Also let $h_{\infty,\delta} \oplus w_{\infty,\delta}$ be a limit of $\left(h_{n,\delta} \oplus w_{n,\delta}\right)$. Write $g_\delta := \exp(h_{\infty,\delta})\exp(w_{\infty,\delta})$. Note that g_δ normalizes U.

As in Chapter 1, we arrive at

$$y_{\infty,\delta} = g_\delta.x_{\infty,\delta} \in \mathrm{Y}_1, \quad x_{\infty,\delta} \in \mathrm{Y}_1.$$

Hence

$$g_\delta\mathrm{Y}_1 = \overline{g_\delta\mathrm{U}.x_{\infty,\delta}} = \overline{\mathrm{U}.y_{\infty,\delta}} = \mathrm{Y}_1.$$

As

$$\left\{g \in \mathrm{G} \,\middle|\, g\mathrm{Y}_1 = \mathrm{Y}_1\right\}$$

is a closed subgroup, if we write $g_\delta = \exp v_\delta$ with $v_\delta \to 0$ in Lie(AV), then there exists some $v_{\neq 0} \in$ Lie(AV) such that

$$\exp(sv)\mathrm{Y}_1 = \mathrm{Y}_1, \quad \forall s \in \mathbb{R}.$$

If v has non-trivial Lie(V)-component then we are done. Otherwise we go back to case 2. Hence the proof completes.

3. Proof of the two lemmas

The reader is encouraged to prove Lem.2.2 and 2.3 on his/her own since the proof presented here has simple ideas but messy details.

Both $\mathfrak{h}$ and $\mathfrak{h}^{\perp}$ are invariant under the adjoint action of H, and hence can be considered separately. In matrix terms,

$$\mathrm{Ad}(g).M = gMg^{-1}, \quad \mathrm{ad}(X).M = XM - MX, \quad \exp(\mathrm{ad}(X)) = \mathrm{Ad}(\exp(X)).$$

3.1. Computation of the Lie algebra.

3.1.1. *Lie algebra of the orthogonal group.* By definition, writing $M_0 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$,

$$\mathfrak{so}_{Q_0} = \left\{X \in \mathfrak{sl}_3 \,\middle|\, M_0 X + X^{tr} M_0 = 0\right\}.$$

Write $X = (x_{ij})$, then we are solving

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix} + \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = 0$$

$$\Longleftrightarrow \begin{bmatrix} x_{31} & x_{32} & x_{33} \\ -x_{21} & -x_{22} & -x_{23} \\ x_{11} & x_{12} & x_{13} \end{bmatrix} + \begin{bmatrix} x_{31} & -x_{21} & x_{11} \\ x_{32} & -x_{22} & x_{12} \\ x_{33} & -x_{23} & x_{13} \end{bmatrix} = 0$$

$$\Longleftrightarrow x_{31} = x_{22} = x_{13} = 0,\ x_{32} = x_{21},\ x_{33} + x_{11} = 0,\ x_{23} = x_{12}.$$

That is to say

$$\mathfrak{so}_{Q_0} = \left\{ \begin{bmatrix} x_{11} & x_{12} & 0 \\ x_{21} & 0 & x_{12} \\ 0 & x_{21} & -x_{11} \end{bmatrix} \right\}.$$

3.1.2. *Computation of its complement.* The notation $\mathfrak{so}_{Q_0}^{\perp}$ below is justified by the fact that it is indeed the orthogonal complement of $\mathfrak{so}_{Q_0}$ in $\mathfrak{sl}_3$ with respect to the killing form (Exercise: check this).

$$\mathfrak{so}_{Q_0}^{\perp} = \left\{X \in \mathfrak{sl}_3 \,\middle|\, M_0 X - X^{tr} M_0 = 0\right\}.$$

Write $X = (x_{ij})$, then we are solving

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ x_{31} & x_{32} & x_{33} \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} - \begin{bmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = 0$$

$$\Longleftrightarrow \begin{bmatrix} x_{31} & x_{32} & x_{33} \\ -x_{21} & -x_{22} & -x_{23} \\ x_{11} & x_{12} & x_{13} \end{bmatrix} = \begin{bmatrix} x_{31} & -x_{21} & x_{11} \\ x_{32} & -x_{22} & x_{12} \\ x_{33} & -x_{23} & x_{13} \end{bmatrix}$$

$$\Longleftrightarrow x_{32} = -x_{21}, x_{11} = x_{33} \text{ and } x_{23} = -x_{12}.$$

That is to say

$$\mathfrak{so}_{Q_0}^{\perp} = \left\{ \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & -2x_{11} & -x_{12} \\ x_{31} & -x_{21} & x_{11} \end{bmatrix} \right\}.$$

3.2. Computation, conjugacy by unipotents. Take $w = (w_{ij}) \in \mathfrak{h}^\perp$, note that

$$\mathrm{Ad}(\mathbf{u}_s)\cdot w = \exp(s\,\mathrm{ad}(u_0))\cdot w = w + s\cdot \mathrm{ad}(u_0)\,w + \frac{s^2}{2}\,\mathrm{ad}(u_0)^2\,w + \frac{s^3}{3!}\,\mathrm{ad}(u_0)^3\,w + \frac{s^4}{4!}\,\mathrm{ad}(u_0)^4\,w$$

where the higher order terms vanish.

Write

$$\begin{aligned} w = &\begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & -2w_{11} & -w_{12} \\ w_{31} & -w_{21} & w_{11} \end{bmatrix} \\ = &w_{31}\cdot \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} + w_{21}\begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} + w_{11}\begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\ &+ \frac{-w_{12}}{3}\begin{bmatrix} 0 & -3 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} + \frac{w_{13}}{6}\begin{bmatrix} 0 & 0 & 6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

The reason why we write it in this form is that

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \xrightarrow{\mathrm{ad}\,u_0} \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \xrightarrow{\mathrm{ad}\,u_0} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \xrightarrow{\mathrm{ad}\,u_0} \begin{bmatrix} 0 & -3 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow{\mathrm{ad}\,u_0} \begin{bmatrix} 0 & 0 & 6 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Using this, one can compute that

$$\mathrm{Ad}(\mathbf{u}_t)\,w = \begin{bmatrix} \frac{t^2}{2}w_{31} + tw_{21} + w_{11} & \frac{t^3}{3!}w_{31} + \frac{t^2}{2}w_{21} + tw_{11} + \frac{-w_{12}}{3} & \frac{t^4}{4!}w_{31} + \frac{t^3}{3!}w_{21} + \frac{t^2}{2}w_{11} + t\frac{-w_{12}}{3} + \frac{w_{13}}{6} \\ tw_{31} + w_{21} & * & * \\ w_{31} & * & * \end{bmatrix} \tag{10}$$

where the terms marked as $*$ are determined by the others, since the matrix is an element in $\mathfrak{h}^\perp$.

3.3. Linear independence of characters. Intuitively, one sees that the upper right corner of Equa.(10) should dominate the rest. To turn this intuition into a solid statement is not so direct due to the possible cancellations between different monomials. By modifying the value of t, though, we can avoid this. For simplicity let

$$\delta_t := \max\left\{\left|\frac{t^4}{4!}w_{31}\right|, \left|\frac{t^3}{3!}w_{21}\right|, \left|\frac{t^2}{2}w_{11}\right|, \left|t\frac{-w_{12}}{3}\right|, \left|\frac{w_{13}}{6}\right|\right\}. \tag{11}$$

Let $\|\cdot\|_{\sup}$ denote the maximal value of the absolute values of entries of a matrix, then

$$\|\mathrm{Ad}(\mathbf{u}_t).w\|_{\sup} \le 5\delta. \tag{12}$$

For simplicity write

$$p_w(t) := \frac{t^4}{4!}w_{31} + \frac{t^3}{3!}w_{21} + \frac{t^2}{2}w_{11} + t\frac{-w_{12}}{3} + \frac{w_{13}}{6}.$$

LEMMA 3.1. *For $t \geq 1$, we have that*

$$\max\left\{\left|p_w(t)\right|, \left|p_w(2t)\right|, \left|p_w(3t)\right|, \left|p_w(4t)\right|, \left|p_w(5t)\right|\right\} \geq \frac{\delta_t}{10^{10}}.$$

To prove this lemma, consider the matrix

$$M_0 := \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2^4 & 2^3 & 2^2 & 2 & 1 \\ 3^4 & 3^3 & 3^2 & 3 & 1 \\ 4^4 & 4^3 & 4^2 & 4 & 1 \\ 5^4 & 5^3 & 5^2 & 5 & 1 \end{bmatrix}.$$

LEMMA 3.2. $\det(M_0) = 4!3!2! \neq 0$, *and coefficients of M_0^{-1} satisfy*

$$|(M_0^{-1})_{ij}| \leq \frac{4!5^4 4^3 3^2 2}{4!3!2!} \leq 10^9$$

for every i, j.

PROOF. M_0 is a Vandermonde matrix. Details left as an exercise. □

PROOF OF LEMMA 3.1.

$$\begin{bmatrix} p_w(t) \\ p_w(2t) \\ p_w(3t) \\ p_w(4t) \\ p_w(5t) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2^4 & 2^3 & 2^2 & 2 & 1 \\ 3^4 & 3^3 & 3^2 & 3 & 1 \\ 4^4 & 4^3 & 4^2 & 4 & 1 \\ 5^4 & 5^3 & 5^2 & 5 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{t^4}{4!} w_{31} \\ \frac{t^3}{3!} w_{21} \\ \frac{t^2}{2} w_{11} \\ \frac{-w_{12}}{3} \\ \frac{w_{13}}{6} \end{bmatrix}$$

And

$$\delta = \left\| \begin{bmatrix} \frac{t^4}{4!} w_{31} \\ \frac{t^3}{3!} w_{21} \\ \frac{t^2}{2} w_{11} \\ \frac{-w_{12}}{3} \\ \frac{w_{13}}{6} \end{bmatrix} \right\|_{\sup} = \left\| M_0^{-1} \cdot \begin{bmatrix} p_w(t) \\ p_w(2t) \\ p_w(3t) \\ p_w(4t) \\ p_w(5t) \end{bmatrix} \right\|_{\sup} \leq 5 \left\| M_0^{-1} \right\|_{\sup} \cdot \left\| \begin{bmatrix} p_w(t) \\ p_w(2t) \\ p_w(3t) \\ p_w(4t) \\ p_w(5t) \end{bmatrix} \right\|_{\sup}$$

Hence

$$\left\| \begin{bmatrix} p_w(t) \\ p_w(2t) \\ p_w(3t) \\ p_w(4t) \\ p_w(5t) \end{bmatrix} \right\|_{\sup} \geq \frac{\delta}{5 \left\| M_0^{-1} \right\|_{\sup}} \geq \frac{\delta}{10^{10}}.$$

□

3.4. Proof of Lemma 2.2. Let $w_{ij}(n)$ be the (i,j)-th coefficient of w_n. Let $\delta > 0$, for n large, we can find $t \in \mathbb{R}$ such that

$$\delta := \max\left\{\left|\frac{t^4}{4!} w_{31}(n)\right|, \left|\frac{t^3}{3!} w_{21}(n)\right|, \left|\frac{t^2}{2} w_{11}(n)\right|, \left|t\frac{-w_{12}(n)}{3}\right|, \left|\frac{w_{13}(n)}{6}\right|\right\}, \tag{13}$$

namely, Equa.(11) holds with $\delta_t = \delta$ and $w_{ij} = w_{ij}(n)$. Let $t_{n,\delta} \in \{t, 2t, ..., 5t\}$ such that the maximum in Lem.3.1 is attained. By Lem.3.1,

$$\left\| \mathrm{Ad}\,\mathbf{u}_{t_{n,\delta}}.w_n \right\|_{\sup} \geq \frac{\delta}{10^{10}}.$$

Also note that as $n \to \infty$, $t_{n,\delta}$ necessarily goes to $+\infty$. Equa.(12) says that

$$\left\| \mathrm{Ad}\,\mathbf{u}_{t_{n,\delta}}.w_n \right\|_{\sup} \leq 5\delta.$$

From Equa.(10), one sees that for $(i,j)\neq(1,3)$,

$$\left|\left(\mathrm{Ad}\,\mathbf{u}_{t_{n,\delta}}.w_n\right)_{i,j}\right|\leq\frac{4!\delta}{t_{n,\delta}},$$

which shows that as n goes to the infinity, only $\left(\mathrm{Ad}\,\mathbf{u}_{t_{n,\delta}}.w_n\right)_{1,3}$ survives. Now the proof is complete.

3.5. From SL2 to SO(Q). In this subsection, we provide an explicit morphism from $\mathrm{SL}_2(\mathbb{R})$ to $\mathrm{SO}_{Q_0}(\mathbb{R})$.

3.5.1. *$sl_2(\mathbb{R})$ as a quadratic space.* Note that $\mathrm{SL}_2(\mathbb{R})$ acts on

$$\mathfrak{sl}_2(\mathbb{R})=\{2\times 2\text{ trace zero matrices}\}$$

via the adjoint representation. And this action preserves the symmetric bilinear form

$$-\mathrm{Tr}:(X,Y)\mapsto-\mathrm{Tr}(X\cdot Y).$$

To identify $\mathfrak{sl}_2(\mathbb{R})$ with $\mathbb{R}^3$, consider the basis

$$E_1=\begin{bmatrix}0&1\\0&0\end{bmatrix},\quad E_2=\frac{1}{\sqrt{2}}\begin{bmatrix}1&0\\0&-1\end{bmatrix},\quad E_3=\begin{bmatrix}0&0\\-1&0\end{bmatrix}.$$

And we fix an isomorphism $\mathbb{R}^3\cong\mathfrak{sl}_2(\mathbb{R})$ by sending e_i to E_i where (e_1,e_2,e_3) is the standard basis of $\mathbb{R}^3$.

Then one can check that

$$\left(-\mathrm{Tr}(E_i\cdot E_j)\right)_{i,j}=\begin{bmatrix}0&0&1\\0&-1&0\\1&0&0\end{bmatrix},$$

which means that $-\mathrm{Tr}$ is identified with Q_0 under the fixed isomorphism.

3.5.2. *adjoint action of sl2 in basis.* Denote by $\rho:\mathrm{SL}_2(\mathbb{R})\to\mathrm{SO}_{Q_0}(\mathbb{R})$ the morphism obtained by the above identification of $\mathfrak{sl}_2(R)\cong\mathbb{R}^3$. Let us compute $\mathrm{d}\rho:\mathfrak{sl}_2(R)\to\mathfrak{so}_{Q_0}(\mathbb{R})$.

Let

$$X=\begin{bmatrix}a&b\\c&-a\end{bmatrix}\in\mathfrak{sl}_2(\mathbb{R}).$$

Then

$$\begin{aligned}\mathrm{ad}(X)E_1&=\begin{bmatrix}a&b\\c&-a\end{bmatrix}\begin{bmatrix}0&1\\0&0\end{bmatrix}-\begin{bmatrix}0&1\\0&0\end{bmatrix}\begin{bmatrix}a&b\\c&-a\end{bmatrix}\\&=\begin{bmatrix}0&a\\0&c\end{bmatrix}-\begin{bmatrix}c&-a\\0&0\end{bmatrix}=\begin{bmatrix}-c&2a\\0&c\end{bmatrix}=2aE_1+(-\sqrt{2}c)E_2+0E_3,\end{aligned}$$

$$\begin{aligned}\mathrm{ad}(X)E_2&=\frac{1}{\sqrt{2}}\begin{bmatrix}a&b\\c&-a\end{bmatrix}\begin{bmatrix}1&0\\0&-1\end{bmatrix}-\frac{1}{\sqrt{2}}\begin{bmatrix}1&0\\0&-1\end{bmatrix}\begin{bmatrix}a&b\\c&-a\end{bmatrix}\\&=\frac{1}{\sqrt{2}}\begin{bmatrix}a&-b\\c&a\end{bmatrix}-\frac{1}{\sqrt{2}}\begin{bmatrix}a&b\\-c&a\end{bmatrix}=\begin{bmatrix}0&-\sqrt{2}b\\\sqrt{2}c&0\end{bmatrix}=(-\sqrt{2}b)E_1+0E_2+(-\sqrt{2}c)E_3\end{aligned}$$

and

$$\begin{aligned}\mathrm{ad}(X)E_3&=\begin{bmatrix}a&b\\c&-a\end{bmatrix}\begin{bmatrix}0&0\\-1&0\end{bmatrix}-\begin{bmatrix}0&0\\-1&0\end{bmatrix}\begin{bmatrix}a&b\\c&-a\end{bmatrix}\\&=\begin{bmatrix}-b&0\\a&0\end{bmatrix}-\begin{bmatrix}0&0\\-a&-b\end{bmatrix}=\begin{bmatrix}-b&0\\2a&b\end{bmatrix}=0E_1+(-\sqrt{2}b)E_2+(-2a)E_3\end{aligned}$$

Hence we have that

$$\mathrm{d}\rho : \begin{bmatrix} a & b \\ c & -a \end{bmatrix} \mapsto \begin{bmatrix} 2a & -\sqrt{2}b & 0 \\ -\sqrt{2}c & 0 & -\sqrt{2}b \\ 0 & -\sqrt{2}c & -2a \end{bmatrix}.$$

Sanity check: RHS is indeed a matrix in $\mathfrak{so}_{Q_0}(\mathbb{R})$.

3.6. Image of a unipotent flow. Let

$$\mathbf{u}'_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}.$$

Then

$$\rho(\mathbf{u}'_s) = \exp\left(\mathrm{d}\rho \begin{bmatrix} 0 & s \\ 0 & 0 \end{bmatrix}\right) = \exp\left(\begin{bmatrix} 0 & -\sqrt{2}s & 0 \\ 0 & 0 & -\sqrt{2}s \\ 0 & 0 & 0 \end{bmatrix}\right) = \mathbf{u}_{-\sqrt{2}s}$$

3.7. Exponential of a lower triangular matrix. Say we have

$$\exp \begin{bmatrix} x & 0 \\ y & -x \end{bmatrix} = \begin{bmatrix} (1+a)^{-1} & 0 \\ b & (1+a) \end{bmatrix},$$

we would like to express x, y in terms of a, b.

Indeed, by definition of exp,

$$\text{LHS} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} x & 0 \\ y & -x \end{bmatrix} + \frac{1}{2}\begin{bmatrix} x & 0 \\ y & -x \end{bmatrix}^2 + \frac{1}{3!}\begin{bmatrix} x & 0 \\ y & -x \end{bmatrix}^3 + \dots$$

So we should compute the powers of this matrix first.

$$\begin{bmatrix} x & 0 \\ y & -x \end{bmatrix}^2 = \begin{bmatrix} x^2 & 0 \\ 0 & (-x)^2 \end{bmatrix} \implies \begin{bmatrix} x & 0 \\ y & -x \end{bmatrix}^{2n} = \begin{bmatrix} x^{2n} & 0 \\ 0 & (-x)^{2n} \end{bmatrix}.$$

And odd powers are

$$\begin{bmatrix} x & 0 \\ y & -x \end{bmatrix}^{2n+1} = \begin{bmatrix} x^{2n+1} & 0 \\ yx^{2n} & (-x)^{2n+1} \end{bmatrix}.$$

Thus

$$\exp \begin{bmatrix} x & 0 \\ y & -x \end{bmatrix} = \begin{bmatrix} e^x & 0 \\ y\left(\frac{e^x - e^{-x}}{2x}\right) & e^{-x} \end{bmatrix} = \begin{bmatrix} (1+a)^{-1} & 0 \\ b & (1+a) \end{bmatrix}.$$

And thus

$$x = \ln(1+a),\ y = b\left(\frac{2\ln(1+a)}{(1+a)-(1+a)^{-1}}\right). \tag{14}$$

The equality for y is not needed. Also note that for $|a| < 1$

$$|\ln(1+a) - a| \le 2|a|^2 \tag{15}$$

and that if (a_n) is a sequence contained in some fixed compact sub-interval of $(-1, +\infty)$ and (b_n) is a sequence converging to 0 then the corresponding (y_n) should converge to 0.

3.8. Proof of Lemma 2.3. For convenience, let us repeat Lem.2.3.

LEMMA 3.3. *For $\delta > 0$ small enough and n large enough, there exist $t_{n,\delta}$ and $s_{n,\delta}$ such that*

$$\mathbf{u}_{s_{n,\delta}} \cdot \mathbf{u}_{t_{n,\delta}} \exp(h_n)\exp(w_n)\mathbf{u}_{t_{n,\delta}}^{-1} = \exp(h_{n,\delta})\exp(w_{n,\delta}), \tag{16}$$

for some $h_{n,\delta} \in \mathfrak{h}$, $w_{n,\delta} \in \mathfrak{h}^{\perp}$ with

$$\max\{\|h_{n,\delta}\|, \|w_{n,\delta}\|\} \in [\frac{\delta}{10^{100}}, 10^{100}\delta] \tag{17}$$

and every limit point of $(h_{n,\delta} \oplus w_{n,\delta})$ lies in Lie(A) $\oplus$ Lie(V).

PROOF. Define $h'_n := \mathrm{d}\rho^{-1}(h_n)$. Write

$$\exp(h'_n) = \begin{bmatrix} 1+a_n & b_n \\ c_n & 1+d_n \end{bmatrix}.$$

For $s, t \in \mathbb{R}$, write $s' := s/(-\sqrt{2})$, $t' = t/(-\sqrt{2})$. Hence $\rho(\mathbf{u}'_{s'}) = \mathbf{u}_s$ and $\rho(\mathbf{u}'_{t'}) = \mathbf{u}_t$. Choose $s_{n,\delta}$ depending on $t_{n,\delta}$ (to be determined later) such that

$$\mathbf{u}'_{s'_{n,\delta}}\mathbf{u}'_{t'_{n,\delta}}\exp(h'_n)(\mathbf{u}'_{t'_{n,\delta}})^{-1} = \begin{bmatrix} (1+d_n - t'_{n,\delta}c_n)^{-1} & 0 \\ c_n & 1+d_n - t'_{n,\delta}c_n \end{bmatrix}. \tag{18}$$

See Chapter 1 for details. Define $h'_{n,\delta}$ by

$$\exp\left(h'_{n,\delta}\right) = \begin{bmatrix} (1+d_n - t'_{n,\delta}c_n)^{-1} & 0 \\ c_n & 1+d_n - t'_{n,\delta}c_n \end{bmatrix}. \tag{19}$$

Write $w_n = (w_{ij}(n))$. Choose t such that

$$\delta = \max\left\{|d_n - t'c_n|, \left|\frac{t^4}{4!}w_{31}(n)\right|, \left|\frac{t^3}{3!}w_{21}(n)\right|, \left|\frac{t^2}{2}w_{11}(n)\right|, \left|t\frac{-w_{12}(n)}{3}\right|, \left|\frac{w_{13}(n)}{6}\right|\right\}.$$

Also let

$$\delta' := \max\left\{\left|\frac{t^4}{4!}w_{31}(n)\right|, \left|\frac{t^3}{3!}w_{21}(n)\right|, \left|\frac{t^2}{2}w_{11}(n)\right|, \left|t\frac{-w_{12}(n)}{3}\right|, \left|\frac{w_{13}(n)}{6}\right|\right\}.$$

We choose $t_{n,\delta}$ from $t, 2t, ..., 5t$ such that the maximum in Lem.3.1 is attained (with δ_t replaced by δ').

Define $h_{n,\delta} := \mathrm{d}\rho(h'_{n,\delta})$ and

$$w_{n,\delta} := \mathrm{Ad}(\mathbf{u}_{t_{n,\delta}}).w_n.$$

Now everything is defined and it remains to check the conclusion of Lem.2.3.

First one can verify Equa.(16). By Equa.(18) and (19),

$$\begin{aligned} &\mathbf{u}'_{s'_{n,\delta}}\mathbf{u}'_{t'_{n,\delta}}\exp(h'_n)(\mathbf{u}'_{t'_{n,\delta}})^{-1} = \exp\left(h'_{n,\delta}\right) \\ (\text{apply } \rho) \implies &\mathbf{u}_{s_{n,\delta}}\mathbf{u}_{t_{n,\delta}}\exp(h_n)\mathbf{u}_{t_{n,\delta}}^{-1} = \exp\left(h_{n,\delta}\right) \\ \implies &\mathbf{u}_{s_{n,\delta}}\mathbf{u}_{t_{n,\delta}}\exp(h_n)\exp\left(w_n\right)\mathbf{u}_{t_{n,\delta}}^{-1} \\ &= \exp\left(h_{n,\delta}\right)\exp\left(\mathrm{Ad}(\mathbf{u}_{t_{n,\delta}}).w_n\right) = \exp\left(h_{n,\delta}\right)\exp\left(w_{n,\delta}\right). \end{aligned}$$

That $h_{n,\delta} \in \mathfrak{h}$, $w_{n,\delta} \in \mathfrak{h}^{\perp}$ follows from their definition. It remains to verify Equa.(17) and that every limit of $(h_{n,\delta})$ is in Lie(A), every limit of $(w_{n,\delta})$ is in Lie(V).

By the discussion below Equa.(15) and that $c_n \to 0$, we find

$$\left|\left(h'_{n,\delta}\right)_{2,1}\right| \to 0.$$

Thus every limit of $(h_{n,\delta})$ is in Lie(A). That every limit of $(w_{n,\delta})$ is in Lie(V) follows from the proof of Lem.2.2.

For n sufficiently large such that $|d_n| \le \delta$,

$$\left|d_n - t'_{n,\delta} c_n\right| \le |d_n| + 5|t' c_n| \le 6|d_n| + 5|d_n - t' c_n| \le 11\delta.$$

Combined with Equa.(15) we see that (assume $\delta \le 1$)

$$\left|\left(h'_{n,\delta}\right)_{2,2}\right| \le \left|d_n - t'_{n,\delta} c_n\right| + 2\left|d_n - t'_{n,\delta} c_n\right|^2 = 11\delta + 2\cdot 11^2\delta^2 \le 300\delta.$$

This shows that

$$\|h_{n,\delta}\| \le 10^{100}\delta.$$

If $\delta = \delta'$, then the remaining conclusions follow from the proof of Lem.2.2.

If $\delta > \delta'$, then $\delta = \left|d_n - t' c_n\right|$. For n large enough such that $|d_n| \le 0.1\delta$,

$$\left|d_n - t'_{n,\delta} c_n\right| \ge \left|t'_{n,\delta} c_n\right| - |d_n| \ge \left|t' c_n\right| - |d_n| \ge \left|d_n - t' c_n\right| - 2\,|d_n| \ge 0.8\delta.$$

Now by Equa.(15),

$$\left|\left(h'_{n,\delta}\right)_{2,2}\right| \ge \left|d_n - t'_{n,\delta} c_n\right| - 2\left|d_n - t'_{n,\delta} c_n\right|^2 \ge 0.8\delta - 2\cdot(11)^2\delta^2.$$

If δ is sufficiently small such that $2\cdot(11)^2\delta \le 0.1$, then

$$\left|\left(h'_{n,\delta}\right)_{2,2}\right| \ge 0.5\delta \implies \|h_{n,\delta}\| \ge 0.5\delta.$$

Again, the rest of the claim follows by arguments in Lem.2.2.

Now we are done. □

4. Exercises

4.1. orbits of diagonal groups. We say that a matrix $g \in \mathrm{SL}_2(\mathbb{R})$ is $\mathbb{R}$-*diagonalizable* iff there exists $h \in \mathrm{SL}_2(\mathbb{R})$ such that hgh^{-1} is a diagonal matrix. Note that for a matrix $X_{\neq \pm \mathrm{id}} \in \mathrm{SL}_2(\mathbb{R})$, being $\mathbb{R}$-*diagonalizable* is equivalent to being hyperbolic in the sense that $\mathrm{Tr}(X) > 2$. Fix a discrete subgroup Γ of $\mathrm{SL}_2(\mathbb{R})$, an $\mathbb{R}$-diagonalizable matrix $\gamma \in \Gamma$ is said to be *primitive* iff it can not be written as $(\gamma')^n$ for some $n \in \mathbb{Z}$, $n \neq \pm 1$ and some other $\gamma' \in \Gamma$ that is $\mathbb{R}$-diagonalizable. By definition $\pm\,\mathrm{id}$ is never primitive. Let

$$\mathrm{Prim}(\Gamma) := \left\{\gamma \text{ is } \mathbb{R}\text{-diagonalizable and primitive}\right\}.$$

EXERCISE 4.1. *Assume* $\Gamma \le \mathrm{SL}_2(\mathbb{R})$ *is a discrete subgroup containing* $\{\pm\,\mathrm{id}\}$. *Find a bijection between*

$$\left\{\textit{compact } \{a_t\}\textit{-orbits}\right\} \cong \mathrm{Prim}(\Gamma)/\sim_\Gamma$$

where $\sim_\Gamma$ *is the equivalence relation defined by* $g \sim_\Gamma h$ *iff* $g = \gamma h \gamma^{-1}$ *for some* $\gamma \in \Gamma$.

EXERCISE 4.2. *Classify all compact* $\{a_t\}_{t\in\mathbb{R}}$*-orbits on* $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$.

EXERCISE 4.3. *Classify all divergent* $\{a_t\}_{t\in\mathbb{R}}$*-orbits on* $\mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$.

Recall that an orbit $\{a_t.x\}$ is said to be divergent iff for every compact set in $C \subset \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$ there exists $t_0 > 0$ such that for all $|t| > t_0$, we have $a_t.x \notin C$.

CHAPTER 4

Nondivergence of unipotent flows on X_2

Back to the Top.
The main reference for this chapter is Kleinbock's Clay notes [Kle10].
Notation:

- $\mathrm{U} := \left\{ \mathbf{u}_s := \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle| \, s \in \mathbb{R} \right\}, \mathrm{A} := \left\{ \mathbf{a}_t := \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix} \middle| \, t \in \mathbb{R} \right\};$
-
- $\mathrm{X}_2 = \left\{ \text{unimodular lattices in } \mathbb{R}^2 \right\}$

1. Summary

DEFINITION 1.1. *For $\varepsilon > 0$, define*

$$\mathscr{C}_\varepsilon := \left\{ \Lambda \in \mathrm{X}_2 \,\middle|\, \mathrm{sys}(\Lambda) \geq \varepsilon \right\}.$$

By Lem.2.9 from Ch.2 (Mahler's criterion), $\mathscr{C}_\varepsilon$ is a compact set and every compact set in X_2 is contained in $\mathscr{C}_\varepsilon$ for some $\varepsilon > 0$.

THEOREM 1.2. *[Uniform non-divergence of unipotent flows for X_2] For every compact set $K \subset \mathrm{X}_2$ and $\varepsilon \in (0,1)$, there exists $\delta = \delta(K, \varepsilon) > 0$ such that the following holds. For every interval (a,b) with $a < b$ in $\mathbb{R}$ and $\Lambda \in \mathrm{X}_2$ satisfying $\mathbf{u}_{s_0}.\Lambda \in K$ for some $s_0 \in (a,b)$, we have that*

$$\frac{1}{b-a} \mathrm{Leb}\{s \in (a,b) \,|\, \mathbf{u}_s.\Lambda \notin \mathscr{C}_\delta\} \leq \varepsilon.$$

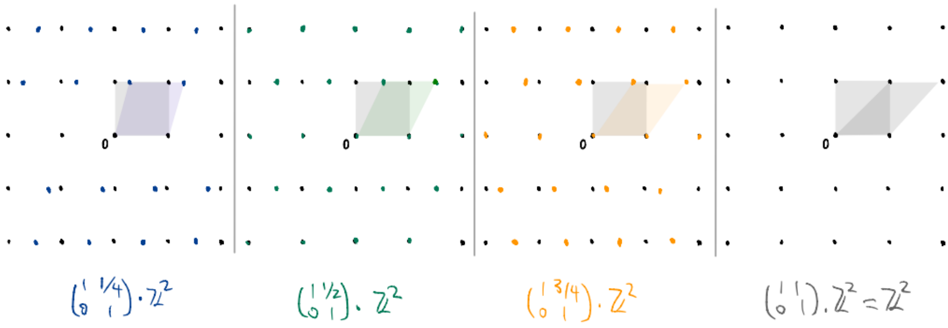
Actually the choice of δ is also independent of the unipotent flow we use — you may replace $\mathbf{u}_s$ everywhere by its conjugates.

THEOREM 1.3. *If $\varepsilon \leq 1$ and $\Lambda \in \mathrm{X}_2$ are such that $\mathbf{u}_s.\Lambda \notin \mathscr{C}_\varepsilon$ for every s in some interval of infinite length (i.e., something like $(a,+\infty), (-\infty, b), (-\infty, +\infty)$), then Λ contains a horizontal vector of length less than ε. That is to say, $(v_1, 0) \in \Lambda$ for some $0 < |v_1| < \varepsilon$.*

COROLLARY 1.4. *For every two x_1, x_2 with compact U-orbits, there exists $u \in \mathrm{U}$ and $a \in \mathrm{A}$ such that $x_2 = au.x_1$.*

The reader might have noticed that the converse also holds since U-action fixes the horizontal direction. Also note that such U-orbits are closed and compact. In this case, one may think of U-action on Λ as "Dehn-twist" along the closed geodesic represented by $(v_1, 0) \in \Lambda \cong \pi_1(\mathbb{R}^2/\Lambda)$.

Here is a compact U-orbit $\mathrm{U}.\mathbb{Z}^2$:



2. The proof

LEMMA 2.1. *There exist $C_1 > 0$ and $\alpha_1 > 0$ such that for every interval (a,b) in $\mathbb{R}$, every $\mathbf{v} = (v_1, v_2) \in \mathbb{R}^2$ and every $\rho \in (0,1)$, we have*

$$\frac{1}{b-a}\operatorname{Leb}\left\{s \in (a,b) \,\middle|\, \|\mathbf{u}_s.v\| < \rho M_0\right\} \le C_1\rho^{\alpha_1}.$$

where $M_0 := \sup_{s\in(a,b)} \|\mathbf{u}_s.v\|$.

PROOF. Take $C_1 = 2\sqrt{2}$ and $\alpha_1 = 1$.

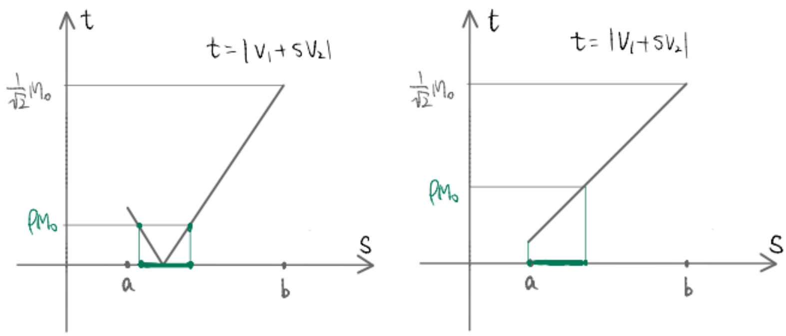
Note $\mathbf{u}_s.(v_1, v_2) = (v_1 + sv_2, v_2)$.

If $|v_2| \ge \frac{1}{\sqrt{2}}M_0$ then for every $s \in (a,b)$, $\|\mathbf{u}_s.v\| \ge |v_2| \ge \frac{1}{\sqrt{2}}M_0$. So if $\rho \le \frac{1}{\sqrt{2}}$, then we are already done. Otherwise, $C_1\rho^{\alpha_1} \ge 1$. Also ok.

So now we are left with the case when $|v_1 + s_0 v_2| \ge \frac{1}{\sqrt{2}}M_0$ for some $s_0 \in (a,b)$. Refer to the picture below, we see that

$$\frac{1}{b-a}\operatorname{Leb}\left\{s \in (a,b) \,\middle|\, |v_1 + sv_2| < \rho M_0\right\} \le 2\frac{\rho}{1/\sqrt{2}} = C_1\rho.$$

It remains only to note that $|v_1 + sv_2| < \|\mathbf{u}_s.(v_1, v_2)\|$. $\square$

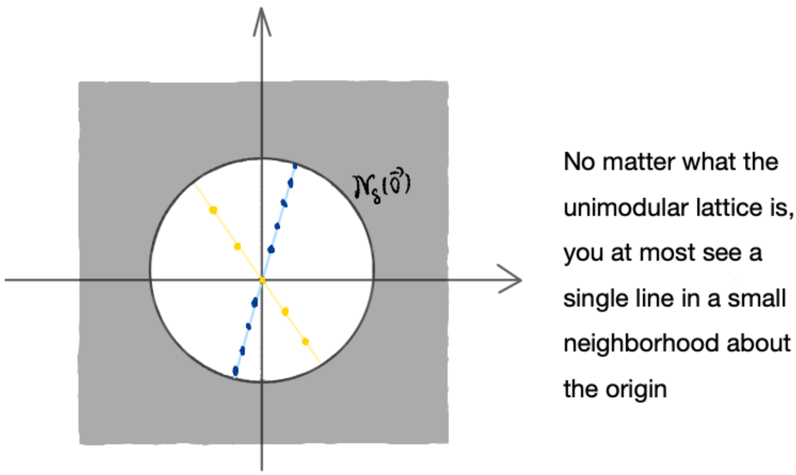


To go from this Lemma to Thm.1.2 or 1.3, we need

Key observation. A rank 2 unimodular lattice $\Lambda \in X_2$ is not allowed to contain two linearly independent vector of length strictly smaller than 1. For otherwise, if v, w is such a pair,

$$\|\Lambda\| \le \|\mathbb{Z}v \oplus \mathbb{Z}w\| \le \|v\|\,\|w\| < 1,$$

contradicting against the assumption that Λ is unimodular.



Let

$$\mathrm{Prim}(\Lambda) := \left\{ v_{\neq 0} \in \Lambda \,\middle|\, \mathbb{R}.v \cap \Lambda = \mathbb{Z}.v \right\}$$

be the set of primitive vectors.

PROOF OF THM.1.2. Find $\delta_0 \in (0,1)$ such that $K \subset \mathscr{C}_{\delta_0}$. We shall determine δ later, depending on δ_0 and ε.

Take $s_0 \in (a,b)$ such that $\mathbf{u}_{s_0}.\Lambda_0 \in K \subset \mathscr{C}_{\delta_0}$. Let

$$I(\Lambda_0, \delta_0) := \left\{ s \in (a,b) \,\middle|\, \mathrm{sys}(\mathbf{u}_s.\Lambda_0) < \delta_0 \right\}$$

which decomposes into a disjoint union of open intervals

$$I(\Lambda_0, \varepsilon_0) = \bigsqcup_{\alpha \in \mathscr{A}} I_\alpha$$

with certain index set $\mathscr{A}$.

Take one $I_\alpha = (x_\alpha, y_\alpha)$. By the remark right before the proof, for $s \in I_\alpha$, there exists a unique v_s (up to ± 1) in $\mathrm{Prim}(\Lambda_0)$ with

$$\|\mathbf{u}_s.v_s\| < \delta_0.$$

By connectedness, this v_s has to be independent of $s \in I_\alpha$. For this reason denote it by v_α. By Lem.3.1 and the assumption that $\mathbf{u}_{s_0}.\Lambda_0 \in \mathscr{C}_{\delta_0}$,

$$\frac{1}{|I_\alpha|} \mathrm{Leb}\left\{ s \in I_\alpha \,\middle|\, \|\mathbf{u}_s.v_\alpha\| < \rho\delta_0 \right\} < C_1 \rho^{\alpha_1}.$$

We take $\rho = \rho(\varepsilon)$ such that $C_1 \rho^{\alpha_1} < \varepsilon$. Let $\delta := \rho\delta_0$.

$$\{ s \in (a,b) \mid \|\mathbf{u}_s.v_\alpha\| < \delta \} = \bigsqcup_{\alpha \in \mathscr{A}} \left\{ s \in I_\alpha \,\middle|\, \|\mathbf{u}_s.v_\alpha\| < \rho\delta_0 \right\}$$

implying

$$\mathrm{Leb}\{ s \in (a,b) \mid \|\mathbf{u}_s.v_\alpha\| < \delta \} = \sum_{\alpha \in \mathscr{A}} \mathrm{Leb}\left\{ s \in I_\alpha \,\middle|\, \|\mathbf{u}_s.v_\alpha\| < \rho\delta_0 \right\} < \sum_{\alpha \in \mathscr{A}} |I_\alpha| \cdot \varepsilon \leq (b-a)\varepsilon.$$

$\square$

PROOF OF LEM.1.3 . Let I be this infinite interval. Since for each $s \in I$ there exists a unique (up to ± 1) v_s in $\mathrm{Prim}(\Lambda)$ with $\|\mathbf{u}_s.v_s\| < 1$. By connectedness argument, this $v = v_s$ is independent of $s \in I$. Thus $\|\mathbf{u}_s.v\| < 1$ for all $s \in I$. But coordinates of $\mathbf{u}_s.v$ are polynomials in s, being bounded then implies that $\mathbf{u}_s.v$ is constant. Therefore U fixes v and we are done. □

3. Exercises

3.1. Nondivergence in rank 1, a number field example. In these set of exercises, it is more convenient to write $\mathbb{R}^4$ as $\mathbb{R}^2 \oplus \mathbb{R}^2$.

EXERCISE 3.1. *Show that $\mathbb{Z}[\sqrt{2}]$ is a principal ideal domain.*

Thus every torsion free (finitely generated) $\mathbb{Z}[\sqrt{2}]$-module is free.

Fix an embedding of $\mathbb{Q}(\sqrt{2})$ in $\mathbb{R}$. Let σ be the other embedding of $\mathbb{Q}(\sqrt{2})$ in $\mathbb{R}$. Consider the action of $\mathbb{Q}(\sqrt{2})$ on $\mathbb{R}^2 \oplus \mathbb{R}^2$ given by

$$x.(v,w) := (x.v, \sigma(x).w).$$

EXERCISE 3.2. *This is a linear action. Write down the matrix representation of this action. Namely, for every $x = a + b\sqrt{2} \in \mathbb{Q}(\sqrt{2})$, write down a 4-by-4 matrix representing the action of x on $\mathbb{R}^2 \oplus \mathbb{R}^2$ with respect to the standard basis.*

Let Δ be a rank-1 $\mathbb{Z}[\sqrt{2}]$-submodule in $\mathbb{R}^2 \oplus \mathbb{R}^2$. We may write $\Delta = \mathbb{Z}[\sqrt{2}].(v,w)$. Let $\|\Delta\| := \|v\| \cdot \|w\|$.

EXERCISE 3.3. *Show that $\|\Delta\|$ is independent of the choice of generator for the $\mathbb{Z}[\sqrt{2}]$-module Δ.*

Define

$$\mathrm{X}_4'(\mathbb{Z}[\sqrt{2}]) := \left\{ \Lambda \leq \mathbb{R}^2 \oplus \mathbb{R}^2 \text{ lattice }, \Lambda \text{ is preserved by } \mathbb{Z}[\sqrt{2}] \right\}.$$

EXERCISE 3.4. *Show that such a lattice is a rank-2 $\mathbb{Z}[\sqrt{2}]$-module.*

Thus for $\Lambda \in \mathrm{X}_4'(\mathbb{Z}[\sqrt{2}])$, we can find a $\mathbb{Z}[\sqrt{2}]$-basis (v_1, w_1) and (v_2, w_2) in $\mathbb{R}^2 \oplus \mathbb{R}^2$. Define $\|\Lambda\| := \|v_1 \wedge v_2\| \cdot \|w_1 \wedge w_2\|$. Define $\det(\Lambda) := (v_1 \wedge v_2, w_1 \wedge w_2) \in (\mathbb{R} \oplus \mathbb{R})/\mathbb{Z}[\sqrt{2}]^\times$. Here $\mathbb{Z}[\sqrt{2}]^\times$ denotes the invertible elements in this ring $\mathbb{Z}[\sqrt{2}]$.

EXERCISE 3.5. *Show that indeed, the value of $\det(\Lambda)$ in $(\mathbb{R} \oplus \mathbb{R})/\mathbb{Z}[\sqrt{2}]^\times$ is independent of the choice of bases. Thus $\|\Lambda\|$ is also independent of the choice of bases.*

EXERCISE 3.6. *Find the relation between this newly defined $\|\Lambda\|$ and the old $\|\Lambda\|_{\mathrm{Old}}$ defined as the volume of $\mathbb{R}^4/\Lambda$.*

Define

$$\mathrm{X}_4(\mathbb{Z}[\sqrt{2}]) := \left\{ \Lambda \in \mathrm{X}_4'(\mathbb{Z}[\sqrt{2}]) \,\middle|\, \det \Lambda = 1 \right\}.$$

Here "1" is the image of $(1,1)$ in $(\mathbb{R} \oplus \mathbb{R})/\mathbb{Z}[\sqrt{2}]^\times$. Equip $\mathrm{X}_4(\mathbb{Z}[\sqrt{2}])$ with the Chabauty topology, viewing it as a collection of closed subgroups of $\mathbb{R}^2 \oplus \mathbb{R}^2$.

EXERCISE 3.7. *Show that the free $\mathbb{Z}[\sqrt{2}]$-module with basis $\{(e_1,e_1),(e_2,e_2)\}$ (denote this module as Λ_0) belongs to $\mathrm{X}_4(\mathbb{Z}[\sqrt{2}])$ and that $g \mapsto g.\Lambda_0$ induces a homeomorphism*

$$\mathrm{SL}_2(\mathbb{R}) \times \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z}[\sqrt{2}]) \cong \mathrm{X}_4(\mathbb{Z}[\sqrt{2}]).$$

For $\Lambda \in \mathrm{X}_4(\mathbb{Z}[\sqrt{2}])$, define

$$\mathrm{sys}_{\mathbb{Z}[\sqrt{2}]}(\Lambda) := \inf_{\Delta \leq \Lambda} \|\Delta\|$$

where Δ varies over all rank-1 $\mathbb{Z}[\sqrt{2}]$-submodule of Λ. For every $\varepsilon > 0$, let

$$\mathscr{C}_\varepsilon := \left\{ \Lambda \in \mathrm{X}_4(\mathbb{Z}[\sqrt{2}]) \,\middle|\, \mathrm{sys}_{\mathbb{Z}[\sqrt{2}]}(\Lambda) \geq \varepsilon \right\}.$$

EXERCISE 3.8. *For every* $\varepsilon > 0$, $\mathscr{C}_\varepsilon$ *is a compact subset of* $\mathrm{X}_4(\mathbb{Z}[\sqrt{2}])$.

EXERCISE 3.9. *Conversely, every compact subset of* $\mathrm{X}_4(\mathbb{Z}[\sqrt{2}])$ *is contained in* $\mathscr{C}_\varepsilon$ *for some* $\varepsilon > 0$.

EXERCISE 3.10. *For* $\varepsilon > 0$ *small enough, for every* $\Lambda \in \mathrm{X}_4(\mathbb{Z}[\sqrt{2}])$, *the set*

$$\{(v, w) \in \Lambda \mid \|v\| \, \|w\| < \varepsilon\}$$

is either $\{0\}$ *or generates a rank*-1 $\mathbb{Z}[\sqrt{2}]$*-submodule of* Λ.

Let $\mathbf{u}_t := \left(\begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \right)$ and $U := \{\mathbf{u}_t,\ t \in \mathbb{R}\}$.

EXERCISE 3.11. *Prove the following. For every* $\varepsilon > 0$, *there exists* $\delta > 0$ *such that for every* $\Lambda \in \mathrm{X}_4(\mathbb{Z}[\sqrt{2}])$,

- *either* Λ *contains a* $\mathbb{Z}[\sqrt{2}]$*-submodule preserved by* U *with norm smaller than* ε,
- *or*

$$\limsup_{T \to +\infty} \frac{1}{T} \operatorname{Leb}\{t \in [0, T] \mid \mathbf{u}_t.\Lambda \notin \mathscr{C}_\delta\} \leq \varepsilon.$$

CHAPTER 5

Nondivergence on X_3 and the strong form of Oppenheim conjecture

Back to the Top.

Notations:

- $\mathrm{H} := \mathrm{SO}_{Q_0}(\mathbb{R})$ with $Q_0(x_1, x_2, x_3) := 2x_1x_3 - x_2^2$;
- $\mathrm{X}_3 := \{$ unimodular lattices in $\mathbb{R}^3\}$;
- $\mathbf{u}_s := \begin{bmatrix} 1 & s & \frac{s^2}{2} \\ 0 & 1 & s \\ 0 & 0 & 1 \end{bmatrix} = \exp\left(s \cdot \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \right)$, $\mathrm{U} := \{\mathbf{u}_s : s \in \mathbb{R}\} \subset \mathrm{H}$;
- $\mathbf{v}_s := \begin{bmatrix} 1 & 0 & s \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \exp\left(s \cdot \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right)$, $\mathrm{V} := \{\mathbf{v}_s : s \in \mathbb{R}\} \not\subseteq \mathrm{H}$;
- $\mathbf{a}_t := \begin{bmatrix} e^t & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-t} \end{bmatrix} = \exp\left(t \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \right)$, $\mathrm{A} := \{\mathbf{a}_t : t \in \mathbb{R}\} \subset \mathrm{H}$;
- $\mathrm{B} := \{\mathbf{a}_t\mathbf{u}_s : s, t \in \mathbb{R}\} \subset \mathrm{H}$.

1. Summary

Finally, in this section we prove the strong form of Oppenheim conjecture. The general case can be reduced to the case of three variables, which we now state

THEOREM 1.1. *Let Q be a non-degenerate indefinite ternary real quadratic form that is not proportional to a rational quadratic form. Then $Q(\mathbb{Z}^3)$ is dense in $\mathbb{R}$. Actually $Q(\mathrm{Prim}(\mathbb{Z}^3))$ is dense in $\mathbb{R}$.*

THEOREM 1.2. *For every non-closed orbit of H on X_3, its closure contains a $\{\mathbf{v}_s\}_{s\geq 0}$ or $\{\mathbf{v}_s\}_{s\leq 0}$-orbit.*

A stronger statement will be proved later. See Ch.12, Thm.1.1.

By similar arguments presented in Chapter 2, Thm.1.1 would follow from Thm.1.2 and the following (whose proof is left as an exercise):

THEOREM 1.3. *If an H-orbit is closed, then the stabilizer in H is discrete and of finite covolume in H. Also the corresponding quadratic form is a scalar multiple of some rational quadratic form.*

To promote the weak version to this one the following non-divergence theorem is needed.

THEOREM 1.4. *For every $\varepsilon > 0$, there exists a compact subset $\mathscr{C}$ of X_3 such that for every $\Lambda \in \mathrm{X}_3$, at least one of the followings is true*

1. *The portion of time for $\mathbf{u}_s.\Lambda$ to spend outside $\mathscr{C}$ is smaller than ε, i.e.,*

$$\limsup_{T\to+\infty} \frac{1}{T} |\{s \in [0, T] \mid \mathbf{u}_s.\Lambda \notin \mathscr{C}\}| \leq \varepsilon;$$

2. $\Lambda \cap \{(x,0,0),\, x \in \mathbb{R}\}$ *contains a non-zero vector of length smaller than* ε*;*
3. $\Lambda \cap \{(x,y,0),\, x,y \in \mathbb{R}\}$ *contains a lattice (of* $\mathbb{R}e_1 \oplus \mathbb{R}e_2$*) of covolume smaller than* ε*.*

COROLLARY 1.5. *Let* $\varepsilon \in (0,1)$ *and pick* $\mathscr{C}$ *as in the above theorem. Then every orbit of* B *intersects non-trivially with* $\mathscr{C}$*.*

Finally let us make a convenient definition. Let $e_1 := (1,0,0)$ and $e_2 = (0,1,0)$.

DEFINITION 1.6. *We say that* $\mathbb{R}.e_1$ *is* Λ*-rational iff* $\Lambda \cap \mathbb{R}e_1$ *is a lattice in* $\mathbb{R}e_1$*, and* $\mathbb{R}e_1 \oplus \mathbb{R}e_2$ *is* Λ*-rational iff* $\Lambda \cap \mathbb{R}e_1 \oplus \mathbb{R}e_2$ *is a lattice in* $\mathbb{R}e_1 \oplus \mathbb{R}e_2$*. In either of these two cases, we say that the orbit* $\mathrm{U}.\Lambda$ ***degenerates****.*

This notion is justified by the fact that in these cases the orbit is essentially contained in certain (embedded) $\mathrm{SL}_2(\mathbb{R}) \ltimes \mathbb{R}^2 / \mathrm{SL}_2(\mathbb{Z}) \ltimes \mathbb{Z}^2$, which is interpreted as the space of lattices of lower rank with a fixed volume together with a marked point in the quotient torus.

2. Proof of the theorem

Now let us prove Thm.1.2. Start with Λ_0 with $\mathrm{H}.\Lambda_0$ non-closed. Let $\mathrm{Y}_0 := \overline{\mathrm{H}.\Lambda_0}$. Define $\mathscr{O}$ as in Chapter 3, the union of all H-orbits in Y_0 that is open in Y_0. Note that $\mathscr{O} \neq \mathrm{Y}_0$.

The old argument takes care of the case when $\mathrm{Y}_0 \setminus \mathscr{O}$ contains no degenerate U-orbits.

Indeed under this assumption every U-orbit in $\mathrm{Y}_0 \setminus \mathscr{O}$, by Thm.1.4, intersects with some fixed compact set non-trivially. Hence we can find a nonempty U-minimal set Y_1 in $\mathrm{Y}_0 \setminus \mathscr{O}$. As in Ch.3, there are two cases:

1. Y_1 is A-stable, we consider $\mathrm{Map}(\mathrm{Y}_1, \mathrm{Y}_0) := \{g\mathrm{Y}_1 \subset \mathrm{Y}_0\}$;
2. Y_1 is not A-stable, we consider $\mathrm{Aut}(\mathrm{Y}_1) := \{g\mathrm{Y}_1 = \mathrm{Y}_1\}$.

The arguments in Ch.3 should go quite smoothly here. In case 1, you may need to do a further perturbation to guarantee the sequence you get has a convergent subsequence.

2.1. New story, general assumption. However, it is unavoidable that $\mathrm{Y}_0 \setminus \mathscr{O}$ may contain some degenerate U-orbit. Let us take a nonempty B minimal set $\mathrm{Y}_1 \subset \mathrm{Y}_0 \setminus \mathscr{O}$ whose existence is guaranteed by the nondivergence corollary Coro.1.5. Take some $\Lambda_1 \in \mathrm{Y}_1$ such that $\mathrm{U}.\Lambda_1$ degenerates. We will assume $\mathbb{R}e_1 \oplus \mathbb{R}e_2$ is Λ_1-rational and leave the other case when $\mathbb{R}e_1$ is Λ_1-rational to the reader.

2.2. Case 1, no closed U-orbits. Assume Y_1 contains no closed U-orbit.

As we assumed, $\mathrm{U}.\Lambda_1$ is stuck in the following closed set (for simplicity write $\mathbb{R}e_{1,2} := \mathbb{R}e_1 \oplus \mathbb{R}e_2$)

$$\mathrm{X}_3(\mathbb{R}e_{1,2}, c_1) := \left\{\Lambda \in \mathrm{X}_3 \,\middle|\, \mathbb{R}e_{1,2} \text{ is } \Lambda\text{-rational, } \left\|\Lambda \cap \mathbb{R}e_{1,2}\right\| = c_1\right\}$$

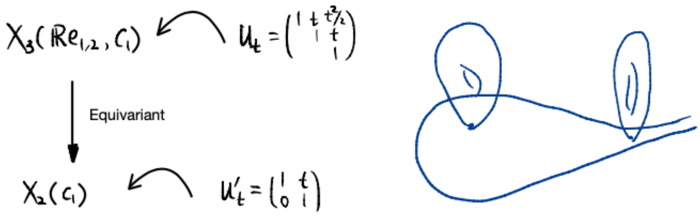
where $c_1 := \left\|\Lambda_1 \cap \mathbb{R}e_{1,2}\right\|$. Also let

$$\mathrm{X}_2(c_1) := \left\{ \text{lattices in } \mathbb{R}^2 \text{ of covolume } c_1 \right\}.$$

Then we have a natural continuous surjection $\pi : \mathrm{X}_3(\mathbb{R}e_{1,2}, c_1) \to \mathrm{X}_2(c_1)$ with compact fibres that is equivariant with respect to

$$\begin{aligned} \rho_\pi : \{g \in \mathrm{SL}_3(\mathbb{R}),\, g \text{ preserves } \mathbb{R}e_{1,2},\, \det(g|_{\mathbb{R}e_{1,2}}) = 1\} &\to \mathrm{SL}_2(\mathbb{R}) \\ g &\mapsto g|_{\mathbb{R}e_{1,2}}. \end{aligned}$$

In particular we have



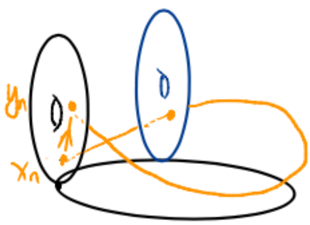
Now we wish to find a U-minimal set in Y_1.

2.3. Case 1.1, some $\pi(U.\Lambda_2)$ is compact. Assume for some $\Lambda_2 \in Y_1$, $\pi(U.\Lambda_2)$ is closed and hence compact.

Then $\overline{U.\Lambda_2}$ is compact (since π is a proper map) and let $\overline{U.\Lambda_3}$ be a nonempty minimal U-set in $Y_2 := \overline{U.\Lambda_2}$.

Then we can find pairs (x_n, y_n) in Y_2 such that $y_n = \exp(w_n)x_n$ with

- $w_n = \begin{bmatrix} 0 & 0 & * \\ 0 & 0 & * \\ 0 & 0 & 0 \end{bmatrix}$, $w_n \neq 0$, $w_n \to 0$.



First assume $w_n \in \mathrm{Lie}(V)$ for infinitely many n, then $\exp(w_n) \in \mathrm{Aut}(Y_2) := \{g \in SL_3(\mathbb{R}),\ gY_2 = Y_2\}$ and since the latter is a closed subgroup, we have the full $V \subset \mathrm{Aut}(Y_2)$.

Otherwise w_n is not fixed by $\mathrm{Ad}(U)$ and for any $\delta > 0$ and for n large enough we can find $t_{n,\delta}$ such that

- $\|\mathrm{Ad}(\mathbf{u}_{t_{n,\delta}}).w_n\| \asymp \delta$ (i.e. for some constant $C > 1$, LHS belongs to $(C^{-1}\delta, C\delta)$);
- every limit of $\big(\mathrm{Ad}(\mathbf{u}_{t_{n,\delta}}).w_n\big)$ is in $\mathrm{Lie}(V)$.

And by taking a limit we find

- $x_{\infty,\delta}, y_{\infty,\delta} \in Y_2$ and $w_{\infty,\delta} \in \mathrm{Lie}(V)$ such that $y_{\infty,\delta} = \exp(w_{\infty,\delta})x_{\infty,\delta}$;
- $w_{\infty,\delta} \neq 0$, $w_{\infty,\delta} \to 0$ as $\delta \to 0$.

Arguing as above, we have $V \subset \mathrm{Aut}(Y_2)$.

2.4. Case 1.2, $\pi(U.\Lambda)$ is never compact. Assume for every $\Lambda \in Y_1$, $\pi(U.\Lambda)$ is not compact.

Then there is some compact set such that every $U.\Lambda$ intersects non-trivially for every $\Lambda \in Y_1$. Therefore there is a nonempty U-minimal set in Y_1 and the rest of the proof is not so different from Sec.2.3.

2.5. Case 2, exists a closed U-orbit. Assume Y_1 contains a closed U-orbit $U.\Lambda_2$.

2.6. Case 2.1, recurrence in non-centralizer direction. Assume that there exists a sequence $(y_n) \subset Y_0$ converging to Λ_2 such that

- $y_n = \exp(w_n)\Lambda_2$ with $w_n \in \mathfrak{h}^\perp$, $w_n \notin \mathrm{Lie}(V)$.

[Recall from Ch.3, $\mathfrak{h}^{\perp}$ is a complement of $\mathfrak{h} = \mathfrak{so}_{Q_0}(\mathbb{R})$ in $\mathfrak{sl}_3(\mathbb{R})$ that is invariant under the adjoint action of $\mathrm{SO}_{Q_0}(\mathbb{R})$]

Without loss of generality assume $\| w_n \| < 1$ for all n. Let

$$t_n := \inf\{t \geq 0 \mid \|\mathrm{Ad}\,\mathbf{u}_t.w_n\| = 1\}.$$

Let $\varphi_n : [0, t_n] \to \left\{ w \in \mathfrak{h}^{\perp} \;\middle|\; \| w \| \leq 1 \right\}$ defined by

$$\varphi_n(t) := \mathrm{Ad}(\mathbf{u}_t).w_n.$$

By passing to a subsequence, assume $\varphi_n([0, t_n])$ converges (in Chabauty topology, or equivalently, w.r.t. Hausdorff distance). Let E_∞ denote the limit set. It is connected and closed.

[Side remark: We want something more than the perturbation method as in Ch.3 could possibly provide.]

2.6.1. *Lucky case.* Assume there is some $\delta_0 > 0$ such that

$$E_\infty \cap \{\| w \| \leq \delta_0\} \subset \mathrm{Lie}(\mathrm{V}).$$

For n large enough s.t. $\| w_n \| < \delta_0$, we define

$$t_n(\delta_0) := \inf\{t \geq 0 \mid \|\mathrm{Ad}\,\mathbf{u}_t.w_n\| = \delta_0\}.$$

By passing to a further subsequence we assume $\varphi_n([0, t_n(\delta_0)])$ converges to $E_\infty(\delta_0)$. By our assumption, $E_\infty(\delta_0)$ is a connected subset of $\{w \in \mathrm{Lie}(\mathrm{V}),\ \| w \| \leq \delta_0\}$. So it is an interval. Since U commutes with V, for every $w \in E_\infty$, $\exp(w)\mathrm{U}.\Lambda_2 \subset \mathrm{Y}_0$. Therefore,

$$\mathbf{v}_{[0,\delta_0]}.\Lambda_2 \subset \mathrm{Y}_0, \text{ or } \mathbf{v}_{[-\delta_0,0]}.\Lambda_2 \subset \mathrm{Y}_0.$$

W.L.O.G, assume $\mathbf{v}_{[0,\delta_0]}.\Lambda_2 \subset \mathrm{Y}_0$. Hence for every $t, s \in \mathbb{R}$,

$$\mathbf{v}_{[0,e^{2t}\delta_0]}\mathbf{a}_t\mathbf{u}_s.\Lambda_2 = \mathbf{a}_t\mathbf{u}_s\mathbf{v}_{[0,\delta_0]}.\Lambda_2 \subset \mathbf{a}_t\mathbf{u}_s\mathrm{Y}_0 = \mathrm{Y}_0. \tag{20}$$

By Thm.1.4, there exists a compact set such that for every $t > 0$, there exists $s_t > 0$ such that $\mathbf{a}_t\mathbf{u}_{s_t}.\Lambda_2$ lives in this compact set. In particular we may select $t_n \to +\infty$ and $s_n \in \mathbb{R}$ such that $\lim a_{t_n} s_n.\Lambda_2$ exists and call it Λ_∞. Then by Equa.(20) and a continuity argument, we have

$$\mathbf{v}_{[0,+\infty)}.\Lambda_\infty \subset \mathrm{Y}_0.$$

So we are done.

2.6.2. *Unlucky, try again!* If the assumption in Sec.2.6.1 does not hold, then we can repeat what is done above the Sec.2.6.1. So we get some $E_\infty^{(2)}$. If lucky, then we go back to Sec.2.6.1. If not, then we can repeat this process again to get $E_\infty^{(3)}$. It suffices to note that this process should stop.

Indeed recall the computation we made in Ch.3, Equa.(10)

$$\mathrm{Ad}(\mathbf{u}_t)\,w =$$

$$\begin{bmatrix} \frac{t^2}{2} w_{31} + t w_{21} + w_{11} & \frac{t^3}{3!} w_{31} + \frac{t^2}{2} w_{21} + t w_{11} + \frac{-w_{12}}{3} & \frac{t^4}{4!} w_{31} + \frac{t^3}{3!} w_{21} + \frac{t^2}{2} w_{11} + t\frac{-w_{12}}{3} + \frac{w_{13}}{6} \\ t w_{31} + w_{21} & * & * \\ w_{31} & * & * \end{bmatrix}. \tag{21}$$

If this process would continue, from the computation we sees right away that for $w_\infty^{(i)} \in E_\infty^{(i)}$, $w_\infty \in E_\infty$,

$$\begin{aligned} (w_\infty)_{3,1} &= 0; \\ \big(w_\infty^{(2)}\big)_{3,1} &= \big(w_\infty^{(2)}\big)_{2,1} = 0; \\ \big(w_\infty^{(3)}\big)_{3,1} &= \big(w_\infty^{(3)}\big)_{2,1} = \big(w_\infty^{(3)}\big)_{1,1} = 0; \\ \big(w_\infty^{(4)}\big)_{3,1} &= \big(w_\infty^{(4)}\big)_{2,1} = \big(w_\infty^{(4)}\big)_{1,1} = \big(w_\infty^{(4)}\big)_{1,2} = 0 \implies w_\infty^{(4)} \in \mathrm{Lie}(\mathrm{V}). \end{aligned}$$

Thus we are always lucky at some point.

2.7. Case 2.2, recurrence only in centralizer direction. Assume the assumption made in Sec.2.6 is wrong. This can be rephrased as saying that there exists some $\delta_0 > 0$, assumed to be much smaller than $\mathrm{InjRad}(\Lambda_2)$, such that

$$\mathrm{Obt}^{-1}\left(\mathrm{Map}(\Lambda_2, \mathrm{Y}_0) \cap \mathcal{N}_{\delta_0}(\mathrm{id})\right) \subset \mathfrak{h} \oplus \mathrm{Lie}(\mathrm{V})$$

where $\mathrm{Obt}: \mathfrak{h} \oplus \mathfrak{h}^{\perp} \to \mathrm{SL}_3(\mathbb{R})$ is a local diffeomorphism (around $(0,0)$) defined by

$$\mathrm{Obt}(h, w) := \exp(h)\exp(w).$$

This is the last and the most annoying case. It is here that we are using the fact that Y_1 is B-minimal. We are going to derive a contradiction and show that this case is not allowed. The argument below is a more-or-less reproduction of [**BM00**, Page 182].

2.7.1. *Step 1.* Y_1 is not a closed B-orbit.

Indeed, otherwise, one sees that Y_1 is even compact by Thm.1.4. But this is impossible by considering $\mathbf{a}_t.\Lambda_2$ as $t \to -\infty$.

2.7.2. *Step 2.* Step 1 together with minimality imply that there exists $b_n = a_n u_n \in B$ with $a_n \to \infty$ such that $b_n.\Lambda_2 \to \Lambda_2$. Note that if we write $a_n = \mathbf{a}_{t_n}$ then $t_n \to +\infty$.

2.7.3. *Step 3.* Since $\mathrm{Y}_1 \subset \mathrm{Y}_0 \setminus \mathscr{O}$ and by our assumption made in this subsection, we find $(v_n) \subset \mathrm{Lie}(\mathrm{V})$ such that $v_n \neq 0$, $v_n \to 0$ and $\exp(v_n).\Lambda_2 \in \mathrm{Y}_0$ for all n.

2.7.4. *Step 4.* This is the key step.

Since $b_n.\Lambda_2 \to \Lambda_2$, we can find for every large n, a unique λ_n close to id such that $b_n.\Lambda_2 = \lambda_n.\Lambda_2$. By assumption one can write $\lambda_n = h_n \exp(v(\lambda_n))$ for some $h_n \in \mathrm{H}$ and $v(\lambda_n) \in \mathrm{Lie}(\mathrm{V})$. We want to argue that $h_n \in \mathrm{B}$.

Now fix some large n and will take l large compared to n. We have

$$b_n.(\exp(v_l).\Lambda_2) = \exp(v_l').b_n.\Lambda_2 = \exp(v_l').\lambda_n.\Lambda_2$$

where $v_l' = \mathrm{Ad}(b_n).v_l \in \mathrm{Lie}(\mathrm{V})$. When l is large compared to n, v_l' is small.

By assumption for n large and l larger,

$$\exp(v_l') \cdot \lambda_n = \exp(v_l') h_n \exp(v(\lambda_n)) \in \mathcal{N}_{\mathrm{id}}(\delta_0) \cap \mathrm{H} \cdot \mathrm{V}.$$

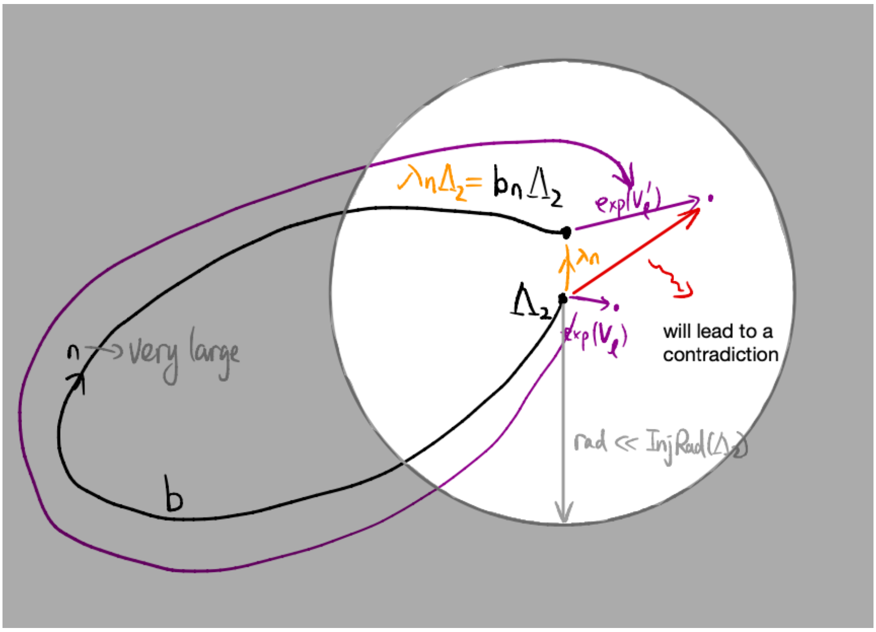
Although the computation of $\log(\exp(X)\exp(Y))$ is usually hard, we still have (again, for l large)

$$\begin{aligned}\exp(v_l') h_n \exp(v(\lambda_n)) \in \mathcal{N}_{\mathrm{id}}(\delta_0) \cap \mathrm{H} \cdot \mathrm{V} &\Longrightarrow \exp(\mathrm{Ad}(h_n^{-1}).v_l') \in \mathcal{N}_{\mathrm{id}}(\delta_0) \cap \mathrm{H} \cdot \mathrm{V} \\ \Longrightarrow \exp(\mathrm{Ad}(h_n^{-1}).v_l') \in \mathcal{N}_{\mathrm{id}}(\delta_0) \cap \mathrm{V} &\Longrightarrow \mathrm{Ad}(h_n^{-1}).v_l' \in \mathrm{Lie}(\mathrm{V}).\end{aligned}$$

[some δ_0 should be smaller than the others, we leave it to the reader to fill in the details]

As l varies, v_l' spans $\mathrm{Lie}(\mathrm{V})$. Thus h_n preserves $\mathrm{Lie}(\mathrm{V})$ and is contained in $\pm\mathrm{B}$, the normalizer of V in G. Since h_n is close to the identity, h_n belongs to B. [Rmk: since h_n is close to identity, this is a Lie algebraic calculation of $\mathfrak{n}_{\mathfrak{g}}(\mathrm{Lie}(\mathrm{V}))$, the normalizer of $\mathrm{Lie}(\mathrm{V})$ in $\mathfrak{g} := \mathfrak{sl}_3(\mathbb{R})$. That is, it suffices to compute the connected component of $N_{\mathrm{G}}(\mathrm{V})$.]

Here is a pictorial summary:



2.7.5. *Step 5.* Step 4 says that

$$b_n.\Lambda_2 = h_n \exp(v(\lambda_n)).\Lambda_2$$

for some $h_n \in \mathrm{B}$ close to the identity. This is impossible! Why? Note that Λ_2 is a periodic U-orbit and everything here normalizes U. Hence both sides are U-periodic. However, the centralizer of U preserves the period but $\mathbf{a}_{t_n}$ (recall $b_n = \mathbf{a}_{t_n} u_n$ with $t_n \to +\infty$) makes the period much larger. This is a contradiction.

3. Proof of Theorem 1.4

From now on we discuss how Thm.1.4 is proved. From the start assume item 2 and 3 do not happen and want to prove item 1 holds.

A direct computation shows that a vector $v \in \mathbb{R}^3 \oplus \wedge^2\mathbb{R}^3$ is fixed by U iff $v \in \mathbb{R}e_1 \oplus \mathbb{R}e_1 \wedge e_2$. For a primitive subgroup Δ of $\Lambda_0 \in \mathrm{X}_3$, we still denote by Δ the vector (well-defined up to ± 1) representing Δ. For instance if $\Delta = \mathbb{Z}v \oplus \mathbb{Z}w$, then Δ is viewed as a vector $\pm v \wedge w \in \wedge^2\mathbb{R}^3$. Now assume $\mathrm{U}.\Lambda_0$ does not degenerate, then every nonzero subgroup Δ is not fixed by U and by the feature of polynomials,

$$\lim_{t\to-\infty} \|\mathbf{u}_t.\Delta\| = \lim_{t\to+\infty} \|\mathbf{u}_t.\Delta\| = +\infty.$$

We can ensure at least the trajectory under U of each subgroup can not be small for a long time:

LEMMA 3.1. *There exist $C_2 > 0$ and $\alpha_2 > 0$ such that for every interval $[a,b]$ in $\mathbb{R}$, every $\mathbf{x} \in \mathbb{R}^3 \oplus \wedge^2\mathbb{R}^3$ and every $\rho \in (0,1)$, if $M_0 := \sup_{s\in[a,b]} \|\mathbf{u}_s\mathbf{x}\|$, then*

$$\frac{1}{b-a}\mathrm{Leb}\left\{s \in [a,b] \,\middle|\, \|\mathbf{u}_s\mathbf{x}\| < \rho M_0\right\} \le C_2\rho^{\alpha_2};$$

The proof is left as an exercise.

The key observation we made last time does not hold anymore. The following notion is aimed to save the situation, providing a sufficient condition for being contained in a compact set.

DEFINITION 3.2. *For $\delta,\rho\in(0,1)$, $\Lambda\in X_3$ is said to be (δ,ρ)-protected (by the flag $\{\{0\}\subset\mathbb{Z}v\subset\Delta\subset\Lambda\}$) iff there exists $0\subset\mathbb{Z}v\subset\Delta\subset\Lambda$ where $\mathbb{Z}v$ and Δ are primitive subgroups of rank 1 and 2 such that*

$$\|v\|,\|\Delta\|\in(\rho\delta,\delta).$$

LEMMA 3.3. *Take $\delta,\rho\in(0,1)$. If $\Lambda\in X_3$ is (δ,ρ)-protected then $\Lambda\in\mathscr{C}_\rho$.*

PROOF. It suffices to prove that every non-zero vector w in Λ has norm at least ρ. So we may assume that $\|w\|<1$.

Pick v and Δ as in the definition. Because Λ is of covolume one, w has to be contained in Δ since $\|w\|<1$. Moreover

$$\rho\delta\le\|\Delta\|\le\|v\|\cdot\|w\|\le\delta\|w\|\implies\|w\|\ge\rho.$$

□

Key observation. Here we have already employed the special feature of X_3 (not valid for $X_{\ge4}$): once we find $\mathbb{Z}v$ and Δ two primitive subgroups such that $\|\mathbb{Z}v\|,\|\Delta\|<1$, then it is automatic that $\mathbb{Z}v$ is contained in Δ. Therefore, in searching for a flag that (δ,ρ)-protects Λ we may look for $\mathbb{Z}v$ and Δ in an independent way (the condition of $\mathbb{Z}v\subset\Delta$ automatically holds).

Thus Thm.1.4 follows from Lem.3.3, the key observation and the following:

LEMMA 3.4. *For every $\varepsilon>0$, there exist $\varepsilon',\rho,\delta\in(0,1)$ such that for every Λ nondegenerate, there exists T_0 such that for all $T\ge T_0$,*

$$\frac{1}{T}\mathrm{Leb}\left\{t\in[0,T]\,\middle|\,\nexists\mathbf{x}\in\mathrm{Prim}^1(\mathbf{u}_t\Lambda),\ \|\mathbf{x}\|\in(\rho\delta,\delta),\ \mathbf{u}_t.\Lambda\notin\mathscr{C}_{\varepsilon'}\right\}\le\varepsilon,$$

and

$$\frac{1}{T}\mathrm{Leb}\left\{t\in[0,T]\,\middle|\,\nexists\Delta\in\mathrm{Prim}^2(\mathbf{u}_t\Lambda),\ \|\Delta\|\in(\rho\delta,\delta),\ \mathbf{u}_t.\Lambda\notin\mathscr{C}_{\varepsilon'}\right\}\le\varepsilon.$$

If we fix a compact set in X_3 from the beginning and allow ε',ρ,δ to depend on this compact set, then conclusion holds for $T_0=0$ and all Λ contained in this compact set.

PROOF OF LEMMA 3.4. Let C_2,α_2 be as in Lem.3.1.

Fix some $\delta\in(0,1)$. Take $\varepsilon':=\delta/2$. Choose $\rho\in(0,1)$ small enough such that $C_2(2\rho)^{\alpha_2}<0.5\varepsilon$. Assume that Λ contains no degenerate vectors. We are going to prove the first inequality and the second one can be proved similarly, which is left as an exercise.

By taking T_0 large enough, we assume that for every $\mathbb{Z}v\in\mathrm{Prim}^1(\Lambda)$, for some $t\in(0,T_0)$, $\|\mathbf{u}_t.v\|\ge\delta$ (and we can forget about the non-degeneracy condition from now on).

Indeed, take $t=1$, there are only finitely many $\mathbb{Z}v\in\mathrm{Prim}^1(\Lambda)$ such that $\|\mathbf{u}_t.v\|<\delta$. List them as $\{\mathbb{Z}v_1,...,\mathbb{Z}v_l\}$. Since $\mathbf{u}_t$ does not fix v_i for every i by non-degeneracy condition, we have that $\|\mathbf{u}_t.v_i\|\to+\infty$ as $t\to+\infty$. So we can pick T_0 such that $\left\|\mathbf{u}_{T_0}.v_i\right\|>\delta$ for every i and this would do the job.

Consider the set $\{t\in(0,T),\mathbf{u}_t.\Lambda\notin\mathscr{C}_{\delta/2}\}$, which is open and hence can be written as a disjoint union of open intervals. Take one of them, say (a,b). At the moment, we have not excluded the possibility of $(a,b)=(0,T)$ yet.

For every $t\in(a,b)$, by definition, there is some $\mathbb{Z}v\in\mathrm{Prim}^1(\Lambda)$ such that $\|\mathbf{u}_t.v\|<\delta/2$. For every such $\mathbb{Z}v$ and t, define $\mathscr{O}(\mathbb{Z}v,t)$ to be the maximal open interval in $\mathbb{R}$ containing t such that

$$s\in\mathscr{O}(\mathbb{Z}v,t)\implies\|\mathbf{u}_t.v\|<\delta.$$

From the definition, it is possible that $\mathscr{O}(\mathbb{Z}v,t)$ is not contained in (a,b), or even $(0,T)$. On the other hand, it is impossible for $(0,T)$ to be contained in $\mathscr{O}(\mathbb{Z}v,t)$ by the choice of T_0. Thus,

$$\sup_{(0,T)\cap\mathscr{O}(\mathbb{Z}v,t)}\|\mathbf{u}_t.v\|\ge\delta.$$

If (a,b) contains some end point of $\mathscr{O}(\mathbb{Z}v,t)$ then this also holds replacing $(0,T)$ by (a,b). Otherwise, we must have for $t=a$ or $t=b$, $\mathrm{sys}(\mathbf{u}_t.\Lambda)=\delta/2$. No matter what, the following always holds

$$\sup_{(a,b)\cap\mathscr{O}(\mathbb{Z}v,t)} \|\mathbf{u}_t.v\| \geq \frac{\delta}{2}. \tag{22}$$

As $\mathbb{Z}v,t$ varies, $\{\mathscr{O}(\mathbb{Z}v,t)\cap(a,b)\}$ covers (a,b). Now we claim that it is possible to select a subcovering with multiplicity at most 2 (the number 2 is not important, but it should be an absolute constant). The *multiplicity* of a covering refers to the maximal number of possible overlaps. Here is one possible way of proving the claim, you may wish to find your own.

Since each of a,b belongs to some $\mathscr{O}(\mathbb{Z}v,t)$, we can find a finite collection of $\{\mathscr{O}(\mathbb{Z}v,t)\}$ that covers (a,b). By passing to a further sub-covering if necessary, we assume it is minimal and is given by $\{\mathscr{O}(\mathbb{Z}v_i,t_i)=(a_i,b_i)\}$ with $a_i<a_{i+1}$. Then we must have

$$a_1<a<a_2<b_1<a_3<b_2<a_4<....<a_l<b_{l-1}<b<b_l,$$

and the claim follows.

Let $I_i:=(a_i,b_i)\cap(a,b)$. By Equa.(22), $\sup_{s\in I_i}\|\mathbf{u}_s.v_i\|\geq\frac{\delta}{2}$. Then by Lem.3.1,

$$\frac{1}{|I_i|}\left|\left\{s\in I_i\,\middle|\,\|\mathbf{u}_s.v_i\|\leq(2\rho)\cdot\frac{\delta}{2}\right\}\right|\leq C_2(2\rho)^{\alpha_2}\leq 0.5\varepsilon.$$

Adding them together completes the proof. □

4. Exercises

4.1. (C,α)-good property of polynomials of bounded degree. Let $C,\alpha>0$ and J be an interval in $\mathbb{R}$, recall a function $f:J\to\mathbb{R}$ is said to be (C,α)-good on J iff for every interval $I\subset J$ of finite length and every $\rho\in(0,1)$,

$$\frac{1}{|I|}\operatorname{Leb}\left\{t\in I\,\middle|\,|f(t)|\leq\rho M_I\right\}\leq C\rho^{\alpha}. \tag{23}$$

where $M_I:=\sup_{t\in I}|f(t)|$.

In this set of exercises we show that there are constants (C,α) such that every polynomial of degree at most three is (C,α)-good on $\mathbb{R}$. The general case would follow from the same proof with some constant depending only on the degree.

Given four distinct points $\mathbf{v}=(v_0,v_1,v_2,v_3)$ in $\mathbb{R}$, for $k=0,1,2,3$, define

$$L^k_{\mathbf{v}}(x):=\prod_{i\neq k}\frac{x-v_i}{v_k-v_i}.$$

EXERCISE 4.1. *Fix such a $\mathbf{v}$ as above. Prove that for any choice of four real numbers (w_0,w_1,w_2,w_3), there exists at most one polynomial p of degree at most 3 such that $p(v_i)=w_i$.*

EXERCISE 4.2. *Same assumption as in last exercise. Show that $p(x):=\sum_{k=0}^3 w_k\cdot L^k_{\mathbf{v}}(x)$ satisfies $p(v_i)=w_i$ for every $i=0,1,2,3$.*

EXERCISE 4.3. *Same assumption as in last exercise. Let $\varepsilon,\delta>0$ be two positive real numbers. Assume further that $|v_i-v_j|\geq\delta$ for every pair (i,j) with $i\neq j$. Also assume $|w_i|\leq\varepsilon$ for all i. Show that for every $x\in[0,1]$, $|p(x)|\leq 4\varepsilon\delta^{-3}$ where p is as in the last exercise.*

EXERCISE 4.4. *Let $I\subset[0,1]$ be a measurable subset with $\operatorname{Leb}(I)=9\delta>0$. Show that there exists four points (v_0,v_1,v_2,v_3) in I such that $|v_i-v_j|\geq\delta$ for every pair (i,j) with $i\neq j$.*

EXERCISE 4.5. *Find $C,\alpha>0$ such that for every polynomial of degree at most three and $\rho\in(0,1)$, Equa.(26) holds when $I=[0,1]$.*

EXERCISE 4.6. *Show that every polynomial of degree at most three is (C,α)-good on $\mathbb{R}$ with C,α same as in the last exercise.*

CHAPTER 6

Nondivergence of unipotent flows on X_N

Back to the Top.
Notations

- $\mathrm{X}_N := \left\{ \text{unimodular lattices in } \mathbb{R}^N \right\} \cong \mathrm{SL}_N(\mathbb{R}) / \mathrm{SL}_N(\mathbb{Z})$;
- for a discrete subgroup Δ in $\mathbb{R}^N$, let $\|\Delta\| := \mathrm{Vol}(\Delta_{\mathbb{R}}/\Delta)$ where $\Delta_{\mathbb{R}}$ denotes the $\mathbb{R}$-linear span of Δ in $\mathbb{R}^N$;
- for $\Lambda \leq \mathbb{R}^N$, $\mathrm{sys}(\Lambda) := \inf_{v \neq 0 \in \Lambda} \|v\|$, and for $\delta > 0$, $\mathscr{C}_\delta := \{\Lambda \in \mathrm{X}_N : \mathrm{sys}(\Lambda) \geq \delta\}$;
- $\mathrm{Prim}^k(\Lambda) := \left\{ \text{primitive subgroups of } \Lambda \text{ of rank } k \right\}$, $\mathrm{Prim}(\Lambda) := \bigcup_{k=0}^{\mathrm{rank}(\Lambda)} \mathrm{Prim}^k(\Lambda)$.

1. Summary and definitions

We would like to illustrate the main ideas behind [Kle10, Section 3] using X_4 as an example. The discussion can be generalized to X_N and even to $\mathbf{G}(\mathbb{R})/\mathbf{G}(\mathbb{Z})$ for other semisimple algebraic groups $\mathbf{G}$. Warning: our presentation and sometimes definitions differ from [Kle10, Section 3] and is "less careful" in many ways.

The discussion is useful beyond unipotent flows on X_N. We would like to mention [EMS97, MW02] here.

DEFINITION 1.1. *Fix (C,α) two positive constants. A map $\phi : I \to \mathrm{SL}_N(\mathbb{R})$ is said to be (C,α)-good at $\Lambda \in \mathrm{X}_N$ if for every primitive subgroup Δ of Λ, every interval $J \subset I$, every $\rho \in (0,1)$ (the case $\rho \geq 1$ is rather trivial), define $M(J,\Delta) := \sup_{s\in J} \left\|\phi_s.\Delta\right\|$, then we have*

$$\frac{1}{|J|} \left|\left\{ s \in J \,\middle|\, \left\|\phi_s.\Delta\right\| \leq \rho \cdot M(J,\Delta) \right\}\right| \leq C \cdot \rho^\alpha.$$

The main examples for us are unipotent flows.

LEMMA 1.2. *There are constants $C_N, \alpha_N > 0$, depending only on N such that for every nilpotent matrix u in $\mathfrak{sl}_N(\mathbb{R})$ and for every (finite or infinite) interval I in $\mathbb{R}$, $\phi(t) := \exp(t.u)$ is (C_N,α_N)-good at every $\Lambda \in \mathrm{X}_N$.*

PROOF. Exercise or see [Kle10]. □

THEOREM 1.3. *Fix $C,\alpha,\varepsilon,\delta$ positive constants. There exists a constant $\kappa = \kappa(C,\alpha,\varepsilon,\delta) > 0$ such that the following holds. Let $\Lambda \in \mathrm{X}_N$ and $\phi : I \to \mathrm{SL}_N(\mathbb{R})$. Assume*

- *ϕ is (C,α)-good at Λ;*
- *$\sup_{t\in I} \left\|\phi_t.\Delta\right\| \geq \delta$ for every $\Delta \in \mathrm{Prim}(\Lambda)$,*

then

$$\frac{1}{|I|} \mathrm{Leb}\left\{ s \in I \,\middle|\, \phi_s.\Lambda \notin \mathscr{C}_\kappa \right\} \leq \varepsilon.$$

In the case of unipotent flows and an interval I of infinite length, if the condition fails, then Λ contains a primitive subgroup fixed by the unipotent flow with small norm.

2. Nondivergence and flags

The key notion is being (δ,ρ)-protected, which provides a sufficient condition to guarantee non-divergence.

DEFINITION 2.1.
*A subset $\mathscr{F}$ of $\mathrm{Prim}(\Lambda)$ is said to be a **flag** if for every two elements Δ_1 and Δ_2 in $\mathscr{F}$, either $\Delta_1 \subset \Delta_2$ or $\Delta_1 \supset \Delta_2$. The **length** of a flag $\mathscr{F}$ is simply the cardinality of $\mathscr{F}$.*

DEFINITION 2.2.
*Let $\delta,\rho \in (0,1)$. Let $\Lambda \in \mathrm{X}_N$ and $\mathscr{F} = \{\Delta_1 \subsetneq \Delta_2 \subsetneq ... \subsetneq \Delta_l\}$ be a flag in $\mathrm{Prim}(\Lambda)$. We say that Λ is weakly (δ,ρ)-**protected** by $\mathscr{F}$ iff*

1. *$\rho\cdot\delta \le \|\Delta_i\| \le \delta$ for every $i = 1,..,l$;*
2. *$\|\Delta\| \ge 0.5\delta$ for every $\Delta \notin \mathscr{F}$ **comparable** with $\mathscr{F}$, i.e. $\mathscr{F}\cup\{\Delta\}$ is still a flag.*

*Now given a map $\phi: I \to \mathrm{SL}_N(\mathbb{R})$. We say that $s \in I$ is weakly (δ,ρ)-**protected** by $\mathscr{F}$ iff*

1. *$\rho\cdot\delta \le \left\|\phi_s.\Delta_i\right\| \le \delta$ for every $i = 1,..,l$;*
2. *$\left\|\phi_s.\Delta\right\| \ge 0.5\delta$ for every $\Delta \notin \mathscr{F}$ comparable with $\mathscr{F}$.*

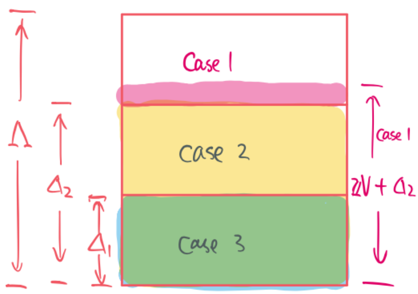
That is to say, $\phi_s.\Lambda$ is weakly (δ,ρ)-protected by $\phi_s.\mathscr{F}$.

I shall drop the word "weakly" later. But keep in mind our definition is different from [Kle10] where 0.5δ is replaced by δ.

From the definition, such a flag is not allowed to contain $\{0\}$ or Λ. Thus the maximal possible length is $N-1$.

One may wish to compare with the definition of Siegel sets.

LEMMA 2.3 (Criterion of non-divergence in terms of flags). *Fix $\delta,\rho \in (0,1)$. Assume for some reason that $\rho < 0.5$. Then there exists a constant $\theta = \theta(\delta,\rho) > 0$ (from the proof, can take $\theta = \rho^N\delta$) such that if $\Lambda \in \mathrm{X}_N$ is (δ,ρ)-protected by some flag $\mathscr{F}$ of $\mathrm{Prim}(\Lambda)$, then $\|\Delta\| \ge \theta$ for every primitive subgroup $\Delta \le \Lambda$. In particular $\mathrm{sys}(\Lambda) \ge \theta$.*



PROOF OF A SPECIAL CASE. Say $\mathscr{F} = \{\Delta_1 \le \Delta_2\}$, which gives a filtration of Λ. For $v \in \Lambda$, there are three cases. We will show $\|v\| \ge \rho\delta$.

Case 1. $v \in \Lambda \setminus \Delta_2$.
Then $\Delta_2 + \mathbb{Z}.v$ is compatible with $\mathscr{F}$, though it may not be primitive. $((\Delta_2)_{\mathbb{R}} + \mathbb{R}.v) \cap \Lambda$ is a primitive subgroup compatible with $\mathscr{F}$ and contains $\Delta_2 + \mathbb{Z}.v$. Thus

$$\|\Delta_2 + \mathbb{Z}.v\| \ge \|((\Delta_2)_{\mathbb{R}} + \mathbb{R}.v) \cap \Lambda\| \ge 0.5\delta.$$

On the other hand

$$\|\Delta_2 + \mathbb{Z}.v\| \le \|\Delta_2\| \cdot \|v\| \le \delta\,\|v\|.$$

Combined together gives $\|v\| \ge 0.5$.

Case 2. $v \in \Delta_2 \setminus \Delta_1$.
Either $\Delta_1 + \mathbb{Z}.v$ has the same rank as Δ_2 or not. Anyway, we always have,

$$\|\Delta_1 + \mathbb{Z}.v\| \geq \min\{\rho\delta, 0.5\delta\} = \rho\delta.$$

On the other hand

$$\|\Delta_1 + \mathbb{Z}.v\| \leq \|\Delta_1\| \cdot \|v\| \leq \delta \|v\|.$$

Combined together gives $\|v\| \geq \rho$.

Case 3. $v \in \Delta_1$.
Then either $\mathbb{Z}.v$ has the same rank as Δ_1, in which case $\|\mathbb{Z}.v\| \geq \|\Delta_1\| \geq \rho\delta$, or $\mathbb{Z}.v$ has smaller rank than Δ_1, in which case $\|\mathbb{Z}.v\| \geq 0.5\delta \geq \rho\delta$.

□

PROOF IN GENERAL. [Read this only if you feel necessary!] Let $\mathscr{F} = \{\Delta_1 \subsetneq \Delta_2 \subsetneq ... \subsetneq \Delta_l\}$ be the flag and Δ is a primitive subgroup of Λ. Let $V_k := \mathbb{R}^N/(\Delta_k)_{\mathbb{R}}$ and π_k be the natural quotient map $\mathbb{R}^N \to V_k$.

Note that if $\Delta' \leq \Lambda$ is contained in Δ_k for some $k \in \{1, ..., l\}$, then

$$\left\|\pi_{k-1}(\Delta')\right\|_{V_{k-1}} = \left\|\pi_{k-1}(\Delta' + \Delta_{k-1})\right\|_{V_{k-1}} = \frac{\left\|\Delta' + \Delta_{k-1}\right\|}{\|\Delta_{k-1}\|} \geq \frac{\min\{\rho\delta, 0.5\delta\}}{\delta} \geq \rho. \tag{24}$$

Let a be the largest index such that Δ_a is contained in Δ. By default, $\Delta_0 := \{0\}$ if $\Delta_1 \neq \{0\}$. If $a = l$, then we are done with $\theta = \rho\delta$. Assume otherwise.

$$\begin{aligned}
\|\Delta\| &= \|\pi_{a+1}(\Delta)\|_{V_{a+1}} \cdot \|\Delta \cap \Delta_{a+1}\| = \|\pi_{a+1}(\Delta + \Delta_{a+1})\|_{V_{a+1}} \cdot \|\pi_a(\Delta \cap \Delta_{a+1})\|_{V_a} \cdot \|\Delta_a\| \\
&= \|\pi_{a+2}(\Delta)\|_{V_{a+2}} \cdot \|\pi_{a+1}(\Delta + \Delta_{a+1}) \cap \pi_{a+1}(\Delta_{a+2})\|_{V_{a+1}} \cdot \|\pi_a(\Delta \cap \Delta_{a+1})\|_{V_a} \cdot \|\Delta_a\| \\
&= \|\pi_{a+2}(\Delta)\|_{V_{a+2}} \cdot \|\pi_{a+1}\left((\Delta + \Delta_{a+1}) \cap \Delta_{a+2}\right)\|_{V_{a+1}} \cdot \|\pi_a(\Delta \cap \Delta_{a+1})\|_{V_a} \cdot \|\Delta_a\| \\
&= \|\pi_{a+2}(\Delta)\|_{V_{a+2}} \cdot \|\pi_{a+1}\left(\Delta \cap \Delta_{a+2}\right)\|_{V_{a+1}} \cdot \|\pi_a(\Delta \cap \Delta_{a+1})\|_{V_a} \cdot \|\Delta_a\| \\
&\quad \ldots\ldots \\
&= \|\pi_{a+k-1}\left(\Delta \cap \Delta_{a+k}\right)\|_{V_{a+k-1}} \cdot \ldots \cdot \|\pi_{a+1}\left(\Delta \cap \Delta_{a+2}\right)\|_{V_{a+1}} \cdot \|\pi_a(\Delta \cap \Delta_{a+1})\|_{V_a} \cdot \|\Delta_a\| \\
&= \|\pi_{a+k-1}\left(\Delta \cap \Delta_{a+k}\right)\|_{V_{a+k-1}} \cdot \ldots \cdot \|\pi_{a+1}\left(\Delta \cap \Delta_{a+2}\right)\|_{V_{a+1}} \cdot \|\pi_a(\Delta \cap \Delta_{a+1})\|_{V_a} \cdot \|\Delta_a\|
\end{aligned}$$

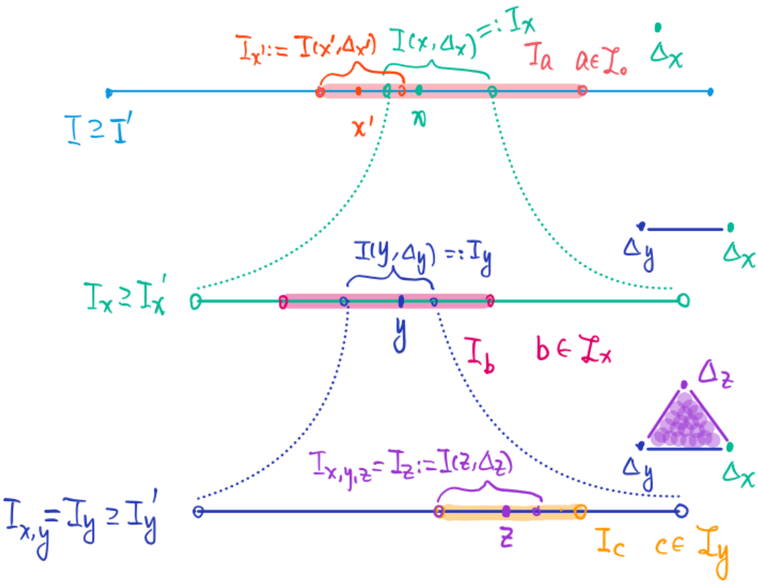
where k is the smallest positive integer such that Δ is contained in Δ_{a+k}. By invoking Equa.(24),

$$\|\Delta\| \geq \rho^k \delta.$$

So we are done by taking $\theta := \rho^N \delta$. □

3. The proof

Instead of proving by induction, we have decided to unfold this process. This makes the proof much longer but hopefully less mysterious. Here is a guide for Step 1-3.



Step 1. By assumption for every $\Delta \in \mathrm{Prim}(\Lambda)$,

$$\sup_{s\in I} \|\phi_s.\Delta\| \geq \delta.$$

Consider the open subset

$$I' := \{s \in I \,|\, \exists \Delta \in \mathrm{Prim}(\Lambda),\ \|\phi_s.\Delta\| < 0.9\delta\}$$

Write it as a disjoint union of open intervals

$$I' = \bigsqcup_{a\in\mathscr{I}_0} I_a.$$

Thus for every $\Delta \in \mathrm{Prim}(\Lambda)$,

$$\sup_{t\in I'} \|\phi_s.\Delta\| \geq 0.9\delta.$$

For $a \in \mathscr{I}_0$, consider (the 0.9 here is just to get a finite cover later, but it is not necessary to do so)

$$\mathscr{A}_a := \{(x,\Delta) \in I_a \times \mathrm{Prim}(\Lambda) \,|\, \|\phi_x.\Delta\| < 0.9\delta\}.$$

For each $(x,\Delta) \in \mathscr{A}_a$, define

$$I(x,\Delta) := \text{ the connected component of } \{s \in I_a \,|\, \|\phi_s.\Delta\| < \delta\} \text{ containing } x.$$

For every $x \in I_a$, pick some Δ_x such that $I_x := I(x,\Delta_x)$ is maximal among (the finitely many) $I(x,\Delta)$ as (x,Δ) varies in $\mathscr{A}_a$. By this choice, I_x and Δ_x satisfy

1. for every $\Delta \in \mathrm{Prim}(\Lambda)$, $\sup_{s\in I_x} \|\phi_s.\Delta\| \geq 0.9\delta$;
2. $\sup_{s\in I_x} \|\phi_s.\Delta_x\| \leq \delta$.

I_a admits a finite sub-covering by I_x's and by passing to a further sub-covering, we assume

$$I_a = \bigcup_{x\in\mathscr{I}_a} I_x \quad \text{with multiplicity } \leq 2$$

where $\mathscr{I}_a$ is certain finite subset of I_a (finiteness is not important, multi ≤ 2 is). Also define

$$\mathscr{P}_x := \{\Delta \in \mathrm{Prim}(\Lambda) \,|\, \Delta \text{ is comparable to } \Delta_x\}.$$

Step 2. Consider the open subset of I_x:

$$I'_x := \left\{ s \in I_x \,\middle|\, \exists \Delta \in \mathscr{P}_x,\ \left\| \phi_s.\Delta \right\| < 0.8\delta \right\}.$$

Write it as a disjoint union of open intervals

$$I'_x = \bigsqcup_{b \in \mathscr{I}_x} I_b.$$

For $b \in \mathscr{I}_x$, consider

$$\mathscr{A}_b := \left\{ (y, \Delta) \in I_b \times \mathscr{P}_x \,\middle|\, \left\| \phi_y.\Delta \right\| < 0.8\delta \right\}.$$

For each $(y, \Delta) \in \mathscr{A}_b$, define

$$I(y, \Delta) := \text{ the connected component of } \left\{ s \in I_b \,\middle|\, \left\| \phi_s.\Delta \right\| < 0.9\delta \right\} \text{ containing } y.$$

For every $y \in I_b$, pick some Δ_y such that $I_{x,y} := I(y, \Delta_y)$ is maximal among (the finitely many) $I(y, \Delta)$ as (y, Δ) varies in $\mathscr{A}_b$. By this choice, $I_{x,y}$ and Δ_y satisfy

1. for every $\Delta \in \mathscr{P}_x$, $\sup_{s \in I_{x,y}} \left\| \phi_s.\Delta \right\| \geq 0.8\delta$;
2. $\sup_{s \in I_{x,y}} \left\| \phi_s.\Delta_y \right\| \leq 0.9\delta$.

Similarly,

$$I_b = \bigcup_{y \in \mathscr{I}_b} I_{x,y} \quad \text{with multiplicity } \leq 2$$

where $\mathscr{I}_b$ is some finite subset of I_b. Also define

$$\mathscr{P}_{x,y} := \left\{ \Delta \in \mathrm{Prim}(\Lambda) \,\middle|\, \Delta \text{ is comparable to } \{\Delta_x, \Delta_y\} \right\}.$$

Step 3. Consider the open subset of $I_{x,y}$:

$$I'_{x,y} := \left\{ s \in I_{x,y} \,\middle|\, \exists \Delta \in \mathscr{P}_{x,y},\ \left\| \phi_s.\Delta \right\| < 0.7\delta \right\}.$$

Write it as a disjoint union of open intervals

$$I'_{x,y} = \bigsqcup_{c \in \mathscr{I}_{x,y}} I_c.$$

For $c \in \mathscr{I}_{x,y}$, consider

$$\mathscr{A}_c := \left\{ (z, \Delta) \in I_c \times \mathscr{P}_{x,y} \,\middle|\, \left\| \phi_z.\Delta \right\| < 0.7\delta \right\}.$$

For each $(z, \Delta) \in \mathscr{A}_c$, define

$$I(z, \Delta) := \text{ the connected component of } \left\{ s \in I_c \,\middle|\, \left\| \phi_s.\Delta \right\| < 0.8\delta \right\} \text{ containing } z.$$

For every $z \in I_c$, pick some Δ_z such that $I_{x,y,z} := I(z, \Delta_z)$ is maximal among (the finitely many) $I(z, \Delta)$ as (z, Δ) varies in $\mathscr{A}_c$. By this choice, $I_{x,y,z}$ and Δ_z satisfy

1. for every $\Delta \in \mathscr{P}_{x,y}$, $\sup_{s \in I_{x,y,z}} \left\| \phi_s.\Delta \right\| \geq 0.7\delta$;
2. $\sup_{s \in I_{x,y,z}} \left\| \phi_s.\Delta_z \right\| \leq 0.8\delta$.

Similarly,

$$I_c = \bigcup_{z \in \mathscr{I}_c} I_{x,y,z} \quad \text{with multiplicity } \leq 2$$

where $\mathscr{I}_c$ is certain finite subset of I_c. Now $\{\Delta_x, \Delta_y, \Delta_z\}$ is already a complete flag modulo $\{0\}$ and Λ.

Good and bad points 1. For x, a, y, b, z, let

$$I_{x,y,z}(\mathrm{Good}) := \left\{ s \in I_{x,y,z} \,\middle|\, \left\|\phi_s.\Delta_z\right\| \geq \rho\delta \right\}, \quad I_{x,y,z}(\mathrm{Bad}) := I_{x,y,z} \setminus I_{x,y,z}(\mathrm{Good}).$$

By (C,α)-goodness, we choose $\rho \in (0,1)$ such that

$$\left|I_{x,y,z}(\mathrm{Bad})\right| \leq (0.01\varepsilon)\left|I_{x,y,z}\right|.$$

Thus

$$\begin{aligned}\left|I'_{x,y}(\mathrm{Bad})\right| &:= \left|\bigsqcup_{c \in \mathscr{I}_{x,y}} \bigcup_{z \in \mathscr{I}_c} I_{x,y,z}(\mathrm{Bad})\right| \leq \sum_c \sum_z \left|I_{x,y,z}(\mathrm{Bad})\right| \leq \sum_c \sum_z (0.01\varepsilon) \cdot \left|I_{x,y,z}\right| \\ &\leq \sum_c 2(0.01\varepsilon) \cdot |I_c| = (0.02\varepsilon) \cdot \left|I'_{x,y}\right|.\end{aligned}$$

Define $I'_{x,y}(\mathrm{Good}) := I'_{x,y} \setminus I'_{x,y}(\mathrm{Bad})$, so $I'_{x,y} = I'_{x,y}(\mathrm{Good}) \sqcup I'_{x,y}(\mathrm{Bad})$.

So far, we have the following regarding each $I_{x,y}$:

1. $s \in I_{x,y} \setminus I'_{x,y} \implies \left\|\phi_s.\Delta\right\| \geq 0.7\delta,\ \forall \Delta \in \mathscr{P}_{x,y}$;
2. $s \in I'_{x,y}(\mathrm{Good}) \implies \exists \Delta_z \in \mathscr{P}_{x,y},\ \rho\delta \leq \left\|\phi_s.\Delta_z\right\| \leq 0.8\delta$;
3. $\left|I'_{x,y}(\mathrm{Bad})\right| \leq 2\delta \cdot \left|I'_{x,y}\right|$.

Good and bad points 2. Define

$$I_{x,y}(\mathrm{Good}) := \left\{ s \in I_{x,y} \,\middle|\, \left\|\phi_s.\Delta_y\right\| \geq \rho\delta \right\}, \quad I_{x,y}(\mathrm{Bad}) := I_{x,y} \setminus I_{x,y}(\mathrm{Good}).$$

And ρ is chosen such that

$$\left|I_{x,y}(\mathrm{Bad})\right| \leq (0.01\varepsilon)\left|I_{x,y}\right|.$$

Thus,

$$\begin{aligned}\left|I'_x(\mathrm{Bad})\right| &:= \left|\bigsqcup_{b \in \mathscr{I}_x} \bigcup_{y \in \mathscr{I}_b} I_{x,y}(\mathrm{Bad})\right| \leq \sum_b \sum_y \left|I_{x,y}(\mathrm{Bad})\right| \leq \sum_b \sum_y (0.01\varepsilon) \cdot \left|I_{x,y}\right| \\ &\leq \sum_b 2(0.01\varepsilon) \cdot |I_b| = (0.02\varepsilon) \cdot \left|I'_x\right|.\end{aligned}$$

Define $I'_x(\mathrm{Good})$ by imposing $I'_x = I'_x(\mathrm{Good}) \sqcup I'_x(\mathrm{Bad})$.

So far, regarding I_x we have:

1. $s \in I_x \setminus I'_x \implies \left\|\phi_s.\Delta\right\| \geq 0.8\delta,\ \forall \Delta \in \mathscr{P}_x$;
2. $s \in I'_x(\mathrm{Good}) \cap I_{x,y} \implies \rho\delta \leq \left\|\phi_s.\Delta_y\right\| \leq 0.9\delta$;
3. $\left|I'_x(\mathrm{Bad})\right| \leq 2\delta \cdot \left|I'_x\right|$.

Good and bad points 3. Finally, define

$$I_x(\mathrm{Good}) := \left\{ s \in I_x \,\middle|\, \left\|\phi_s.\Delta_x\right\| \geq \rho\delta \right\}, \quad I_x(\mathrm{Bad}) := I_x \setminus I_x(\mathrm{Good}).$$

And ρ is chosen such that

$$|I_x(\mathrm{Bad})| \leq 0.01\varepsilon\, |I_x|.$$

Thus,

$$\begin{aligned}\left|I'(\mathrm{Bad})\right| &:= \left|\bigsqcup_{a \in \mathscr{I}_0} \bigcup_{x \in \mathscr{I}_a} I_x(\mathrm{Bad})\right| \leq \sum_a \sum_x |I_x(\mathrm{Bad})| \leq \sum_a \sum_x 0.01\varepsilon \cdot |I_x| \\ &\leq \sum_a 2 \cdot 0.01\varepsilon \cdot |I_a| = (0.02\varepsilon) \cdot \left|I'\right|.\end{aligned}$$

Define $I'(\mathrm{Good})$ by imposing $I' = I'(\mathrm{Good}) \sqcup I'(\mathrm{Bad})$. Here we have:

1. $s \in I \setminus I' \implies \left\|\phi_s.\Delta\right\| \geq 0.9\delta,\ \forall \Delta \in \mathrm{Prim}(\Lambda)$;
2. $s \in I'(\mathrm{Good}) \cap I_x \implies \rho\delta \leq \left\|\phi_s.\Delta_x\right\| \leq \delta$;
3. $\left|I'(\mathrm{Bad})\right| \leq 2\delta \cdot \left|I'\right|$.

Warp-up. Now we collect all the bad points together and let

$$I(\mathrm{Bad}) := I'(\mathrm{Bad}) \cup \left(\bigcup_{a \in \mathscr{I}_0, x \in \mathscr{I}_a} I'_x(\mathrm{Bad}) \right) \cup \left(\bigcup_{a \in \mathscr{I}_0, x \in \mathscr{I}_a} \bigcup_{b \in \mathscr{I}_x, y \in \mathscr{I}_b} I'_{x,y}(\mathrm{Bad}) \right)$$

We have

$$\begin{aligned} \left| \bigcup_{a,x,b,y} I'_{x,y}(\mathrm{Bad}) \right| \leq \sum_{a,x,b \in \mathscr{I}_x} \sum_{y \in \mathscr{I}_b} \left| I'_{x,y}(\mathrm{Bad}) \right| &\leq (0.02\varepsilon) \cdot \sum_{a,x,b} \sum_{y} \left| I_{x,y} \right| \\ &\leq (0.04\varepsilon) \cdot \sum_{a,x,b \in \mathscr{I}_x} |I_b| \leq (0.04\varepsilon) \cdot \sum_{a,x} |I_x| \\ &\leq (0.08\varepsilon) \cdot |I| \end{aligned}$$

and

$$\begin{aligned} \left| \bigcup_{a \in \mathscr{I}_0, x \in \mathscr{I}_a} I'_x(\mathrm{Bad}) \right| \leq \sum_{a \in \mathscr{I}_0} \sum_{x \in \mathscr{I}_a} \left| I'_x(\mathrm{Bad}) \right| &\leq (0.02\varepsilon) \cdot \sum_{a \in \mathscr{I}_0} \sum_{x \in \mathscr{I}_a} |I_x| \\ &\leq (0.04\varepsilon) \cdot \sum_{a \in \mathscr{I}_0} |I_a| \leq (0.04\varepsilon) \cdot |I|. \end{aligned}$$

Hence

$$|I(\mathrm{Bad})| \leq (0.14\varepsilon) \cdot |I| < \varepsilon |I|. \tag{25}$$

Let $s \in I \setminus I(\mathrm{Bad})$.

Case 1. $s \in I \setminus I'$, then $\|\phi_s.\Delta\| \geq 0.9\delta > 0.5\delta$, $\forall \Delta \in \mathrm{Prim}(\Lambda)$ so it is (δ, ρ)-protected by the trivial flag.

Case 2. $s \in I' \setminus I(\mathrm{Bad}) = (\sqcup I_a) \setminus I(\mathrm{Bad}) = (\sqcup_a \cup_x I_x) \setminus I(\mathrm{Bad})$. Say $s \in I_x \setminus I(\mathrm{Bad})$. Then

$$\rho\delta \leq \|\phi_s.\Delta_x\| \leq \delta.$$

Case 2.1. $s \in I_x \setminus I'_x$. Then $\|\phi_s.\Delta\| \geq 0.8\delta > 0.5\delta$ for all $\Delta \in \mathscr{P}_x$. This means that s is (ρ, δ)-protected by $\{\Delta_x\}$.

Case 2.2. $s \in I'_x \setminus I(\mathrm{Bad}) = (\sqcup I_b) \setminus I(\mathrm{Bad}) = (\sqcup_b \cup_y I_{x,y}) \setminus I(\mathrm{Bad})$. Say $s \in I_{x,y} \setminus I(\mathrm{Bad})$. Then

$$\rho\delta \leq \|\phi_s.\Delta_y\| \leq \delta.$$

Case 2.2.1. $s \in I_{x,y} \setminus I'_{x,y}$. Then $\|\phi_s.\Delta\| \geq 0.7\delta > 0.5\delta$ for all $\Delta \in \mathscr{P}_{x,y}$. This means that s is (ρ, δ)-protected by $\{\Delta_x, \Delta_y\}$.

Case 2.2.2. $s \in I'_{x,y} \setminus I(\mathrm{Bad}) = \sqcup_c I_c \setminus I(\mathrm{Bad}) = \sqcup_c \cup_z I_{x,y,z} \setminus I(\mathrm{Bad})$. Say $x \in I_{x,y,z} \setminus I(\mathrm{Bad})$, then

$$\rho\delta \leq \|\phi_s.\Delta_z\| \leq \delta.$$

Thus s is (δ, ρ)-protected by $\{\Delta_x, \Delta_y, \Delta_z\}$.

Now every $s \in I \setminus I(\mathrm{Bad})$ falls into one of the cases 1, 2.1, 2.2.1 and 2.2.2, so it is (δ, ρ)-protected. Hence Lem.2.3 implies if $s \in I \setminus I(\mathrm{Bad})$ then $\phi_s.\Lambda \in \mathscr{C}_\theta$ with $\theta = \theta(\delta, \rho)$. Now we take $\kappa := \theta$. Combining with Equa.(25), we are done.

4. Exercises

4.1. More examples of (C, α)-good functions. Let $C, \alpha > 0$ and J be an interval in $\mathbb{R}$, recall a function $f : J \to \mathbb{R}$ is said to be (C, α)-good on J iff for every interval $I \subset J$ of finite length and every $\rho \in (0, 1)$,

$$\frac{1}{|I|} \mathrm{Leb}\left\{ t \in I \,\middle|\, |f(t)| \leq \rho M_I \right\} \leq C\rho^\alpha. \tag{26}$$

where $M_I := \sup_{t \in I} |f(t)|$.

Let J be an interval of finite length. Let

$$\mathscr{A} := \{f = ae^x + be^{-x},\ a, b \in \mathbb{R}\}.$$

EXERCISE 4.1. *Show that there exist* $C,\alpha>0$ *(depending on* J *and* $\mathscr{A}$*) such that for every function* $f\in\mathscr{A}$ *is* (C,α)*-good on* J*.*

EXERCISE 4.2. *If* f_1, f_2 *are* (C,α)*-good on* J*, then* $x\mapsto \max\{|f_1(x)|,|f_2(x)|\}$ *is also* (C,α)*-good on* J*.*

CHAPTER 7

Ergodicity and Mixing

Back to the Top.

1. Basic constructions

For details the reader may consult [EW11], especially chapter 8 and appendices therein.

Let G be a "nice" (σ-compact locally compact metrizable) topological group and X a "nice" (σ-compact locally compact metrizable) topological space. Assume G acts on X continuously, i.e. we have a continuous map $G \times X \to X$ satisfying some compatibility conditions.

Let $\mathscr{B}_X$ be the σ-algebra on X generated by open sets in X. This is termed the Borel σ-algebra. Then the G-action is also measurable with respect to $\mathscr{B}_X$. Thus G naturally acts on measures on $(X, \mathscr{B}_X)$.

DEFINITION 1.1. *A measure μ on $\mathscr{B}_X$ is called a Borel measure. It is called a probability measure iff $\mu(X) = 1$. The collection of all probability measures is denoted as* $\mathrm{Prob}(X)$. *We view* $\mathrm{Prob}(X)$ *as a topological space equipped with the weak-$*$ topology.*

More precisely, we embed $\mathrm{Prob}(X)$ with the weakest topology such that

$$\mu \mapsto \int f(x)\mu(x)$$

is continuous for every

$$f \in C_c(X) := \{\text{compactly supported real-valued continuous functions on } X\}.$$

Being real-valued or complex-valued is not important.

Let

$$\mathrm{Meas}(X)^{\leq 1} := \{\text{finite measures } \mu \text{ on } X,\ \mu(X) \leq 1\},$$

also equipped with weak-$*$ topology. We also let

$$\mathrm{LFM}(X) := \{\text{locally finite measures on } X\},$$

be equipped with weak-$*$ topology. Note that $C_c(X)$ admits a countable dense subset.

LEMMA 1.2. *With weak-$*$ topology,* $\mathrm{Meas}(X)^{\leq 1}$ *is a compact metrizable space. If X is compact, then so is* $\mathrm{Prob}(X)$.

REMARK 1.3. *If we forget about the topological structure on X, and take some probability measure μ, then up to completion, $(X, \mathscr{B}_X, \mu)$ is "isomorphic" to a convex combination of the natural measure on* $[0,1]$ *interval and atomic measures supported on single points (see* [Wal82, Theorem 2.1]*). Thus the study of $(X, \mathscr{B}_X, \mu)$ is rather boring without a group action, unlike the topological space X, when the classification of X is already a huge problem.*

We naturally has an action of G on $\mathrm{Prob}(X)$, $\mathrm{Meas}^{\leq 1}(X)$ and $\mathrm{LFM}(X)$ defined by

$$g_*\mu(E) := \mu(g^{-1}E)$$

for every measurable set E and measure μ.

LEMMA 1.4. *The induced map* $G \times \mathrm{LFM}(X) \to \mathrm{LFM}(X)$ *is continuous.*

A measure μ is said to be G-invariant iff $g_*\mu = \mu$ for all $g \in G$. The collection of G-invariant probability measures is denoted as $\mathrm{Prob}(X)^G$. Similarly define $\mathrm{Meas}^{\le 1}(X)^G$ and $\mathrm{LFM}(X)^G$.

To distinguish different p.m.p(= probability measure preserving) actions of G, a convenient functor is given by taking the associated unitary representation.

Take a $\mu \in \mathrm{LFM}(X)^G$. Then the associated unitary representation is given by

$$\begin{aligned} G \times L^2(X,\mu) &\to L^2(X,\mu) \\ (g,\phi) &\mapsto g\cdot\phi(x) := \phi(g^{-1}x) \end{aligned}$$

LEMMA 1.5. *This is indeed a unitary representation:*

1. *for each $g \in G$, the action on $L^2(X,\mu)$ is a unitary;*
2. *the representation is continuous*

where $\mathscr{U}(L^2(X,\mu))$, the set of unitary operators on $L^2(X,\mu)$, is equipped with the strong operator topology.

In more concrete terms, using the following lemma, the continuity claim just asserts that if $g_n \to g$ in G and $\phi_n \to \phi$ in $L^2(X,\mu)$, then $g_n\cdot\phi_n \to g\cdot\phi$ in $L^2(X,\mu)$.

LEMMA 1.6. $L^2(X,\mu)$ *admits a countable dense subset.*

For two p.m.p. G-actions to be isomorphic, it is necessary for the associated unitary representations to be isomorphic. Properties of p.m.p. G-actions defined via the associated unitary representation are sometimes called "spectral properties".

2. Ergodicity and mixing

We assume G and X to be nice in this section.

DEFINITION 2.1. *A p.m.p. G-action on $(X,\mathscr{B}_X,\mu)$ is said to be ergodic iff every G-invariant measurable subset E of X is either μ-null ($\mu(E)=0$) or μ-conull ($\mu(X\setminus E)=0$).*

So ergodicity is something like irreducibility.

LEMMA 2.2. *If a p.m.p. G-action on $(X,\mathscr{B}_X,\mu)$ is ergodic, then every μ-almost invariant measurable subset of X is either μ-null or μ-conull.*

A measurable subset $E \subset X$ is said to be μ-**almost invariant** iff for every $g \in G$,

$$\mu(gE\Delta E) = \mu((gE\setminus E)\cup(E\setminus gE)) = 0.$$

Since our group could be uncountable, this lemma is not obvious. Using this lemma, one can show that

LEMMA 2.3. *A p.m.p. G-action on $(X,\mathscr{B}_X,\mu)$ is ergodic iff the associated unitary representation has no fixed vector orthogonal to constants.*

Hint: Starting from a set E, one has the characteristic function 1_E. Starting from a function f, one considers its level sets.

By this lemma, being ergodic is a spectral property. Another spectral property we need is mixing.

DEFINITION 2.4. *A p.m.p. G-action on $(X,\mathscr{B}_X,\mu)$ is said to be mixing iff for every two measurable subsets $E,F\subset X$ and every divergent sequence (g_n) in G, we have*

$$\lim_{n\to\infty}\mu(g_n^{-1}E\cap F) = \mu(E)\mu(F).$$

This notion is useless for compact groups.

LEMMA 2.5. *A p.m.p. G-action on $(X, \mathscr{B}_X, \mu)$ is mixing iff for every two $\phi, \psi \in L^2(X,\mu)$ orthogonal to constants and every divergent sequence (g_n) in G, we have*

$$\lim_{n\to\infty} \langle g_n \cdot \phi, \psi\rangle = 0.$$

Here $\langle \phi, \psi\rangle := \int \phi(x)\overline{\psi(x)}\mu(x)$.

3. Unitary representations of $\mathrm{SL}_2(\mathbb{R})$ are mixing

Notations

- $\mathrm{G} := \mathrm{SL}_2(\mathbb{R})$ and Γ is a discrete subgroup of G;
- $\mathrm{A} := \left\{ \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}, \ t \in \mathbb{R} \right\} = \{\mathbf{a}_t, \ t \in \mathbb{R}\}$;
- $\mathrm{U} := \left\{ \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}, \ s \in \mathbb{R} \right\} = \{\mathbf{u}_s, \ s \in \mathbb{R}\}$;
- $\mathrm{B} := \mathrm{A} \cdot \mathrm{U}$.

For convenience let us make the following definition

DEFINITION 3.1. *A unitary representation $\pi : G \to \mathscr{U}(\mathscr{H})$ is mixing iff for every $v, w \in \mathscr{H}$ and every divergent sequence (g_n) in G, $\lim_{n\to\infty}\langle g_n.v, w\rangle = 0$.*

Unitary representations, if containing no non-zero vectors fixed by G, of $\mathrm{G} = \mathrm{SL}_2(\mathbb{R})$ are always mixing.

THEOREM 3.2. *Let π be a unitary representation of* G *on a separable Hilbert space $\mathscr{H}$. Assume there is no non-zero* G*-fixed vectors. Then π is mixing.*

PROOF. By "KAK-decomposition" (see the comment after the proof for an explanation), it suffices to show that $\pi|_{\mathrm{A}}$ is mixing. So take (a_n) to be a divergent sequence in A. By a diagonal argument, we find an infinite subsequence (a_{n_k}) such that for every $\phi, \psi \in \mathscr{H}$,

$$\lim_{k\to\infty} \langle a_{n_k}\phi, \psi\rangle \text{ exists.}$$

This defines a linear map $E : \mathscr{H} \to \mathscr{H}$ such that the above limit is equal to $\langle E\phi, \psi\rangle$. One can check $\|E\|_{\mathrm{op}} \le 1$ where $\|\cdot\|_{\mathrm{op}}$ stands for the operator norm. It suffices to show that $E = 0$, which is going to be achieved by showing that every vector contained in the image of E is fixed by G.

By passing to a further subsequence we assume either $\left(\log(a_{n_k})\right)_{1,1} \to +\infty$ or $-\infty$.

Define

$$U^- := \left\{ x \in \mathrm{G} \,\middle|\, \lim_{k\to+\infty} a_{n_k} x a_{n_k}^{-1} = 1 \right\},\ U^+ := \left\{ x \in \mathrm{G} \,\middle|\, \lim_{k\to+\infty} a_{n_k}^{-1} x a_{n_k} = 1 \right\}$$

There are two things we firstly note. Let E^* be the adjoint of E.

1. $E \circ u = E$ for every $u \in U^-$. Indeed, for every pair ϕ, ψ in $\mathscr{H}$,

$$\begin{aligned}\langle Eu\phi, \psi\rangle &= \lim\langle a_{n_k} u\phi, \psi\rangle = \lim\langle a_{n_k} u a_{n_k}^{-1} a_{n_k}\phi, \psi\rangle \\ &= \lim\langle a_{n_k}\phi, a_{n_k} u^{-1} a_{n_k}^{-1}\psi\rangle = \langle E\phi, \psi\rangle.\end{aligned}$$

The last step is because $\left(a_{n_k} u^{-1} a_{n_k}^{-1}\psi\right)$ converges to ψ in norm. Hence $E \circ u = E$. By taking the adjoint, we get $u^{-1} \circ E^* = E^*$. Thus the image of E^* is fixed by U^-.

2. $u \circ E = E$ for every $u \in U^+$. For every pair ϕ, ψ,

$$\begin{aligned}\langle uE\phi, \psi\rangle &= \langle E\phi, u^{-1}\psi\rangle = \lim\langle a_{n_k}\phi, u^{-1}\psi\rangle = \lim\langle a_{n_k} a_{n_k}^{-1} u a_{n_k}\phi, \psi\rangle \\ &= \lim\langle a_{n_k}\phi, \psi\rangle = \langle E\phi, \psi\rangle.\end{aligned}$$

Hence $u \circ E = E$.

Next is the trick. As the $*$ operation is continuous with respect to W.O.T., $(a_{n_k}^{-1}) = (a_{n_k}^*)$ converges in W.O.T. to E^*.

3. $\ker E = \ker E^*$. Indeed,

$$\langle E\phi, E\phi\rangle = \lim_l \lim_k \langle a_{n_k}\phi, a_{n_l}\phi\rangle = \lim_l \lim_k \langle a_{n_l}^{-1}\phi, a_{n_k}^{-1}\phi\rangle = \langle E^*\phi, E^*\phi\rangle$$

(Exercise: show that in general $\ker E \neq \ker E^*$ for a bounded linear operator on a Hilbert space.)

Now we can finish the proof. 1. says that $E(1-u) = 0$, $\forall u \in U^-$. Combined with 3., we get $E^*(1-u) = 0$, $\forall u \in U^-$. Taking $*$ of this, we get $E = u^{-1}E$. Thus the image of E is fixed by U^-. 2. asserts that the image of E is fixed by U^+. Since U^- and U^+ generates G, we are done.

$\square$

Let us quickly explain, using linear algebra, why you can write a matrix $g \in \mathrm{SL}_2(\mathbb{R})$ as $k_1 a k_2$ with k_i in $\mathrm{SO}_2(\mathbb{R})$ and a being diagonal. This fact was used to reducing the mixing in general to mixing of A. First we claim that we can write $g = k_1|g|$ where k_1 is orthogonal and $|g|$ is symmetric. Assuming the claim, since $|g|$ can be diagonalized under an orthogonal basis, we are done. Now let us prove the claim. The matrix gg^{tr} is symmetric and hence diagonalizable. Moreover it has positive eigenvalues. Hence makes sense to take $|g| := \sqrt{gg^{tr}}$. Then one defines $k_1 := g|g|^{-1}$ and it is direct to check that $\langle k_1 v, k_1 v\rangle = \langle v, v\rangle$ for every vector v. And we are done.

4. Examples

EXAMPLE 4.1. *Let $G := \mathbb{Z}$ generated by $1 := R_\alpha$ act on $\mathbb{R}/\mathbb{Z}$ by $R_\alpha \cdot x := x + \alpha$ for some real number α. Then this action preserves the natural Lebesgue measure m on $\mathbb{R}/\mathbb{Z}$. It is ergodic iff $\alpha \in \mathbb{R} \setminus \mathbb{Q}$. Moreover, this action is not mixing.*

SKETCH OF PROOF. There are two proofs. Assume $\alpha \notin \mathbb{Q}$.

Either you can argue that R_α generates a dense subgroup of $\mathbb{R}/\mathbb{Z}$ and then by continuity, m has to be invariant under the full $\mathbb{R}/\mathbb{Z}$. Then argue that m is the unique $\mathbb{R}/\mathbb{Z}$-invariant probability measure.

Or you can argue that there are no invariant L^2 functions by expanding them under the basis $\{x \mapsto e^{2\pi i n x}\}_{n\in\mathbb{Z}}$.

I leave it to you to show that R_α is not mixing. $\square$

EXAMPLE 4.2. *Let $G := \mathbb{Z}$ act on $\mathbb{R}^2/\mathbb{Z}^2$ where the generator 1 acts by $(x, y) \mapsto (x + y, x + 2y)$. Then G preserves the natural Lebesgue measure on $\mathbb{R}^2/\mathbb{Z}^2$ and the action is ergodic and mixing.*

SKETCH OF PROOF. Two ways: 1. Fourier analysis; 2. use the idea presented in last section (you need something contracted by the G action to make the argument work, what is this?).

$\square$

Let $M := \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$. For $t \in \mathbb{R}$, let $M^t := \exp(t \cdot \log M)$. The above example is about the induced action of $M^{\mathbb{Z}}$ on $\mathbb{R}^2/\mathbb{Z}^2$. The reason why you have such an induced action is of course $\mathbb{Z}^2$ is preserved by $M^{\mathbb{Z}}$. For other t, this is not true. Nevertheless, each M^t defines a homeomorphism

$$\mathbb{R}^2/\mathbb{Z}^2 \mapsto \mathbb{R}^2/M^t \cdot \mathbb{Z}^2.$$

Let

$$X = \left\{(x, t) \,\middle|\, t \in \mathbb{R}/\mathbb{Z},\ x \in \mathbb{R}^2/M^t \cdot \mathbb{Z}^2\right\}.$$

EXAMPLE 4.3. *Show that X has a natural measure m. Moreover, the action of $M^{\mathbb{R}}$ is ergodic but not mixing.*

This example tells you that in general an ergodic B-action (that is not extendable to an $\mathrm{SL}_2(\mathbb{R})$ p.m.p. action) may not be mixing. However, this B-action is not "totally ergodic" in the sense that some infinite subgroup does not act ergodically. I do not know an example of totally ergodic B-action that is not mixing. Note that by argument from the last section, it must be A-mixing.

5. Exercises

5.1. Non-commensurable lattices in $\mathrm{SL}_2(\mathbb{R})$, II. This is a continuation of Exercise 2.1–2.6 from Exercise Sheet 2. Notations are inherited and here are a few more:

- Let $X := \mathrm{G}/\Gamma$ and $\widehat{\mathrm{m}}_X$ the unique G-invariant probability measure on X;
- Let Ω be a nonempty open bounded subset of UV^+ (or UV^-);
- Let $\widetilde{\mu}_0$ be the restriction of the Haar measure on UV to Ω. Fix $x_0 \in X$, let μ_0 be the push-forward of $\widetilde{\mu}_0$ under the map $g \mapsto g.x_0$. By multiplying by a scalar, we normalize μ_0 to be a probability measure $\widehat{\mu}_0$.

EXERCISE 5.1. *Show that $\widehat{\mathrm{m}}_X$ is A-mixing.*

EXERCISE 5.2. *Using mixing to show that* $\lim_{t\to+\infty}(\mathbf{a}_t)_*\widehat{\mu}_0 = \widehat{\mathrm{m}}_X$.

EXERCISE 5.3. *Let* Y_0 *be as in Exer 2.3 from Exer. Sheet 2. Show that* $\mathrm{Y}_0 = X$.

Thus we have shown that H-orbits on X are either closed or dense.

Now let Γ_1, Γ_2 be two discrete subgroups in $\mathrm{SL}_2(\mathbb{R})$ (later we will assume them to be cocompact).

EXERCISE 5.4. *The following two are equivalent*

1. $\Gamma_1 \cdot \Gamma_2$ *is closed in* $\mathrm{SL}_2(\mathbb{R})$;
2. $\mathrm{H}\cdot(\Gamma_1 \times \Gamma_2)$ *is closed in* G.

EXERCISE 5.5. *The following two are equivalent*

1. $\Gamma_1 \cdot \Gamma_2$ *is dense in* $\mathrm{SL}_2(\mathbb{R})$;
2. $\mathrm{H}\cdot(\Gamma_1 \times \Gamma_2)$ *is dense in* G.

From now on we assume Γ_1, Γ_2 are both cocompact in $\mathrm{SL}_2(\mathbb{R})$.

EXERCISE 5.6. *The following two are equivalent*

1. $\Gamma_1 \cdot \Gamma_2$ *is closed in* $\mathrm{SL}_2(\mathbb{R})$;
2. Γ_1 *is commensurable with* Γ_2 *(namely,* $\Gamma_1 \cap \Gamma_2$ *is of finite-index in both* Γ_1 *and* Γ_2*).*

[It seems unclear to me how to prove this only assuming Γ_i's are lattices. There is an approach using random walk by Eskin–Margulis.]

EXERCISE 5.7. *The followings are equivalent*

1. Γ_1 *is commensurable with* Γ_2;
2. $\Gamma_1 \cdot [\mathrm{id}]_{\Gamma_2}$ *is a finite subset of* $\mathrm{SL}_2(\mathbb{R})/\Gamma_2$;
3. $\Gamma_1 \cdot \Gamma_2$ *is not dense in* $\mathrm{SL}_2(\mathbb{R})$.

5.2. Totally geodesic hyperbolic planes in H3, II. Notations and assumptions are inherited from Sec.3 from Exercise Sheet 2.

EXERCISE 5.8. *Show that* H*-orbits on* G/Γ *are either closed or dense.*

5.3. Mixing fails for non-semisimple groups. Notations

- $B = A \cdot U$ where $A := \left\{ \mathbf{a}_t = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}, t \in \mathbb{R} \right\}$ and $U = \left\{ \mathbf{u}_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}, s \in \mathbb{R} \right\}$;
- $\mathcal{H}$ is a separable Hilbert space and $\Phi : B \to \mathcal{U}(\mathcal{H})$ is a unitary representation of B.

EXERCISE 5.9. *Show that if $\mathcal{H}$ has no non-zero $\Phi(U)$-fixed vector ("U-ergodic"), then for every $\phi, \psi \in \mathcal{H}$ and $t_n \to +\infty$, $\lim_n \langle \Phi(\mathbf{a}_{t_n}).\phi, \psi \rangle = 0$ ("A^+-mixing").*

EXERCISE 5.10. *Same notations and assumptions as in last exercise. Show that for every $\phi, \psi \in \mathcal{H}$ and $t'_n \to -\infty$, $\lim_n \langle \Phi(\mathbf{a}_{t'_n}).\phi, \psi \rangle = 0$ ("A^--mixing").*

Below is an example showing that "U-mixing" may not be true under the hypothesis made in last two exercises.

Let $\mathcal{H}_0 := L^2(\mathbb{R}_{>0}, \mathrm{Leb})$. Define, for $t, s \in \mathbb{R}$ and $\phi \in \mathcal{H}_0$,

$$(\mathbf{a}_t.\phi)(x) := e^t \phi(e^{2t} x), \quad (\mathbf{u}_s.\phi)(x) := e^{2\pi i s x} \cdot \phi(x).$$

EXERCISE 5.11. *Show that the above defined action of A and U extends to a group homomorphism $\Phi_0 : B \to \mathrm{Hom}(\mathcal{H}_0, \mathcal{H}_0)$.*

Here $\mathrm{Hom}(\mathcal{H}_0, \mathcal{H}_0)$ stands for linear maps from $\mathcal{H}_0$ to $\mathcal{H}_0$.

EXERCISE 5.12. *Show that image of Φ_0 consists of unitary operators.*

EXERCISE 5.13. *Show that Φ_0 defines a unitary representation of B (namely, one should check continuity w.r.t. strong operator topology).*

EXERCISE 5.14. *Show directly that Φ_0 is A-mixing. Namely, for a divergent sequence $(a_n) \subset A$ and $\phi, \psi \in \mathcal{H}_0$, $\lim_n \langle \Phi_0(a_n).\phi, \psi \rangle = 0$.*

EXERCISE 5.15. *Show that there is no non-zero $\Phi_0(U)$-fixed vector. Yet Φ_0 is not U-mixing.*

5.4. Another example of Mautner phenomenon. Notations

- $N := \left\{ \begin{bmatrix} 1 & s & r \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix} \middle| s, t, r \in \mathbb{R} \right\}$, $Z := \left\{ \mathbf{z}_r := \begin{bmatrix} 1 & 0 & r \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \middle| r \in \mathbb{R} \right\}$;
- $W := \left\{ \mathbf{w}_t := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix} \middle| t \in \mathbb{R} \right\}$, $U := \left\{ \mathbf{u}_s := \begin{bmatrix} 1 & s & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \middle| s \in \mathbb{R} \right\}$;
- $\mathcal{H}$ is a separable Hilbert space and $\Phi : N \to \mathcal{U}(\mathcal{H})$ is a unitary representation of N.

EXERCISE 5.16. *Verify the following*

$$\mathbf{w}_t \mathbf{u}_s \mathbf{w}_{-t} = \mathbf{u}_s \mathbf{z}_{-st}, \ \forall s, t \in \mathbb{R}.$$

EXERCISE 5.17. *Show that a $\Phi(W)$-fixed vector is $\Phi(Z)$-fixed.*

[Since $W \cdot Z$ is a normal subgroup of N with quotient group $\mathbb{R}$, there exists a unitary representation $(\Phi, \mathcal{H})$ of N and $v \in \mathcal{H}$ such that its stabilizer in N is exactly $W \cdot Z$.]

Now let Γ be a lattice in N.

EXERCISE 5.18. *Show that Γ is not commutative, and hence, not virtually commutative (namely, every finite-index subgroup of Γ is not commutative).*

EXERCISE 5.19. *Show that $\Gamma \cap Z$ is a lattice in Z.*

Let $p : N \to N/Z$ (Z is normal in N) be the natural quotient map.

EXERCISE 5.20. *Show that $p(\Gamma)$ is a lattice of N/Z.*

Let $\widehat{\mathrm{m}}_X$ be the N-invariant probability measure on N/Γ and let $\widehat{\mathrm{m}}_{\overline{X}}$ be the N/Z-invariant probability measure on $(N/Z)/p(\Gamma)$.

EXERCISE 5.21. *Show that* $\widehat{\mathrm{m}}_X$ *is* W*-ergodic iff* $\widehat{\mathrm{m}}_{\overline{X}}$ *is* W*-ergodic.*

EXERCISE 5.22. *Fix* Γ*, show that there exists some one-parameter unipotent subgroup* $\{\mathbf{v}_s\}$ *of* N *that acts ergodically on* $\widehat{\mathrm{m}}_X$*.*

One more example.

Let $G := \left\{ \begin{bmatrix} a & b & x \\ c & d & y \\ 0 & 0 & 1 \end{bmatrix} \middle| \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{R}),\ x, y \in \mathbb{R} \right\}$.

$\Gamma := \left\{ \begin{bmatrix} a & b & x \\ c & d & y \\ 0 & 0 & 1 \end{bmatrix} \middle| \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z}),\ x, y \in \mathbb{Z} \right\}$.

EXERCISE 5.23. *Use mixing and non-divergence of unipotent flow to show that* $\mathrm{SL}_2(\mathbb{Z})$ *is a lattice in* $\mathrm{SL}_2(\mathbb{R})$*.*

EXERCISE 5.24. *Show that* Γ *is a lattice in* G*.*

Let $\widehat{\mathrm{m}}_{G/\Gamma}$ be the unique G-invariant probability measure on G/Γ.

EXERCISE 5.25. *Show that* $\widehat{\mathrm{m}}_{G/\Gamma}$ *is* $\mathrm{SL}_2(\mathbb{R})$*-ergodic.*

Here we embed $\mathrm{SL}_2(\mathbb{R})$ in the left upper corner of G. By what has been proved in the class, this implies that $\widehat{\mathrm{m}}_{G/\Gamma}$ is $\mathrm{SL}_2(\mathbb{R})$-mixing.

CHAPTER 8

Unipotent invariant finite measures on quotients of $\mathrm{SL}_2(\mathbb{R})$, I

Back to the Top.

In this chapter we introduce pointwise ergodic theorem and start our discussion of classification of unipotent invariant probability measures on quotients of $\mathrm{SL}_2(\mathbb{R})$.

1. Ergodicity and extremality

Unless otherwise specified, we assume G and X are nice. So G is a locally compact and σ-compact metrizable group and X is a σ-compact locally compact metrizable space. The space of probability measures $\mathrm{Prob}(X)$ with the weak$*$ topology is not necessarily compact (unless X is compact) but the $\mathrm{Meas}(X)^{\leq 1}$ is. And $\mathrm{Meas}(X)^{\leq 1}$, $\mathrm{Prob}(X)$ and $\mathrm{LFM}(X)$ are also nice (decompose X into countable union of compact pieces and consider probability or finite measures supported on finite unions of them).

LEMMA 1.1. *A G-invariant probability measure μ is ergodic iff it is extremal in the space of G-invariant probability measures. Or more succinctly,* $\mathrm{Prob}(X)^{G,\mathrm{Erg}} = \mathrm{Extre}(\mathrm{Prob}(X)^G)$.

Being **extremal** means that μ can not be written as convex combination of different invariant probability measures. That is to say, if $\mu = a\nu_1 + (1-a)\nu_2$ for some $a \in (0,1)$ and $\nu_i \in \mathrm{Prob}(X)^G$, then $\nu_1 = \nu_2 = \mu$. In particular, two different ergodic μ_1, μ_2 must be singular w.r.t. each other. Namely, we may partition $X = A \sqcup B$ into two measurable parts such that $\mu_1(B) = 0$ and $\mu_2(A) = 0$.

SKETCH OF PROOF. If μ is not ergodic, then we can pick two complementary invariant measurable sets. Then μ is the sum of the restriction of μ to these two sets and is not extremal. Conversely, if $\mu = a\nu_1 + (1-a)\nu_2$ then ν_1 and ν_2 are absolutely continuous w.r.t. μ. So we find two G-invariant $L^1(\mu)$-functions representing "$\frac{\mathrm{d}\nu_i}{\mathrm{d}\mu}$" which are forced to be constants unless μ is not ergodic. □

By general facts from functional analysis (Hahn–Banach theorem), the convex combinations of $\mathrm{Extre}(\mathrm{Prob}(X)^G)$ are dense in $\mathrm{Prob}(X)^G$ (pretend X to be compact first and then do the general case). A theorem of Choquet says that more precisely (See Thm.4.8 and 8.20 of the book of Einsiedler–Ward [EW11]),

THEOREM 1.2 (Ergodic decomposition). *For every $\mu \in \mathrm{Prob}(X)^G$ there exists a unique Borel probability measure $\lambda \in \mathrm{Prob}(\mathrm{Prob}(X)^G)$ such that*

- $\lambda(\mathrm{Prob}(X)^{G,\mathrm{Erg}}) = 1$;
- $\mu = \int_{\nu \in \mathrm{Prob}(X)^{G,\mathrm{Erg}}} \nu \, \lambda(\nu)$.

Let me add that $\mathrm{Prob}(X)^{G,\mathrm{Erg}}$ is not closed in general (Exercise: find such an example) but in the world of unipotent flows, this is closed due to a theorem of Mozes–Shah.

In virtue of this theorem, to classify invariant probability measures, we often start with ergodic ones.

2. Pointwise ergodic theorem for a flow

We can construct a new invariant probability measure from known ones by convex combination. But how to get one to start with? Well, in general such a measure may not exist (say, the $\mathrm{SL}_2(\mathbb{R})$-action on the space of lines of $\mathbb{R}^2$). But for a flow, namely a continuous $\mathbb{R}$-action (denote the action $\mathbb{R} \times X \to X$ by $(t,x) \mapsto T_t.x$) on a nice X, we can consider

$$\frac{1}{T}\int_0^T (T_t)_*\delta_x \,\mathrm{dt} = \frac{1}{T}\int_0^T \delta_{T_t.x}\,\mathrm{dt}$$

as $T \to +\infty$. Here δ_x denotes the measure defined by $\delta_x(E) = 1$ iff $x \in E$ and is zero otherwise. You can replace the δ-measure supported on $\{x\}$ by any other probability measure. Using this construction, one shows that

LEMMA 2.1. *Let $(T_t)_{t\in\mathbb{R}}$ be a flow on X. If further assume X is compact, then there exists a (T_t)-invariant probability measures.*

Conversely, every ergodic flow-invariant probability measure may be constructed this way from a delta measure. Actually, more is true. This is the pointwise ergodic theorem.

THEOREM 2.2. *Let T_t denote the action of $\mathbb{R}$ on a nice space X. Let μ be an ergodic Borel probability (T_t)-invariant measure on X. Then for every $f \in L^1(X, \mathscr{B}_X, \mu)$ there exists a measurable set E_f of full measure ($\mu(E_f) = 1$) such that for every $x \in E_f$ we have*

$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T f(T_t.x)\,\mathrm{dt} = \int f(x)\mu(x). \tag{27}$$

Using the fact that $C_c(X)$ admits a countable dense subset for a nice X, a diagonal argument shows that

COROLLARY 2.3. *Assumption as in the above theorem. There exists a full measure set E such that for every $x \in E$,*

$$\lim_{T\to+\infty} \frac{1}{T}\int_0^T (T_t)_*\delta_x\,\mathrm{dt} = \mu \tag{28}$$

where the limit is taken with respect to the weak$*$ topology.

There is no such general ergodic theorem beyond the world of amenable groups.

A point x satisfying Equa.(27) (or (28)) is sometimes called f-**generic** (or generic). To emphasize both the group action and the invariant measure, one may also call x a (T_t, μ)-generic point. In general, it may be very difficult to describe the set of generic points. One beauty of unipotent flows is that you do have an explicit description of generic points in this case.

3. Ergodic measures for unipotent flows

Let us start with the easiest case.

DEFINITION 3.1. *Given a continuous action of G on X. We say that the action is **uniquely ergodic** iff the action admits a unique invariant probability measure.*

THEOREM 3.2. *Let Γ be a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$, then there exists a unique up-to-a-scalar $\mathrm{SL}_2(\mathbb{R})$-invariant locally finite measure m_{X} on $\mathrm{X} := \mathrm{SL}_2(\mathbb{R})/\Gamma$.*

A reference is Raghunathan's book [**Rag72**]. For existence and uniqueness of invariant (Haar) measures on a (nice) topological group, one may consult [**DE14**].

Thus when such a measure is finite, we get an example of uniquely ergodic action.

LEMMA 3.3. *Assume a flow T_t on a compact space X is uniquely ergodic with the unique invariant probability measure denoted by μ, then for every $x \in \mathrm{X}$, Equa.(28) holds.*

Proof is left as an exercise. You can not drop the compactness assumption.

Now we go to the world of unipotent flows. Some notations:

- $\mathrm{G} := \mathrm{SL}_2(\mathbb{R})$, Γ is a discrete subgroup of G and $\mathrm{X} := \mathrm{G}/\Gamma$.
- $\mathrm{U} := \left\{ \mathbf{u}_s := \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle| s \in \mathbb{R} \right\}$, $\mathrm{A} := \left\{ \mathbf{a}_t := \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix} \middle| t \in \mathbb{R} \right\}$;
- $\mathrm{B} := \mathrm{A} \cdot \mathrm{U}$.

THEOREM 3.4. *Assume* Γ *is cocompact in* G*. Then the* U*-action on* X *is uniquely ergodic.*

The existence is guaranteed. One needs to prove the uniqueness. This result is due to Furstenberg [**Fur73**].

A more general result is

THEOREM 3.5. *Let* Γ *be a discrete subgroup of* G*. Then every* $\mu \in \mathrm{Prob}(\mathrm{X})^{\mathrm{U,Erg}}$ *is one of the following:*

1. *supported on a closed (necc. compact)* U*-orbit;*
2. $\mathrm{m_X}/|\mathrm{m_X}|$ *with* $|\mathrm{m_X}| < \infty$.

In particular, if X has no compact U-orbit and $|\mathrm{m_X}| = \mathrm{m_X}(\mathrm{X})$ is not finite, then there is no finite U-invariant measure. Though this does not prevent the existence of dynamically interesting infinite U-invariant measures.

A reference for the material presented here is Ratner's paper [**Ra92**].

4. Outline of the proof and step 1

The proof of Thm.3.5 to be presented here consists of two parts

Step 1. Upgrade from U-invariance to B-invariance if the measure is not supported on a compact U-orbit.

Step 2. Show that the action of B is uniquely ergodic unless $\mathrm{m_X}$ is infinite.

The first step is essentially achieved by a combination of ideas from Ch.1 and pointwise ergodic theorem. It might be possible to do the second step by a duality argument in the style of Ch.1. We will do something different.

4.1. Step 1. Compared to Ch.1 we will do the following adjustment

$$\text{compact topological spaces} \longrightarrow \text{probability invariant measures}$$
$$\text{minimal sets} \longrightarrow \text{generic points}$$

We shall actually use compact subsets of generic points so that we can take limits.

LEMMA 4.1. *Let* μ *be an ergodic* U*-invariant probability measure on* X*, then*

1. *either* μ *is supported on a closed* U*-orbit;*
2. *or* μ *is* B*-invariant.*

Before the proof we make the following observation

LEMMA 4.2. *If* x, y *are both* (U,μ) *generic points and* $y = g.x$ *with* $g \in \mathrm{G}$ *normalizing* U*, then* $g_*\mu = \mu$.

PROOF. Since $g \in \mathrm{G}$ normalizes U, we find some constant $c_g > 0$ such that $g\mathbf{u}_t g^{-1} = \mathbf{u}_{c_g t}$. By definition of genericity we have

$$\begin{aligned} g_*\mu &= g_* \lim_{T\to+\infty} \frac{1}{T}\int_{t=0}^{T} (\mathbf{u}_t)_*\delta_x \,\mathrm{dt} = \lim_{T\to+\infty} \frac{1}{T}\int_{t=0}^{T} g_*(\mathbf{u}_t)_*\delta_x \,\mathrm{dt} \\ &= \lim_{T\to+\infty} \frac{1}{T}\int_{t=0}^{T} (\mathbf{u}_{c_g\cdot t})_* g_*\delta_x \,\mathrm{dt} = \lim_{T\to+\infty} \frac{1}{T}\int_{t=0}^{T} (\mathbf{u}_{c_g\cdot t})_*\delta_y \,\mathrm{dt} = \mu. \end{aligned}$$

□

PROOF OF LEMMA 4.1. Without loss of generality assume μ is not supported on a closed U-orbit. In light of Lem.4.2 above, we hope to find a pair x, y that are both (U,μ)-generic and $y = \mathbf{a}_t.x$ with $t \neq 0$; and by varying the pair, we want t to be arbitrarily close to 0.

Recall that the argument from Ch.1 basically goes like:

- Step 1. find two sequences (x_n) and (y_n) with $d(x_n, y_n) \to 0$ and for each n, x_n and y_n are not on the same local U-orbit;
- Step 2. if for infinitely many n, x_n and y_n are on the same local B-orbit, then we are done;
- Step 3. otherwise, depending on $\delta > 0$, we find s_n, t_n such that every limit pair (x_∞, y_∞) of $x'_n := \mathbf{u}_{t_n} x_n$ and $y'_n := \mathbf{u}_{s_n} y_n$ are differed by some $\mathbf{a}_t$ with $t \in [C^{-1}\delta, C\delta]$ for some constant $C > 1$;
- Step 4. as a complement to Step 3, it should be noted that the choice of s_n is determined by t_n and the choice of t_n has the freedom of multiplying by a (multiplicatively) bounded number. This havs the effect of changing the C in step 3 by another C';
- Step 5. so far we have demonstrated $\mathbf{a}_t$ with $|t| \to 0, t \neq 0$ with $\mathbf{a}_t \in \mathrm{G}_\mu$, the stabilizer of μ in G. Since G_μ is a closed subgroup, $A \subset \mathrm{G}_\mu$.

Below is a detailed account of carrying out the above strategy in the measure theoretic setting. You may try to figure out how by yourself.

We need to guarantee the limits (x_∞, y_∞) to be generic. Since the set of generic points is usually not closed, we define

$$E_{\mathrm{U},\mu} := \left\{ (\mathrm{U},\mu)\text{-generic points} \right\}$$

and take E to be a compact subset of $E_{\mathrm{U},\mu}$ such that $\mu(E) > 0.9$.

Take T_0 large enough such that the following set

$$F := \left\{ x \in \mathrm{X} \,\middle|\, \frac{1}{T} \operatorname{Leb}\{t \in [0,T], \mathbf{u}_t.x \in E\} \geq 0.9,\ \forall T \geq T_0 \right\} \tag{29}$$

has $\mu(F) > 0.9$. (how? First by ptws ergodic theorem applied to the indicator function of E, we see that

$$\left\{ x \in \mathrm{X} \,\middle|\, \lim_{T\to+\infty} \frac{1}{T} \operatorname{Leb}\{t \in [0,T], \mathbf{u}_t.x \in E\} = \mu(E) > 0.9 \right\}$$

has full measure 1. Thus as S varies over positive integers, the increasing union of the following sets

$$F_S := \left\{ x \in \mathrm{X} \,\middle|\, \frac{1}{T} \operatorname{Leb}\{t \in [0,T], \mathbf{u}_t.x \in E\} > 0.9,\ \forall T \geq S \right\}$$

has measure 1. Thus we can find some T_0 such that F_{T_0} has measure at least 0.9.)

We claim that there exist pairs (x, y) in F arbitrarily close to each other and yet not on the same local U-orbit (unless μ is supported on a compact U-orbit, which by assumption does not happen).

To be precise, two points x, y are said to be on the same *local* U-orbit if $x = \mathbf{u}_s.y$ for some $s \in (-1,1)$.

If the claim were not true, then there exists $\varepsilon > 0$ such that if $x, y \in F$ and $d(x,y) < \varepsilon$ then $x = \mathbf{u}_s y$ for some $|s| < 1$. Cover F by countably many measurable sets $\{B_i\}$ of diameter smaller than ε. Then $B_i \cap F \subset \mathbf{u}_{(-1,1)}.x_i$ for some x_i. So $F \subset \bigcup_i \mathbf{u}_{(-1,1)}.x_i$. Thus for some x_i,

$$\mu(\{\mathbf{u}_s.x_i \mid |s| < 1\}) > 0.$$

By ergodicity this implies that U.x has to close up (you can invoke ptws ergodic theorem to prove this but you do not have to) and that μ is the U-invariant measure supported on this orbit. Contradiction.

Recall calculation from Ch.1, with $s+t$ replaced by t,

$$u_t A_n u_s^{-1} = \begin{bmatrix} 1+a_n+tc_n & b_n+s(d_n-a_n)-s^2c_n+(t-s)(1+d_n-sc_n) \\ c_n & 1+d_n-sc_n \end{bmatrix} \tag{30}$$

for

$$A_n = \begin{bmatrix} 1+a_n & b_n \\ c_n & 1+d_n \end{bmatrix} \quad \text{with } a_n, b_n, c_n, d_n \to 0.$$

Compared to Ch.1, let us make a little adjustment on the choice of $s_{n,\delta}$ and particularly $t_{n,\delta}$ to simplify matters. Assume $c_n \neq 0$. For a small number $\delta > 0$, choose $s_{n,\delta}$ as before, namely,

$$s_{n,\delta} := \frac{d_n+\delta}{c_n}$$

(choosing $s = (d_n - \delta)/c_n$ is also ok). We also need an additional parameter $\lambda = \lambda_{n,\delta} \in (0.1, 1)$ to be determined in a moment. Let $s'_{n,\delta} := \lambda_{n,\delta} s_{n,\delta}$. Choose $t'_{n,\delta} := \phi_{n,\delta}(\lambda_{n,\delta}) \cdot s'_{n,\delta}$ where

$$\phi_{n,\delta}(\lambda_{n,\delta}) := \frac{a_n-(1-\lambda_{n,\delta})d_n+\lambda_{n,\delta}\delta}{1+(1-\lambda_{n,\delta})d_n-\lambda_{n,\delta}\delta}+1 = \frac{(a_n-d_n)+(d_n+\delta)\lambda_{n,\delta}}{(1+d_n)-(d_n+\delta)\lambda_{n,\delta}}+1. \tag{31}$$

This choice is such that the upper right corner of Equa.(30) converges to 0 asymptotically. Indeed with $\lambda = \lambda_{n,\delta}$, $s = s'_{n,\delta}$ and $t = t'_{n,\delta}$ we have

$$\begin{aligned} & b_n+s(d_n-a_n)-s^2c_n+(t-s)(1+d_n-sc_n) \\ (sc_n = d_n+\delta) \quad =& b_n+s(d_n-a_n)-s\lambda(d_n+\delta)+(t-s)(1+d_n)-(t-s)\lambda(d_n+\delta) \\ (\text{Equa.}(31)) \quad =& b_n+s(-a_n+d_n-\lambda d_n-\lambda\delta)+(t-s)(1+d_n-\lambda d_n-\lambda\delta) = b_n \to 0. \end{aligned}$$

Let us firstly cheat by assuming $\phi_{n,\delta}(\lambda_{n,\delta}) \equiv 1$ in Equa.(31). See Sec.4.2 below to see the true proof. We have

$$s'_{n,\delta} = t'_{n,\delta} = \lambda_{n,\delta} s_{n,\delta}.$$

Take $\delta > 0$ and n large enough such that $s_{n,\delta} > T_0$. Then by the definition of F (see Equa.(29)),

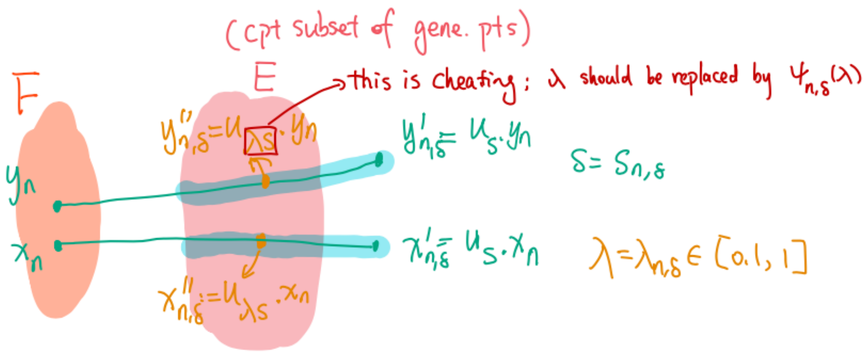
$$\begin{aligned} &\mathrm{Leb}\left(\left\{\lambda \in (0.1,1) \,\middle|\, \mathbf{u}_{\lambda s_{n,\delta}}.x_n \in E\right\}\right) > 0.9-0.1 = 0.8; \\ &\mathrm{Leb}\left(\left\{\lambda \in (0.1,1) \,\middle|\, \mathbf{u}_{\lambda s_{n,\delta}}.y_n \in E\right\}\right) > 0.9-0.1 = 0.8. \end{aligned} \tag{32}$$

In particular, their intersection is nonempty and we take some element $\lambda_{n,\delta}$. Define $x''_{n,\delta} := \mathbf{u}_{s'_{n,\delta}}.x_n$ and $y''_{n,\delta} := \mathbf{u}_{s'_{n,\delta}}.y_n$, then $x''_{n,\delta}, y''_{n,\delta} \in E$. By letting $n \to +\infty$ (pass to a subsequence if necessary) and by Equa.(30) above, we get

$$y_{\infty,\delta} = \begin{bmatrix} (1-\lambda_{\infty,\delta}\delta)^{-1} & 0 \\ 0 & 1-\lambda_{\infty,\delta}\delta \end{bmatrix} x_{\infty,\delta}$$

where $x_{\infty,\delta} := \lim x''_{n,\delta} \in E$, $y_{\infty,\delta} := \lim y''_{n,\delta} \in E$ and $\lambda_{\infty,\delta} := \lim \lambda_{n,\delta} \in [0.1, 1]$. So we get a sequence of non-identity elements in A converging to id that maps some generic point $(x_{\infty,\delta})$ to another one $(y_{\infty,\delta})$. By Lem.4.2, they are contained in G_μ, which is a closed subgroup. Thus A is contained in G_μ and the proof completes.

Finally, here is a summary-by-picture:



By def. of F, we can choose $\lambda = \lambda_{n,\delta} \in [0.1, 1]$ s.t. $x''_{n,\delta}, y''_{n,\delta}$ lie in E, a compact subset of generic points.

rmk: you may think of F as contained in E if you like

□

4.2. To avoid cheating... Here is the honest proof.

As functions on $[0,2]$ indexed by n, δ, we can check that as $n \to \infty$ and $\delta \to 0$, the functions $\phi_{n,\delta}$ (resp., their derivatives) converge to the constant 1 (resp., 0) uniformly. Thus for n sufficiently large and δ sufficiently small, we may and do assume that

$$\phi_{n,\delta}(\lambda) \in [0.99, 1.01], \quad \phi'_{n,\delta}(\lambda) \in [-0.01, 0.01], \quad \forall \lambda \in [0,2].$$

Let $\psi_{n,\delta}(\lambda) := \phi_{n,\delta}(\lambda) \cdot \lambda$. For n large and δ small,

$$\psi'_{n,\delta}(\lambda) = \phi'_{n,\delta}(\lambda) \cdot \lambda + \phi_{n,\delta}(\lambda) \in [0.97, 1.03], \quad \forall \lambda \in [0,2]. \tag{33}$$

So $\psi = \psi_{n,\delta}$ defines a diffeomorphism from $[0.1, 1] \to \psi([0.1, 1])$. Note that

$$[0.15, 0.95] \subset \psi([0.1, 1]) \subset [0.05, 1.05]. \tag{34}$$

Let (abbr. $\psi := \psi_{n,\delta}$ and $s := s_{n,\delta}$)

$$\begin{aligned} & A := \left\{\lambda \in [0.1, 1] \,\middle|\, \mathbf{u}_{\psi(\lambda)s}.y_n \in E\right\} \\ & \Longrightarrow \psi(A) = \left\{\lambda \in \psi([0.1, 1]) \,\middle|\, \mathbf{u}_{\lambda s}.y_n \in E\right\}; \\ & B := \{\lambda \in [0.1, 1] \mid \mathbf{u}_{\lambda s}.x_n \in E\}. \end{aligned} \tag{35}$$

By Equa.(32) and (34), we have $\mathrm{Leb}(\psi(A)), \mathrm{Leb}(B) \geq 0.8 - 0.1 = 0.7$. Also, by Equa.(33), for every $z \in A$, $\left|\psi'(z)\right|^{-1} \geq 1.03^{-1}$. Therefore,

$$\begin{aligned} \mathrm{Leb}(A) &= \int 1_A(x)\,\mathrm{d}x = \int 1_A(\psi^{-1}y)\left|(\psi^{-1})'(y)\right| \mathrm{d}y \\ &= \int 1_{\psi(A)}(y)\left|\psi'(\psi^{-1}y)\right|^{-1} \mathrm{d}y \geq 0.7 \cdot 1.03^{-1} \geq 0.6. \end{aligned}$$

Consequently, $A \cap B \neq \emptyset$ and we choose some $\lambda_{n,\delta} \in A \cap B$. As above, define $s'_{n,\delta} := \lambda_{n,\delta} \cdot s_{n,\delta}$ and $t'_{n,\delta} = \phi_{n,\delta}(\lambda_{n,\delta}) \cdot s'_{n,\delta} = \psi_{n,\delta}(\lambda_{n,\delta}) \cdot s_{n,\delta}$. By Equa.(35), $x''_{n,\delta} := \mathbf{u}_{s'_{n,\delta}}.x_n$ and $y''_{n,\delta} := \mathbf{u}_{s'_{n,\delta}}.y_n$ belongs to E. The rest of the proof is the same as those below Equa.(32).

Now we have completed Step 1.

5. Exercises

5.1. Lattices and closedness of orbits.

- G is a connected Lie group and Γ is a discrete subgroup of G;
- $H \le G$ is a closed subgroup.

EXERCISE 5.1. *Assume $H \cap \Gamma$ is a lattice in H. Show that for a divergent sequence (x_n) in $H/H \cap \Gamma$,* $\mathrm{InjRad}(x_n) \to 0$.

EXERCISE 5.2. *Assume Γ satisfies the conclusion of the last exercise. Show that $H\Gamma/\Gamma$ is closed in G/Γ.*

- $U = \left\{ \mathbf{u}_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}, s \in \mathbb{R} \right\}$, Γ is a discrete subgroup of $\mathrm{SL}_2(\mathbb{R})$.

EXERCISE 5.3. *Assume $U \cap \Gamma$ is cocompact in U, by duality we know that $\Gamma U/U$ is closed in $\mathrm{SL}_2(\mathbb{R})/U$. The latter is homeomorphic to $\mathbb{R}^2 - (0,0)$ under $g \mapsto g.e_1$. Thus $\Gamma.e_1$ is closed in $\mathbb{R}^2 - (0,0)$. Show that, in fact, $\Gamma.e_1$ is closed in $\mathbb{R}^2$.*

EXERCISE 5.4. *Show that the conclusion might fail if we replace "$U \cap \Gamma$ is cocompact in U" by "$U\Gamma$ is closed in $\mathrm{SL}_2(\mathbb{R})$".*

EXERCISE 5.5. *Show that $B = A \cdot U$ with $A := \left\{ \mathbf{a}_t = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}, t \in \mathbb{R} \right\}$ has no lattice.*

5.2. More exercises.

EXERCISE 5.6. *Let Γ be a lattice in $\mathrm{SL}_2(\mathbb{R})$, and assume Γ is not cocompact in $\mathrm{SL}_2(\mathbb{R})$. Let $X := \mathrm{SL}_2(\mathbb{R})/\Gamma$. Let d be a right invariant Riemannian metric on $\mathrm{SL}_2(\mathbb{R})$, which induces a quotient Riemannian metric d_X on X, from which we can define a (volume) measure on X. Accept the fact that such a measure is necessarily the $\mathrm{SL}_2(\mathbb{R})$-invariant finite measure on X. Show that a sequence $(x_n) \subset X$ goes to ∞ iff $\mathrm{InjRad}(x_n) \to 0$ as $n \to \infty$.*

EXERCISE 5.7. *Assume the notations and the conclusion of the exercise above. Show that $(g_n\Gamma/\Gamma) \subset X$ goes to ∞ iff there exists $\gamma_n \in \Gamma$ such that $\mathrm{dist}(\mathrm{id}, g_n\gamma_n g_n^{-1}) \to 0$.*

EXERCISE 5.8. *For a matrix $X = (x_{i,j})$, let $\|X\|_{\sup} := \sup_{i,j} |x_{i,j}|$. By a direct computation, show that there exists a constant $C > 0$, such that for every $\varepsilon > 0$ and $X, Y \in \mathrm{SL}_2(\mathbb{R})$ with $\|\mathrm{id} - X\| \le \varepsilon$ and $\|\mathrm{id} - Y\| \le \varepsilon$, we have that*

$$\left\|\mathrm{id} - XYX^{-1}Y^{-1}\right\| \le C \cdot \varepsilon^2.$$

EXERCISE 5.9. *Notations as in the exercise above. Show that there exists a neighborhood $\mathcal{N}$ of id in $\mathrm{SL}_2(\mathbb{R})$ such that for every discrete subgroup $\Gamma \le \mathrm{SL}_2(\mathbb{R})$, $\Gamma \cap \mathcal{N}$ generates an abelian group.*

EXERCISE 5.10. *Notations as in the exercise above. Show that there exists a neighborhood $\mathcal{N}'$ of id in $\mathrm{SL}_2(\mathbb{R})$ such that for every discrete subgroup $\Gamma \le \mathrm{SL}_2(\mathbb{R})$, there exists $g \in \mathrm{SL}_2(\mathbb{R})$ such that $g\Gamma g^{-1} \cap \mathcal{N}' = \{\mathrm{id}\}$.*

EXERCISE 5.11. *Let Γ in $\mathrm{SL}_2(\mathbb{R})$ be a lattice. Use previous exercises to show that Γ is not cocompact iff it contains non-identity unipotent matrices.*

REMARK 5.1. *The "if" direction is proved in the class. This is a special instance of Kazhdan–Margulis theorem.*

EXERCISE 5.12. *Let $a_t := \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$ and $u_s := \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$. In the class we have seen that for a discrete subgroup $\Gamma \le \mathrm{SL}_2(\mathbb{R})$, if $x \in \mathrm{SL}_2(\mathbb{R})/\Gamma$ belongs to a compact u_s-orbit, then $a_t.x$ diverges as t goes to $-\infty$. Now assume Γ is a lattice. Show that the converse holds. Namely, if $a_t.x$ diverges as t goes to $-\infty$, then $\{u_s.x\}_{s\in\mathbb{R}}$ is compact.*

CHAPTER 9

Unipotent invariant finite measures on quotients of $\mathrm{SL}_2(\mathbb{R})$, II

Back to the Top.

In this chapter we complete "step 2" from the last chapter. To do this we need the notion of conditional measures.

1. Conditional measures

As a reference, see [EW11, Ch.5] and [Cou16, Part IV and Ch.17].

Let X be a nice space and $\mathscr{B}_X$ be its Borel σ-algebra. Let $\mu \in \mathrm{Prob}(X)$. Let $\mathscr{A}$ be a *countably generated* sub-σ-algebra of $\mathscr{B}_X$. Being countably generated means that, by definition, there exists a countable collection of measurable subsets $\mathrm{Gene}(\mathscr{A}) = \{A_i\}$ of X such that $\mathscr{A}$ is the smallest sub-σ-algebra containing them. Assume the complement of every A_i is also contained in $\mathrm{Gene}(\mathscr{A})$. For $x \in X$, let the **atom** containing x be $[x]^{\mathscr{A}} := \bigcap_{A_i \ni x} A_i$.

LEMMA 1.1. *Actually $[x]^{\mathscr{A}} = \bigcap_{A \ni x,\, A \in \mathscr{A}} A$, hence $[x]^{\mathscr{A}}$ is independent of the choice of a (symmetric) countable generator $\mathrm{Gene}(\mathscr{A})$.*

Proof is left as an exercise.

THEOREM 1.2. *(Conditional measures) Let $(X, \mathscr{B}_X, \mu)$ and $\mathscr{A}$ be as above.*

1. *Existence of conditional measures.*

There exists $X' \in \mathscr{A}$ of full measure and a measurable map $X' \to \mathrm{Prob}(X)$ denoted as $x \mapsto \mu_x^{\mathscr{A}}$ such that $\mu_x^{\mathscr{A}}([x]^{\mathscr{A}}) = 1$ and

$$\int_A \int f(y)\mu_x^{\mathscr{A}}(y)\mu(x) = \int_A f(x)\mu(x) \tag{36}$$

for every $A \in \mathscr{A}$ and $f \in L^1(X, \mathscr{B}_X, \mu)$. Implicitly we have claimed that $x \mapsto \int f(y)\mu_x^{\mathscr{A}}(y)$ is integral on A.

2. *Uniqueness of conditional measures.*

If $x \mapsto \nu_x^{\mathscr{A}}$ is another measurable map from a possibly different full measure subset X'' to $\mathrm{Prob}(X)$ *satisfying Equa.(36) for every compactly supp. cont. function $f \in C_c(X)$ and $A = X''$, then for some full measure set $X''' \subset X' \cap X''$ we have $\mu_x^{\mathscr{A}} = \nu_x^{\mathscr{A}}$ for $x \in X'''$.*

EXAMPLE 1.3. *Let $\mathscr{A} = \mathscr{B}_X$. Then $[x]^{\mathscr{A}} = \{x\}$ and $\mu_x^{\mathscr{A}} = \delta_x$ for every $x \in X$.*

EXAMPLE 1.4. *Let $\mathscr{A}$ be the sigma algebra generated by a finite measurable partition $\{P_1, ..., P_l\} \subset \mathscr{B}_X$ of X, then $[x]^{\mathscr{A}} = P_i$ iff $x \in P_i$ and $\mu_x^{\mathscr{A}} = \frac{\mu|_{P_i}}{\mu(P_i)}$.*

EXAMPLE 1.5. *Let $X = [0,1] \times [0,1]$ and $\mu = \mathrm{Leb}$ be the standard Lebesgue measure defined by $|\mathrm{d}x \wedge \mathrm{d}y|$. Let $\mathscr{A} := \left\{A \times [0,1] \,\middle|\, A \in \mathscr{B}_{[0,1]}\right\}$. Then for every $(x,y) \in X$, $[(x,y)]^{\mathscr{A}}_{(x,y)} = \{x\} \times [0,1]$ and $\mu^{\mathscr{A}}_{(x,y)}$ is induced by $|\mathrm{d}y|$.*

This example can be generalized to foliations on manifolds where X is a small open set with a local foliation chart, which provides $\mathscr{A}$.

EXAMPLE 1.6. *Everything same as in the last example except that we let μ be the standard Lebesgue measure supported on $\Delta := \{(x,x),\ x \in [0,1]\}$. Then $[(x,y)]^{\mathscr{A}}_{(x,y)} = \{x\} \times [0,1]$ and $\mu^{\mathscr{A}}_{(x,y)} = \delta_y$.*

EXAMPLE 1.7. *If you have a probability measure preserving map $\pi : (X, \mathscr{B}_X, \mu) \to (Y, \mathscr{B}_Y, \nu)$ with X, Y nice. Let $\mathscr{A} := \pi^{-1}\mathscr{B}_Y$. In this case, Equa.(36) can be viewed as a "fibre integration formula" (you can replace the μ on the LHS by ν). Here atoms are fibres of π. In some sense, all countably generated sub σ-algebra $\mathscr{A}$ arises from such a π.*

EXAMPLE 1.8. *Let G, X both be nice and assume G preserves μ. Let*

$$\mathscr{A} := \{A \in \mathscr{B}_X \mid A \text{ is almost } G\text{-invariant}\}.$$

A measurable subset A is almost G-invariant if $\mu(g.A \Delta A) = 0$ for all $g \in G$. Then Equa.(36) provides an explicit form of ergodic decomposition.

2. Step 2 of the measure classification

Notations:

- $\mathrm{G} := \mathrm{SL}_2(\mathbb{R})$, Γ is a discrete subgroup of G and $\mathrm{X} := \mathrm{G}/\Gamma$.
- let $\mathrm{m_X}$ be a G-invariant locally finite measure on X and let $\widehat{\mathrm{m}}_{\mathrm{X}} := \frac{\mathrm{m_X}}{\mathrm{m_X(X)}}$ if $\mathrm{m_X(X)} < +\infty$;
- $\mathrm{U} := \left\{\mathbf{u}_s := \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \,\middle|\, s \in \mathbb{R}\right\}$, $\mathrm{A} := \left\{\mathbf{a}_t := \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix} \,\middle|\, t \in \mathbb{R}\right\}$;
- $\mathrm{B} := \mathrm{A} \cdot \mathrm{U}$;
- $\mathrm{V} := \left\{\mathbf{v}_s := \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix} \,\middle|\, s \in \mathbb{R}\right\}$.

Recall that we are left to prove the following.

THEOREM 2.1. *If there exists a* B*-invariant,* U*-ergodic probability measure μ on* X*, then Γ is a lattice and μ is equal to $\widehat{\mathrm{m}}_{\mathrm{X}}$.*

By the discussion from Chapter 7, we have the following

LEMMA 2.2. *Same assumption. The measure μ is ergodic (actually mixing) with respect to $a^{\mathbb{Z}}$-action for every $a_{\neq \mathrm{id}} \in \mathrm{A}$.*

Let μ be a B-invariant, a-ergodic probability measure. Here a is a fixed element of A such that $a^n v a^{-n} \to \mathrm{id}$ as $n \to +\infty$ for every $v \in \mathrm{V}$. We need to show that μ coincides with the $\mathrm{m_X}$ (up to a scalar) and in particular, $\mathrm{m_X}$ is finite.

Fix some o in the support of μ. Choose (symmetric) neighborhoods of identity $\mathscr{N}^{\mathrm{B}}_{\varepsilon}$ (resp., $\mathscr{N}^{\mathrm{V}}_{\varepsilon}$) in B (resp., V) that are very small compared to the injectivity radius at o. We say two points x, y are on the same local B (resp., V) orbit iff $x \in \mathscr{N}^{\mathrm{B}}_{\varepsilon}.y$ (resp., $x \in \mathscr{N}^{\mathrm{V}}_{\varepsilon}.y$). Choose $\delta > 0$ even smaller compared to ε.

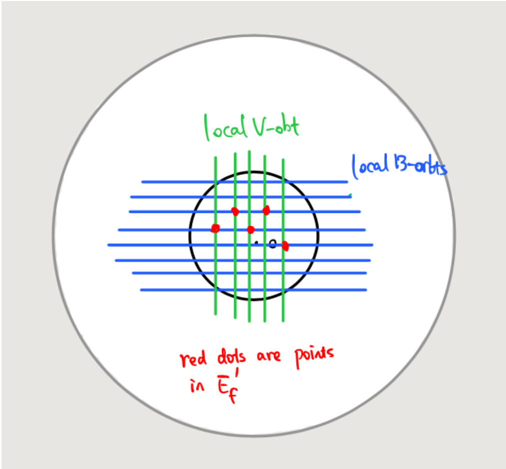
Let

$$\mathrm{Gene}(f,\mu) := \left\{x \in \mathrm{X} \,\middle|\, \lim_{N \to +\infty} \frac{1}{N} \sum_{n=1}^{N} f(a^n x) = \int f(y)\mu(y)\right\}.$$

Note that this set is $\mathrm{V} \cdot \mathrm{A}$-invariant. Let E_f be its intersection with $\mathscr{N}_\delta(o)$.

We define a sub-σ-algebra $\mathscr{A}$ on $\mathscr{N}_\delta(o)$ by specifying its atoms: x and y belong to the same atom iff x and y are on the same local B-orbit. Let $E'_f \subset E_f$ be those x such that the conditional measure $\mu^{\mathscr{A}}_x$ is the restriction of some (left-)B-invariant measure when we identify $[x]^{\mathscr{A}}$ as a subset of $\mathscr{N}^{\mathrm{B}}_{\varepsilon} \subset \mathrm{B}$ via the orbit map. Then μ being B-invariant, E'_f is a conull set in E_f. Let $\widetilde{E}_f$ consist of elements in $\mathscr{N}_\delta(o)$ that are on the local V-orbit of some element in E'_f. Thus $\widetilde{E}_f$ is conull in $\mathscr{N}_\delta(o)$ with respect to μ and $\mathrm{m_X}$.

As an exercise, the reader is invited to fill in the various missing details here. Consult Sec.3 if it helps. Here is a picture.



Now we are ready to conclude the proof.

First assume $\mathrm{m_X} < \infty$. Every point $x \in \widetilde{E}_f$ is a-generic for μ. But since the $a^{\mathbb{Z}}$-action on $\mathrm{m_X}$ is also ergodic and $\mathrm{m_X}(\widetilde{E}_f) > 0$, we can find a point $x \in \widetilde{E}_f$ generic for $\mathrm{m_X}$. Thus $\int f(x)\mu(x) = \int f(x)\mathrm{m_X}(x)$ by pointwise ergodic theorem. Since f is arbitrary we are done.

Now assume $\mathrm{m_X} = \infty$. Then the associated unitary representation is absence of constants. Thus by mixing, for every $\phi, \psi \in L^2(\mathrm{X}, \mathrm{m_X})$, we have

$$\lim_{n\to\infty} \int \phi(a^n.x)\psi(x)\mathrm{m_X}(x) = 0.$$

Take $\phi = f$ and $\psi = 1_{\widetilde{E}_f}$, then

$$\lim_{n\to\infty} \int_{\widetilde{E}_f} f(a^n.x)\mathrm{m_X}(x) = \lim_{n\to\infty} \int f(a^n.x)1_{\widetilde{E}_f}(x)\mathrm{m_X}(x) = 0.$$

Let us compute, for $f \in C_c(\mathrm{X})$,

$$\begin{aligned} \mathrm{m_X}(\widetilde{E}_f) \int f(x)\mu(x) &= \int_{\widetilde{E}_f} \left(\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(a^n.x) \right) \mathrm{m_X}(x) \\ \text{(bounded convergence thm)} &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \left(\int_{\widetilde{E}_f} f(a^n.x)\mathrm{m_X}(x) \right) = 0, \end{aligned}$$

which is impossible if $f > 0$ at some point in $\mathrm{Supp}(\mu)$. Hence $\mathrm{m_X} = \infty$ leads to a contradiction. See [Ra92, Page 27,28] for an alternative way of concluding the proof.

3. More details on locally invariant measures

Let G be a Lie group and $\Omega \subset G$ be a nonempty open subset. A measure μ on Ω is said to be *locally left invariant* under G iff for every measurable subset $A \subset \Omega$ and $g \in G$ such that $gA \subset \Omega$, we have $\mu(gA) = \mu(A)$.

LEMMA 3.1. *A locally left invariant locally finite measure μ is the restriction of some left G-invariant measure on G.*

PROOF. Fix a countable set $(g_i)_{i\in\mathbb{Z}_{\geq 0}}$ in G such that $G = \bigcup g_i.\Omega$. Assume $g_0 = \mathrm{id}$.

$$A_0 := \Omega,\ A_1 := g_1.\Omega \setminus \Omega,\ A_2 := g_2.\Omega \setminus (\Omega \cup g_1\Omega)\ldots.$$

Then $G = \bigsqcup_{i\in\mathbb{Z}_{\geq 0}} A_i$.

Define a measure μ' on G by

$$\mu'(E) := \sum_{i\geq 0} \mu(g_i^{-1}(E\cap A_i)) \tag{37}$$

for every measurable subset E. Then one can prove that μ' is left G-invariant. Here are more details.

Take $g \in G$. For simplicity let $E_i := E\cap A_i$, then

$$\mu'(g.E) = \sum_j \mu'(g.E_j) = \sum_{i,j} \mu(g_i^{-1}(g.E_j \cap A_i)).$$

Note that $g_i^{-1}(g.E_j\cap A_i) \subset \Omega$ and $(g\circ g_j)^{-1}(g.E_j\cap A_i)\subset\Omega$. By local left-invariance we get

$$\mu(g_i^{-1}(g.E_j\cap A_i)) = \mu(g_j^{-1}g^{-1}(g.E_j\cap A_i)).$$

Note that

$$\bigsqcup_i g_j^{-1}g^{-1}(g.E_j\cap A_i) = g_j^{-1}g^{-1}(g.E_j) = g_j^{-1}.E_j.$$

Thus from Equa.(37) we get

$$\mu'(g.E) = \sum_{i,j}\mu(g_j^{-1}g^{-1}(g.E_j\cap A_i)) = \sum_j \mu(g_j^{-1}.E_j) = \mu'(E).$$

So we are done.

□

To check local-invariance, the following is helpful.

LEMMA 3.2. *Let $\delta > 0$. Let Ω be a connected open subset of G and μ be a measure on Ω. Assume μ is locally invariant with respect to $g \in \mathcal{N}_\delta(\mathrm{id}) \subset G$. Then μ is locally invariant.*

Note that being open connected and being open path-connected is equivalent for a subset of a manifold.

PROOF. Fix a countable dense subset $\{\epsilon_i\}$ of $\mathcal{N}_\delta(\mathrm{id})$. For every $g\in G$, the collection of finite-length words in $\{\epsilon_i\}$ representing g is a countable set, we may index it by $\mathbb{Z}_{\geq 1}$. More precisely,

$$\mathscr{W}(g) := \left\{ w_k : \{1,2,\dots,l\} \to \{\epsilon_i\} \;\middle|\; l,k\in\mathbb{Z}_{\geq 1},\ w_k(l)\cdot\ldots\cdot w_k(1) = g \right\}$$

Write l_k for the length of the word w_k. Let $E\subset\Omega$ with $g.E\subset\Omega$. We need to show $\mu(E)=\mu(g.E)$. Now for every $w\in\mathscr{W}(g)$, consider

$$E_1 := \{x\in E \mid w_1(k)\cdot\ldots\cdot w_1(1).x\in\Omega,\ \forall k=1,\dots,l_1\}$$
$$E_2 := \{x\in E\setminus E_1 \mid w_2(k)\cdot\ldots\cdot w_2(1).x\in\Omega,\ \forall k=1,\dots,l_2\}$$
$$\dots$$

Then E_i's are disjoint from each other. Moreover, since Ω is path-connected and $\{\epsilon_i\}$ is dense in $\mathcal{N}_\delta(\mathrm{id})$, we have

$$E = \bigsqcup_{i\in\mathbb{Z}_{\geq 1}} E_i.$$

Moreover for each i,

$$\mu(E_i) = \mu(w_i(1).E_i) = \mu(w_i(2)w_i(1).E_i) = \ldots = \mu(g.E_i).$$

Therefore,

$$\mu(E) = \sum\mu(E_i) = \sum\mu(g.E_i) = \mu(g.E).$$

□

4. Exercises

Equidistribution of unipotent flows on quotients of $\mathbf{SL}_2(\mathbb{R})$

Back to the Top.
Notations

- $\mathrm{X}_2 := \{\text{unimodular lattices in } \mathbb{R}^2\} \cong \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$, $\mathrm{G} := \mathrm{SL}_2(\mathbb{R})$;
- $\mathrm{U} := \left\{ \mathbf{u}_s := \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \,\middle|\, s \in \mathbb{R} \right\}$; $\mathrm{A} := \left\{ \mathbf{a}_t = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix} \,\middle|\, t \in \mathbb{R} \right\}$;
- $\widehat{\mathrm{m}}_{\mathrm{X}_2}$ is the $\mathrm{SL}_2(\mathbb{R})$-invariant probability measure on X_2;
- $\mathrm{Prim}(\Lambda)$ is the set of non-zero primitive vectors in Λ for $\Lambda \leq \mathbb{R}^2$ discrete;
- $\mathrm{Prim}^1(\Lambda)$ is the set of rank-1 primitive subgroups of Λ.

1. Equidistribution on the modular surface

In this section, we illustrate the idea of [DS84] in the case X_2. The general case will be discussed in Sec.3.

THEOREM 1.1. *Let $\Lambda_0 \in \mathrm{X}_2$ be such that $\mathrm{U}.\Lambda_0$ is not compact. Then*

$$\lim_{S \to +\infty} \mu_S := \lim_{S \to +\infty} \int_0^S (\mathbf{u}_s)_* \delta_{\Lambda_0} \,\mathrm{d}s = \widehat{\mathrm{m}}_{\mathrm{X}_2}.$$

Consider

$$\mathscr{T} := \left\{ \Lambda \in \mathrm{X}_2 \,\middle|\, \mathrm{U}.\Lambda \text{ is compact} \right\}.$$

LEMMA 1.2. *The set of compact U-orbits is a tube: $\mathscr{T} = \{\mathbf{a}_t \mathbf{u}_s \cdot \mathbb{Z}^2,\ t \in \mathbb{R},\ s \in \mathbb{R}/\mathbb{Z}\}$. And $\mathrm{U}.\Lambda$ is compact iff Λ contains a non-zero horizontal vector.*

This has been proved in Chapter 4.

Our proof of Thm.1.1 decomposes as:

Step 1. Passing to a subsequence, assume the limit of $(\mu_S)_S$ exists and call it μ. Thanks to the non-divergence theorem (see Ch.4), we also know μ is a probability measure.
Step 2. Also μ is readily seen to be U-invariant since it comes from an averaging process.
Step 3. Show $\mu(\mathscr{T}) = 0$.
Step 4. Use the ergodic decomposition to conclude.

Details of Step 1 and 2 are left as an exercise. Let us take up Step 3.

PROOF OF STEP 3. Fix $t_1 < t_2$, let

$$\mathscr{T}_{[t_1,t_2]} := \left\{ \mathbf{a}_t \mathbf{u}_s.\mathbb{Z}^2,\ t \in [t_1, t_2],\ s \in \mathbb{R}/\mathbb{Z} \right\}.$$

Thus it suffices to show that $\mu(\mathscr{T}_{[t_1,t_2]}) = 0$ for all $-\infty < t_1 < t_2 < +\infty$. By the definition of weak$*$ convergence, it suffices to find an open neighborhood $\mathscr{N}_\varepsilon$, for every $\varepsilon > 0$, of $\mathscr{T}_{[t_1,t_2]}$ such that $\limsup \mu_S(\mathscr{N}_\varepsilon) \leq \varepsilon$. Letting $\varepsilon \to 0$ the finishes the proof.

This will be achieved by Thm.1.5 below. $\square$

Note that $\mathbf{u}_s\Lambda_0$ being close to $\mathscr{T}_{[t_1,t_2]}$ means that, for certain $v \in \mathrm{Prim}(\Lambda_0)$, we have $\mathbf{u}_s.v$ is close to

$$A_{[t_1,t_2]} := \{\mathbf{a}_t\mathbf{u}_s.e_1 \mid t \in [t_1,t_2],\ t \in \mathbb{R}\} = [e^{t_1}, e^{t_2}] \times \{0\}.$$

For $C, \delta > 0$, consider the box

$$\mathrm{Box}_{C,\delta} := [-C,C] \times [-\delta,\delta].$$

Define

$$I(C,\delta) := \left\{s \geq 0 \,\middle|\, \mathrm{Prim}(\mathbf{u}_s.\Lambda_0) \cap \mathrm{Box}_{C,\delta} \neq \emptyset\right\}.$$

For $\mathbb{Z}v \in \mathrm{Prim}^1(\Lambda_0)$, consider

$$I(C,\delta,\mathbb{Z}v) := \left\{s \geq 0 \,\middle|\, \mathbf{u}_s.v \in \mathrm{Box}_{C,\delta}\right\}.$$

Since $-\mathrm{Box}_{C,\delta} = \mathrm{Box}_{C,\delta}$, this is independent of the choice of the generator of $\mathbb{Z}v$. From the definition,

$$I(C,\delta) = \bigcup_{\mathbb{Z}v \in \mathrm{Prim}^1(\Lambda_0)} I(C,\delta,\mathbb{Z}v). \tag{38}$$

The key observation is that

LEMMA 1.3. *Assume $\delta \cdot C \leq 0.1$. Then for two $\mathbb{Z}v \neq \mathbb{Z}w \in \mathrm{Prim}^1(\Lambda_0)$, $I(C,\delta,\mathbb{Z}v) \cap I(C,\delta,\mathbb{Z}w) = \emptyset$. In other words, Equa.(38) above is a disjoint union when $\delta \cdot C \leq 0.1$.*

PROOF. Otherwise the lattice $\mathbf{u}_s.\Lambda_0$ would contain two linearly independent vectors v, w in $[-C,C] \times [-\delta,\delta]$. Thus the triangle spanned by v, w is also contained in $[-C,C] \times [-\delta,\delta]$, implying $\|v \wedge w\| \leq 2(4C\delta) < 1$. This contradicts against the assumption Λ_0 is unimodular. □

For $\varepsilon > 0$, define

$$C_1(\varepsilon) := \varepsilon^{-1}, \quad \delta_1(\varepsilon) := 0.1\varepsilon.$$

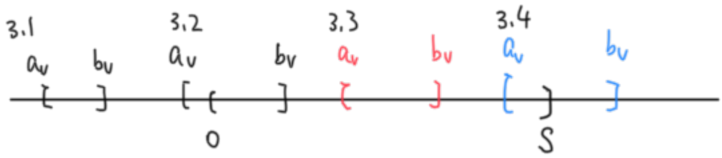
For every $\mathbb{Z}v \in \mathrm{Prim}^1(\Lambda_0)$, there are three cases

Case 1. $I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v) = \emptyset$;
Case 2. $I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v) \neq \emptyset$ and $\mathbb{Z}v \not\subseteq \mathbb{R}e_1$; in this case $I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v) = \mathbb{R}_{\geq 0}$;
Case 3. $I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v) \neq \emptyset$ and $\mathbb{Z}v \not\subseteq \mathbb{R}e_1$; in this case $I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v)$ is a closed interval of the form $[a_v, b_v]$.

Case 2 is excluded since Λ_0 contains no non-zero horizontal vector by assumption (see Lem.1.2).

Now take $S > 0$, there are sub-cases for case 3:

3.1 $S < a_v$ or $b_v < 0$; in this case $[0,S] \cap I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v) = \emptyset$;
3.2 $a_v \leq 0 \leq b_v \leq S$; in this case $[0,S] \cap I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v) = [0,b_v]$;
3.3 $0 < a_v \leq b_v < S$; in this case $[0,S] \cap I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v) = [a_v,b_v]$;
3.4 $0 \leq a_v \leq S \leq b_v$; in this case $[0,S] \cap I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v) = [a_v,S]$;
3.5 $[0,S] \subset [a_v,b_v]$.



PROPOSITION 1.4. *Take C_2 satisfying $1 < C_2 < 0.5C_1(\varepsilon) = 0.5\varepsilon^{-1}$. Then*

$$\limsup_{S\to+\infty} \frac{1}{S} \mathrm{Leb}\left(I(C_2,\delta_1(\varepsilon)) \cap [0,S]\right) \leq 4C_2\varepsilon.$$

From the proof it will be clear that the inequality holds for S large enough without taking the limit.

Only case 3.2, 3.3 and 3.4 above will contribute, for which we have three lemmas Lem.2.2, 2.1, and 2.3 below.

PROOF. If every $\mathbb{Z}v \in \mathrm{Prim}^1(\Lambda)$ falls in case 1 or case 3.1 (for every $S > 0$), then LHS in Prop.1.4 is zero and the inequality trivially holds. Otherwise, find $S > 0$ large enough such that for some $\mathbb{Z}v \in \mathrm{Prim}^1(\Lambda)$, we are in case 3.2–3.5. By choosing S larger, case 3.5 can be excluded.

$$\mathrm{Leb}\,(I(C_2,\delta_1(\varepsilon)) \cap [0,S]) = \left| \bigsqcup_{v\in\mathrm{case}\,3} [0,S] \cap I(C_2,\delta_1(\varepsilon),\mathbb{Z}v) \right|$$
$$(\text{Lem.2.2, 2.1, 2.3}) \le 4C_2\varepsilon \cdot \sum |[0,S] \cap I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v)| \le 4C_2\varepsilon \cdot S$$

□

Now we can prove:

THEOREM 1.5. *For every $\varepsilon > 0$, there is a neighborhood $\mathcal{N}_\varepsilon$ of $\mathcal{T}_{t_1,t_2}$ such that*

$$\limsup_{S\to+\infty} \mu_S(\mathcal{N}_\varepsilon) \le \varepsilon.$$

Consequently for every limit point μ of (μ_S) , $\mu(\mathcal{T}_{t_1,t_2}) = 0$.

PROOF. Take $C_2 > 1$, depending on t_1, t_2, such that $\mathrm{Box}_{C_2,\delta_1(\varepsilon)}$ contains $[e^{t_1}, e^{t_2}] \times \{0\}$. Let $\varepsilon' := \frac{\varepsilon}{4C_2}$. When $\varepsilon > 0$ is small enough, $C_2 < 0.5(\varepsilon')^{-1}$. Define $\mathcal{N}_\varepsilon$ to be those lattices whose primitive vectors intersect non-trivially with $\mathrm{Box}_{C_2,\delta_1(\varepsilon')}$. Then Prop.1.4 concludes the proof.

□

Thus we have completed Step 3.

PROOF OF STEP 4. Say we have a U-invariant probability measure μ with $\mu(\mathcal{T}) = 0$. By classification of ergodic U-invariant probability measures ν on X (see Ch.4 Thm.3.5), either ν is supported on $\mathcal{T}$ or $\nu = \widehat{\mathrm{m}}_{\mathrm{X}_2}$. Let

$$\mu = \int_{\mathrm{Prob}(\mathrm{X}_2)^{\mathrm{U,Erg}}} \nu\, \lambda(\nu)$$

be the ergodic decomposition of μ, then

$$0 = \mu(\mathcal{T}) = \int \nu(\mathcal{T})\, \lambda(\nu).$$

Thus λ-almost every ν, $\nu(\mathcal{T}) = 0 \implies \nu = \widehat{\mathrm{m}}_{\mathrm{X}_2}$. So $\mu = \widehat{\mathrm{m}}_{\mathrm{X}_2}$. □

2. Supplementary lemmas

LEMMA 2.1. *(For case 3.3 above) Assume $\Lambda_0 \cap \mathbb{R}e_1 = \{0\}$, then for $C > 0$,*

$$|I(C,\delta_1(\varepsilon),\mathbb{Z}v)| \le C\varepsilon \cdot |I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v)|.$$

PROOF. If the LHS is 0, then nothing needs to be done. Otherwise, assume w.l.o.g that $v = (v_1, v_2)$ with $v_2 > 0$. Then

$$\begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 + sv_2 \\ v_2 \end{pmatrix},$$

and

$$I(C,\delta_1(\varepsilon),\mathbb{Z}v) = \frac{1}{v_2}[-v_1 - C, -v_1 + C]. \tag{39}$$

Thus

$$|I(C,\delta_1(\varepsilon),\mathbb{Z}v)| = \frac{2C}{v_2} = C\varepsilon \cdot \frac{2\varepsilon^{-1}}{v_2} = C\varepsilon \cdot |I(C_1(\varepsilon),\delta_1(\varepsilon),\mathbb{Z}v)|.$$

□

LEMMA 2.2. *(For Case 3.2 above) Assume* $\Lambda_0 \cap \mathbb{R}e_1 = \{0\}$. *Take* $\mathbb{Z}v \in \mathrm{Prim}^1(\Lambda)$ *and* $S > 0$ *satisfying case 3.2 above. Also let* $1 < C \leq 0.5\varepsilon^{-1}$. *Then*

$$|[0,S] \cap I(C, \delta_1(\varepsilon), \mathbb{Z}v)| \leq 4C\varepsilon \cdot |[0,S] \cap I(C_1(\varepsilon), \delta_1(\varepsilon), \mathbb{Z}v)|.$$

PROOF. W.l.o.g, keep assuming that $v = (v_1, v_2)$ with $v_2 > 0$. In case 3.2,

$$[0,S] \cap I(C_1(\varepsilon), \delta_1(\varepsilon), \mathbb{Z}v) = [0, b_v].$$

If $[0,S] \cap I(C, \delta_1(\varepsilon), \mathbb{Z}v)$ is empty nothing needs to be done. Otherwise, by Equa.(39),

$$0 < -v_1 + C \implies v_1 < C.$$

Then

$$\begin{aligned} |[0,S] \cap I(C, \delta_1(\varepsilon), \mathbb{Z}v)| \leq \frac{2C}{v_2} &= \frac{2C}{-v_1 + \varepsilon^{-1}} \cdot \frac{-v_1 + \varepsilon^{-1}}{v_2} \\ &= \frac{2C}{-v_1 + \varepsilon^{-1}} \cdot |[0,S] \cap I(C_1(\varepsilon), \delta_1(\varepsilon), \mathbb{Z}v)| \end{aligned}$$

It remains to observe

$$\frac{2C}{-v_1 + \varepsilon^{-1}} \leq \frac{2C}{-C_2 + \varepsilon^{-1}} \leq \frac{2C}{-0.5\varepsilon^{-1} + \varepsilon^{-1}} = 4C\varepsilon.$$

□

LEMMA 2.3. *[For Case 3.4 above] Assume* $\Lambda_0 \cap \mathbb{R}e_1 = \{0\}$. *Take* $\mathbb{Z}v \in \mathrm{Prim}^1(\Lambda)$ *and* $S > 0$ *satisfying case 3.4 above. Also let* $1 < C \leq 0.5\varepsilon^{-1}$. *Then*

$$|[0,S] \cap I(C, \delta_1(\varepsilon), \mathbb{Z}v)| \leq 4C\varepsilon \cdot |[0,S] \cap I(C_1(\varepsilon), \delta_1(\varepsilon), \mathbb{Z}v)|.$$

PROOF. W.l.o.g, keep assuming that $v = (v_1, v_2)$ with $v_2 > 0$. In case 3.4

$$[0,S] \cap I(C_1(\varepsilon), \delta_1(\varepsilon), \mathbb{Z}v) = [a_v, S].$$

If $[0,S] \cap I(C, \delta_1(\varepsilon), \mathbb{Z}v)$ is empty, nothing needs to be done. Otherwise, by Equa.(39),

$$\frac{-v_1 - C}{v_2} \leq S \implies v_1 + v_2 S \geq C.$$

Under this condition we have

$$|[0,S] \cap I(C, \delta_1(\varepsilon), \mathbb{Z}v)| \leq \frac{2C}{v_2}$$

and

$$|[0,S] \cap I(C_1(\varepsilon), \delta_1(\varepsilon), \mathbb{Z}v)| = S - \frac{(-v_1 - \varepsilon^{-1})}{v_2} = \frac{\varepsilon^{-1} + v_2 S + v_1}{v_2} \geq \frac{\varepsilon^{-1} - C}{v_2} \geq \frac{0.5\varepsilon^{-1}}{v_2}.$$

Thus,

$$|[0,S] \cap I(C, \delta_1(\varepsilon), \mathbb{Z}v)| \leq \frac{2C}{0.5\varepsilon^{-1}} |[0,S] \cap I(C_1(\varepsilon), \delta_1(\varepsilon), \mathbb{Z}v)|.$$

Note that $\frac{2C}{0.5\varepsilon^{-1}} = 4C\varepsilon$. □

3. [Not readable at the moment]Other non-cocompact lattices

[Needs further revision!]

Let $\Gamma \leq \mathrm{G} := \mathrm{SL}_2(\mathbb{R})$ be a lattice. Let $\mathrm{X} := \mathrm{G}/\Gamma$. We are going to assume some light hyperbolic geometry. Readers who are less familiar with hyperbolic geometry are welcome to take $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. Main ideas are preserved in this case.

The discussion here is more “geometric" compared to the last section.

First we have the non-divergence theorem (compare Chapter 4).

THEOREM 3.1. *For every $\varepsilon > 0$, there exists a compact subset of $\mathscr{C} \subset X$ such that for every $x \in X$, either*

$$\limsup_{S\to+\infty} \frac{1}{S}\mathrm{Leb}\{s \in [0,S],\ \mathbf{u}_s.x \notin \mathscr{C}\} \leq \varepsilon$$

or $\mathrm{U}.x$ is compact.

Proof is left as an exercise.

Let

$$\mathscr{T} := \left\{x \in \mathrm{X} \,\middle|\, \mathrm{U}.x \text{ is compact}\right\}.$$

One can show that

THEOREM 3.2. *There exist finitely many points $y_1,\dots,y_l$ in X with compact U-orbits such that if $\mathscr{T}_i := \mathrm{AU}.y_i$ then*

$$\mathscr{T} = \bigsqcup_{i=1,\dots,l} \mathscr{T}_i.$$

Fix $x_0 \notin \mathscr{T}$, let

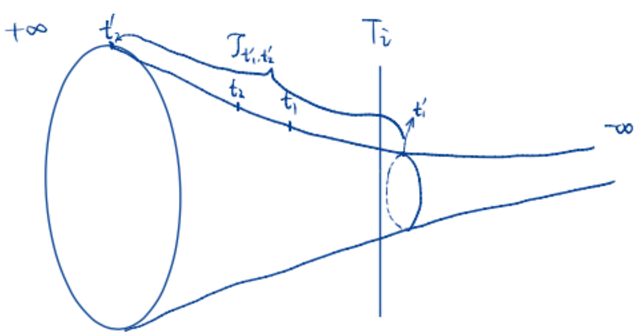
$$\mu_S := \frac{1}{S}\int_0^S (\mathbf{u}_s)_*\delta_{x_0}\,\mathrm{ds}$$

and take μ to be a weak$*$ limit. Let us explain why $\mu(\mathscr{T}) = 0$, which follows if $\mu(\mathscr{T}_i) = 0$ for every $i = 1,\dots,l$. From now on we focus on a single index i_0. W.l.o.g., assume $y_{i_0} = [\mathrm{id}]_\Gamma$, where $[\bullet]_\Gamma$ stands for the image of $\bullet$ in the quotient by Γ.

3.1. Lifts of tubes. Define, for $-\infty \leq t_1 < t_2 \leq +\infty$,

$$\mathscr{T}_{t_1,t_2,i_0} := \left\{\mathbf{a}_t\mathrm{U}.y_{i_0} \,\middle|\, t_1 < t < t_2,\ s \in \mathbb{R}\right\}$$
$$\widetilde{\mathscr{T}}_{t_1,t_2,i_0} := \left\{\mathbf{a}_t\mathrm{U}.\widetilde{y}_{i_0} \,\middle|\, t_1 < t < t_2,\ s \in \mathbb{R}\right\},\ \widetilde{\mathscr{T}}_{i_0} := \mathrm{AU}.\widetilde{y}_{i_0}.$$

where $\widetilde{y}_{i_0} = [\mathrm{id}]_{\Gamma\cap\pm\mathrm{U}} \in \mathrm{G}/\Gamma\cap\pm\mathrm{U}$. In general, one should lift $y_i =: [g_i]_\Gamma$ to $\widetilde{y}_i = [g_i]_{\Gamma\cap\pm g_i\mathrm{U}g_i^{-1}}$.



THEOREM 3.3. *Fix some* $-\infty < t_1 < t_2 < +\infty$. *For every* $\varepsilon > 0$, *there exists a neighborhood* $\mathcal{N}_\varepsilon$ *of* $\mathcal{T}_{t_1,t_2}$ *such that*

$$\limsup_{S\to+\infty} \frac{1}{S} \operatorname{Leb}\{s \in [0,S] \mid \mathbf{u}_s.x_0 \in \mathcal{N}_\varepsilon\} \leq \varepsilon.$$

In light of the case of X_2, we are going to find two neighborhoods $\mathcal{N}_\varepsilon \subset \mathcal{N}'_\varepsilon$ such that the time a noncompact U-orbit spends in $\mathcal{N}_\varepsilon$ is much shorter than that in $\mathcal{N}'_\varepsilon$.

Consider the natural projection $p : \mathrm{G}/\pm\mathrm{U}\cap\Gamma \to \mathrm{G}/\Gamma$. It is an injection restricted to $\widetilde{\mathcal{T}_i}$ and is a closed embedding when restricted to the closure of $\widetilde{\mathcal{T}}_{s,t}$ for every pair $s<t$. For $t'_1 < t_1$ and $t'_2 = t'_2(\varepsilon)$ to be determined, there exists an open neighborhood $\widetilde{\Omega}_\varepsilon$ of $\widetilde{\mathcal{T}}_{t'_1,t'_2}$ such that

$$p|_{\widetilde{\Omega}_\varepsilon} : \widetilde{\Omega}_\varepsilon \to p\left(\widetilde{\Omega}_\varepsilon\right) \tag{40}$$

is a homeomorphism.

DEFINITION 3.4. *For* $T \in \mathbb{R}$,

$$\begin{aligned} \operatorname{cusp}_{i_0}(T) &:= \left\{k\mathbf{a}_t\mathrm{U}.y_{i_0} \mid k \in \mathrm{SO}_2(\mathbb{R}),\ t<T\right\} = \mathrm{SO}_2(\mathbb{R})\mathcal{T}_{-\infty,T,i_0};\\ \widetilde{\operatorname{cusp}}_{i_0}(T) &:= \left\{k\mathbf{a}_t\mathrm{U}.\widetilde{y}_{i_0} \mid k \in \mathrm{SO}_2(\mathbb{R}),\ t<T\right\} = \mathrm{SO}_2(\mathbb{R})\mathcal{T}_{-\infty,T,i_0}, \end{aligned}$$

LEMMA 3.5. *There exists* $T_0 \in \mathbb{R}$ *such that the following holds. Under* p, $\widetilde{\operatorname{cusp}}_{i_0}(T_0)$ *is mapped homeomorphically onto* $\operatorname{cusp}_{i_0}(T_0)$. *For* $s \in \mathbb{R}$, *there exists* $T(s) < T_0$ *such that* $\mathcal{T}_{T_0,s,i_0}$ *does not intersect* $\operatorname{cusp}_{i_0}(T(s))$.

Anticipating the proof, t'_2 will be chosen depend on ε and linear algebra. Then we choose t'_1 to be T'_1+1 where $T'_1 := T(t'_2)$ comes from Lem.3.5. Then $\widetilde{\Omega}_\varepsilon$ is chosen by Equa.(40).

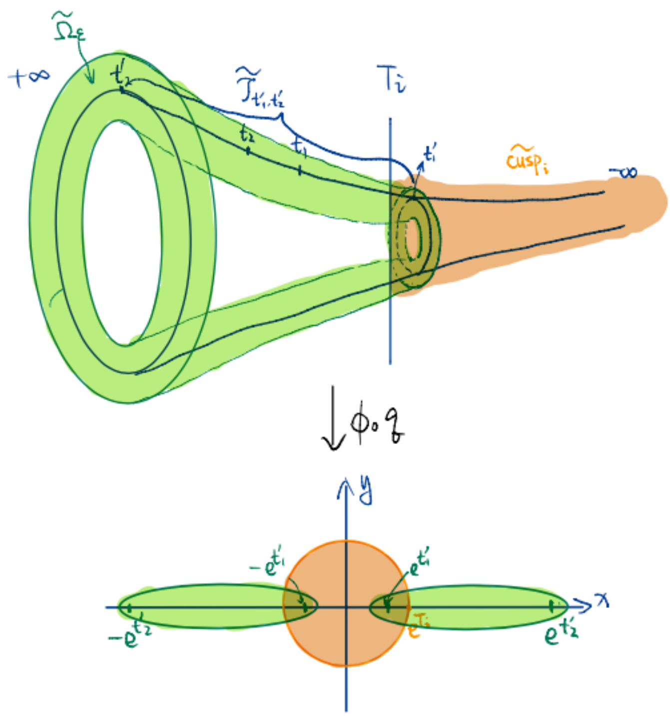
LEMMA 3.6. *The natural projection* p *restricted to* $\operatorname{cusp}_{i_0}(T'_1) \cup \widetilde{\Omega}_\varepsilon$ *is a homeomorphism onto its image.*

3.2. Linearization. Let us define q to be the natural quotient $\mathrm{G}/\Gamma\cap\pm\mathrm{U} \to \mathrm{G}/\pm\mathrm{U}$ and $\phi : \mathrm{G}/\pm\mathrm{U} \to \mathbb{R}^2/\pm1$ by $\phi(g) := g.e_1/\pm1$. For notational convenience, we will be working with $\mathbb{R}^2$ rather than $\mathbb{R}^2/\pm1$. Here is a diagram.

$$\begin{tikzcd} & \mathrm{G}/\Gamma\cap\pm\mathrm{U} \arrow[dl, "p"'] \arrow[dr, "q"] & & \\ \mathrm{G}/\Gamma & & \mathrm{G}/\pm\mathrm{U} \arrow[r, hook, "\phi"] & \mathbb{R}^2/\pm1 \end{tikzcd}$$

The $\widetilde{\operatorname{cusp}}_{i_0}$ is already q-saturated: $q^{-1}q(\widetilde{\operatorname{cusp}}_{i_0}) = \widetilde{\operatorname{cusp}}_{i_0}$. More concretely,

$$\phi\circ q(\widetilde{\operatorname{cusp}}_{i_0}) = \left\{v_{\neq0} \in \mathbb{R}^2 \mid \|v\| < e^{T_i}\right\}/\pm1.$$



$\widetilde{\Omega}_\varepsilon$ may not be q-saturated. However, its image under $\phi\circ q$ is an open neighborhood of

$$\phi\circ q\left(\widetilde{\mathcal{T}}_{t_1',t_2',i_0}\right)=(e^{t_1'},e^{t_2'})\times\{0\}/\pm 1.$$

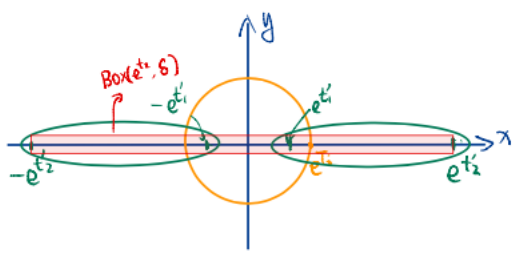
Then one can show that there exists a smaller open nbhd Ω' of $q\left(\widetilde{\mathcal{T}}_{t_1',t_2',i_0}\right)$ such that its preimage under q is contained in $\widetilde{\Omega}_\varepsilon$. Thus we can choose $\delta=\delta(\varepsilon)>0$ small enough such that

$$\widetilde{\Omega}_\varepsilon':=(\phi\circ q)^{-1}\left((e^{t_1'},e^{t_2'})\times(-\delta,\delta)\right)/\pm 1.$$

is contained in $\widetilde{\Omega}_\varepsilon$. To combine $\widetilde{\text{cusp}}_{i_0}$ with $\widetilde{\Omega}_\varepsilon'$, choose a even smaller δ such that

$$\widetilde{\mathcal{N}}_\varepsilon':=(\phi\circ q)^{-1}\left(\text{Box}(e^{t_2'},\delta)\right)/\pm 1.$$

is contained in $\widetilde{\text{cusp}}_{i_0}\cup\widetilde{\Omega}_\varepsilon'$ where $\text{Box}(e^{t_2'},\delta)=[-e^{t_2'},e^{t_2'}]\times[-\delta,\delta]$. So p restricted to $\widetilde{\mathcal{N}}_\varepsilon'$ is injective.



Also let

$$\widetilde{\mathcal{N}}_\varepsilon:=(\phi\circ q)^{-1}\left(\text{Box}(e^{t_2+1},\delta)\right)/\pm 1.$$

Let $\mathcal{N}_\varepsilon':=p(\widetilde{\mathcal{N}}_\varepsilon')$ and $\mathcal{N}_\varepsilon:=p(\widetilde{\mathcal{N}}_\varepsilon)$. They are open neighborhoods of $\mathcal{T}_{t_1,t_2,i_0}$.

3.3. Some linear algebra. At this point, one can adapt the strategy of previous sections to prove Thm.3.3 and hence analogues of Thm.1.1 for other lattices. The constant $t_2' = t_2'(\varepsilon)$ is determined in this process.

4. Exercises

4.1. Equidistribution via mixing.

- $G = \mathrm{SL}_2(\mathbb{R})$, $U = \left\{ \mathbf{u}_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}, s \in \mathbb{R} \right\}$, $A = \left\{ \mathbf{a}_t = \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}, t \in \mathbb{R} \right\}$;
- $V = \left\{ \mathbf{v}_r = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}, r \in \mathbb{R} \right\}$;
- Γ is a lattice in G, let $X := G/\Gamma$ and $\widehat{\mathrm{m}}_X$ be the unique G-invariant probability measure on X;
- Fix a right invariant Riemannian metric on G. Use this metric to induce a distance function $d(\cdot,\cdot)$ on G, let $d_X([g_1]_\Gamma, [g_2]_\Gamma) := \inf_{\gamma_1,\gamma_2 \in \Gamma} d(g_1\gamma_1, g_2\gamma_2)$;
- for every $\delta, s_0 > 0$, let
$$\mathrm{Box}(\delta, s_0) := (-\delta, \delta) \times (-\delta, \delta) \times (0, s_0);$$
- let $\mathrm{Leb}_{\delta,s_0}$ be the restriction of standard Lebesgue measure restricted to $\mathrm{Box}(\delta, s_0)$;
- by abuse of notation we also denote by $\mathrm{Leb}_{\delta,s_0}$ for its push-forward under the map $(r, t, s) \mapsto \mathbf{v}_r \cdot \mathbf{a}_t \cdot \mathbf{u}_s$;
- for $x \in X$, let Obt_x denote the map $G \to X$ defined by $g \mapsto g.x$.

EXERCISE 4.1. *Fix $x \in X$, $\delta, s_0 > 0$. Show that there exists a non-negative function $f \in L^\infty(X, \mathrm{m}_X)$ such that* $(\mathrm{Obt}_x)_* \mathrm{Leb}_{\delta,s_0} = f \cdot \widehat{\mathrm{m}}_X$.

EXERCISE 4.2. *Show that for every $\varepsilon > 0$, there exists $\delta > 0$ such that for every $s_0 > 0$, $t > 0$, $(r, u, s) \in \mathrm{Box}(\delta, s_0)$ and $x \in X$ we have*
$$d_X(\mathbf{a}_t \cdot (\mathbf{v}_r \mathbf{a}_u) \cdot \mathbf{u}_s.x, \mathbf{a}_t \mathbf{u}_s.x) < \varepsilon.$$

Recall that mixing implies that for $\phi, \psi \in L^2(X, \widehat{\mathrm{m}}_X)$,
$$\lim_{t \to \pm\infty} \int \phi(\mathbf{a}_t.x)\psi(x)\widehat{\mathrm{m}}_X(x) = \int \phi(x)\widehat{\mathrm{m}}_X(x) \cdot \int \psi(x)\widehat{\mathrm{m}}_X(x).$$

EXERCISE 4.3. *For every $s_0 > 0$, $x_0 \in X$ and $f \in C_c(X)$, we have*
$$\lim_{t \to +\infty} \frac{1}{s_0} \int_0^{s_0} f(\mathbf{a}_t \mathbf{u}_s.x_0)\,\mathrm{ds} = \int f(x)\widehat{\mathrm{m}}_X(x).$$

EXERCISE 4.4. *Show that if $(U.x_n)$ is a sequence of compact U-orbits of periods $S_n \to +\infty$, then for every compactly supported continuous function f,*
$$\lim_{n \to +\infty} \frac{1}{S_n} \int_0^{S_n} f(\mathbf{u}_s.x_n)\,\mathrm{ds} = \int f(x)\widehat{\mathrm{m}}_X(x).$$

EXERCISE 4.5. *Show that the above convergence (in Exer.4.3) is "uniform" in the following sense. For every $f \in C_c(X)$, $\varepsilon, s_0 > 0$ and $x_0 \in X$, there exists $\delta > 0$ such that for every $y \in X$ with $d_X(x_0, y) < \delta$, we have for all $t > 0$,*
$$\left| \frac{1}{s_0} \int_0^{s_0} f(\mathbf{a}_t \mathbf{u}_s.x_0)\,\mathrm{ds} - \frac{1}{s_0} \int_0^{s_0} f(\mathbf{a}_t \mathbf{u}_s.y)\,\mathrm{ds} \right| < \varepsilon.$$

EXERCISE 4.6. *Use the above exercise to give another proof of the equidistribution of horocycle flows. Show that if $U.x_0$ is not compact in X, then for every $f \in C_c(X)$,*
$$\lim_{S \to +\infty} \frac{1}{S} \int_0^S f(\mathbf{u}_s.x_0)\,\mathrm{ds} = \int f(x)\widehat{\mathrm{m}}_X(x).$$

CHAPTER 11

Ergodic decomposition of unipotent invariant measures

Back to the Top.
The main reference of this chapter is [**Sha91b**] and [**MS95**, Section 2].
Notations

- $\mathrm{X} := \mathrm{G}/\Gamma$ with $\mathrm{G} := \mathrm{SL}_n(\mathbb{R})$ and $\Gamma := \mathrm{SL}_n(\mathbb{Z})$;
- $\mathrm{U} = \{\mathbf{u}_s,\ s \in \mathbb{R}\}$ is a one-parameter unipotent subgroup of G.

DEFINITION 0.1. *A subgroup U of $\mathrm{SL}_n(\mathbb{R})$ is said to be a one-parameter unipotent subgroup iff there exists a nilpotent matrix $u \in \mathfrak{sl}_n(\mathbb{R})$ such that $U = \{\exp(t.u),\ t \in \mathbb{R}\}$.*

In this and the following few chapters, we are going to assume the measure classification theorem of unipotent flows (Thm.1.1 below) and demonstrate how it is applied. Further discussion of its proof is delayed to a later chapter.

1. Ergodic U-invariant measures

The following is the description of ergodic U-invariant probability measures due to Ratner [**Ra91a**].

THEOREM 1.1. *Let μ be an ergodic U-invariant probability measure on X, then there exists $x \in \mathrm{X}$ and a closed connected subgroup $H \leq \mathrm{G}$ containing U such that*

1. *$H.x$ is closed and supports an H-invariant probability measure $\widehat{\mathrm{m}}_{H.x}$;*
2. *$\mu = \widehat{\mathrm{m}}_{H.x}$.*

In short, one says that ergodic U-invariant probability measures are *homogeneous* (the word "algebraic" is also used). By writing $x = [g]_\Gamma$ and replacing H by $g^{-1}Hg$, the theorem may be rephrased as

THEOREM 1.2. *Let μ be an ergodic U-invariant probability measure on X, then there exists $g \in \mathrm{G}$ and a closed connected subgroup $H \leq \mathrm{G}$ containing $g^{-1}\mathrm{U}g$ such that*

1. *$[H]_\Gamma := H\Gamma/\Gamma$ is closed and supports an H-invariant probability measure $\widehat{\mathrm{m}}_{[H]_\Gamma}$;*
2. *$\mu = g_*\widehat{\mathrm{m}}_{[H]_\Gamma}$.*

In particular, $\mathrm{supp}(\mu) = g[H]_\Gamma$.

EXAMPLE 1.3. *If $\mathrm{G} = \mathrm{SL}_2(\mathbb{R})$ and $\mathrm{U} := \left\{\mathbf{u}_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle|\ s \in \mathbb{R}\right\}$, then candidates of H are $\{\mathrm{G}$, conjugates of $\mathrm{U}\}$. If further assume Γ is cocompact, then G is the only candidate. For non cocompact lattices, there are finitely many candidates up to Γ-conjugacy.*

We have discussed this example in depth in Ch.8,9,10.

EXAMPLE 1.4. *If $\mathrm{G} = \mathrm{SL}_2(\mathbb{C})$ $(\Gamma = \mathrm{SL}_2(\mathbb{Z}[i]))$ and $\mathrm{U} := \left\{\mathbf{u}_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle|\ s \in \mathbb{R}\right\}$, then, up to conjugacy, candidates of H are $\{\mathrm{U}, \mathrm{V}, \mathrm{SL}_2(\mathbb{R}), \mathrm{G}\}$, where $\mathrm{V} = \left\{\begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle|\ s \in \mathbb{C}\right\}$.*

Where does the pair (g, H) live?

DEFINITION 1.5. *For two subgroups A, B of G, define*

$$N(A, B) := \left\{g \in \mathrm{G} \,\middle|\, gAg^{-1} \supset B\right\}.$$

In this terminology, g as above belongs to $N(H, \mathrm{U})$. When $H = U$, $N(U, U)$ is just the normalizer of U in G. When $H = \mathrm{G}$, $N(\mathrm{G}, U) = \mathrm{G}$. How about H? Note that we do not want U to appear in the definition of this space.

DEFINITION 1.6. *Let $\mathscr{H}$ be the collection of subgroups L of G satisfying*

1. *L is a connected and closed subgroup;*
2. *$[L]_\Gamma$ is closed and supports an L-invariant probability measure $\widehat{\mathrm{m}}_{[L]_\Gamma}$;*
3. *some one-parameter unipotent subgroup of L acts ergodically on $\widehat{\mathrm{m}}_{[L]_\Gamma}$.*

Thus H as above belongs to $\mathscr{H}$.

From $\mu \in \mathrm{Prob}(\mathrm{X})^{\mathrm{U,Erg}}$ we get a pair (g, H). However, the pair (g, H) is not unique, for

$$g_* \mathrm{m}_{[H]_\Gamma} = (gh)_* \mathrm{m}_{[H]_\Gamma} = (gh\gamma)_* \mathrm{m}_{[\gamma^{-1} H\gamma]_\Gamma}$$

for every $h \in H$ and $\gamma \in \Gamma$. Thus $(gh\gamma, \gamma^{-1} H\gamma)$ and (g, H) correspond to the same μ. The lemma below says that is all.

LEMMA 1.7. *Let H_1, H_2 be two connected closed subgroups of G such that $H_i\Gamma$ (i=1,2) are both closed. Let $g_1, g_2 \in \mathrm{G}$. Then $g_1 H_1 \Gamma = g_2 H_2 \Gamma$ iff there exist $h_2 \in H_2$ and $\gamma_2 \in \Gamma$ such that*

$$g_2 h_2 \gamma_2 = g_1,\ \gamma_2^{-1} H_2 \gamma_2 = H_1.$$

PROOF. It only suffices to prove the " $\Longrightarrow$ " direction. The other direction follows directly. So assume $g_1 H_1 \Gamma = g_2 H_2 \Gamma$. Then

$$H_1 \Gamma = g_3 H_2 \Gamma,\ g_3 := g_1^{-1} g_2.$$

Thus $\mathrm{id} \in g_3 H_2 \Gamma$ and

$$1 = g_3 h_2 \gamma_2,\ \exists h_2 \in H_2, \gamma_2 \in \Gamma_2.$$

This already implies that

$$g_2 h_2 \gamma_2 = g_2 g_3^{-1} = g_1.$$

Now we have

$$H_1 \Gamma = \gamma_2^{-1} h_2^{-1} H_2 \Gamma = \gamma_2^{-1} H_2 \gamma_2 \Gamma.$$

By inspecting a small neighborhood of $[\mathrm{id}]_\Gamma$ and use the fact that $[H_1]_\Gamma$ and $[\gamma_2^{-1} H_2 \gamma_2]_\Gamma$ are both embedded submanifolds, we see that $H_1 = \gamma_2^{-1} H_2 \gamma_2$. □

2. Candidates of homogeneous orbit closure

Take some $x_0 \in \mathrm{X}$. Eventually, we would know that the closure of $\mathrm{U}.x_0$ is homogeneous and the homogeneous measure is finite and U-ergodic. But this does not follow immediately from the Thm.1.1. Nevertheless, we can say something even without appealing to Thm.1.1.

DEFINITION 2.1. *Let $\mathscr{A}$ (depending on $x_0 \in \mathrm{X}$ and U) be the collection of subgroups L of G satisfying*

1. *L is a connected closed subgroup of G containing U;*
2. *$L.x_0$ is closed.*

LEMMA 2.2. *The collection $\mathscr{A}$ has a smallest element. Indeed, if $L_1, L_2 \in \mathscr{A}$ then $(L_1 \cap L_2)^\circ \in \mathscr{A}$.*

PROOF. First we remark that for a closed subgroup $L \leq \mathrm{G}$, if $L\Gamma$ is closed then $L^\circ\Gamma$ is also closed. Indeed, one shows that every orbit of L° on $L/L\cap\Gamma$, which is homeomorphic to $L\Gamma/\Gamma$, is open and hence closed. So it suffices to show that if $L_3 := L_1 \cap L_2$, then $L_3\Gamma$ is closed. This follows from a similar reasoning. Indeed, every orbit of L_3 on $[L_1]_\Gamma \cap [L_2]_\Gamma$ is open and hence closed. To see why every orbit is open, one may take a local neighborhood. □

As an exercise, fill in the missing details in the proof.
Take $g_0 \in \mathrm{G}$ such that $x_0 = [g_0]_\Gamma$.

THEOREM 2.3. *Let $H := H_{\mathscr{A}}$ be the smallest element of $\mathscr{A}$, then*

1. *the closed set $H.x_0$ supports a finite H-invariant measure $\mathrm{m}_{H.x_0}$;*
2. *the measure $\mathrm{m}_{H.x_0}$ is U-ergodic;*
3. *there exists a $\mathbb{Q}$-algebraic subgroup $\mathbf{H}'$ of SL_n such that $g_0^{-1}Hg_0 = \mathbf{H}'(\mathbb{R})^\circ$. Actually, $\mathbf{H}'$ is the smallest $\mathbb{Q}$-algebraic subgroup containing $g_0^{-1}\mathrm{U}g_0$. In particular, H is algebraic.*

The last statement can be skipped if you are allergic to algebraic groups.

Before the proof, note that there is a locally finite measure $\mathrm{m}_{H.x_0}$ that is only "quasi-invariant under H" (for instance, the one induced from a right invariant Riemannian metric). A priori, it is not clear why it is H-invariant. But one can still talk about ergodicity and the associated unitary representation (with suitably twisted action). You may ignore this minor issue by pretending $\mathrm{m}_{H.x_0}$ to be H-invariant from the start.

Here is a sketch of proof.

Step 1. By Mautner's phenomenon (see [**Moo80**, Theorem 1.1] and some supplementary arguments in [**Sha91b**, Proposition 2.7]), there exists a closed normal subgroup $F \lhd H$ containing U such that for every unitary representation of H, every U-fixed vector is F-fixed. We already "know" this if H is semisimple by arguments in Ch.7. See exercises attached to Ch.7 for an example beyong the semisimple case. Thus to show U-ergodicity, suffices to show F-ergodicity.

Step 2. Let Γ_H be the stabilizer of x_0 in H. Explicitly, $\Gamma_H = H \cap g_0\Gamma g_0^{-1}$. Define

$$F' := \overline{F\cdot\Gamma_H}.$$

Since F is normal, F' is a closed subgroup of H. Since F' is right invariant under Γ_H, $F'\Gamma_H/\Gamma_H$ is closed in H/Γ_H. Thus $F'.x_0$ is closed. And F' contains U. By minimality of H, $F' = H$.

Step 3. Now we show F-ergodicity of $\mathrm{m}_{H.x_0}$. Let Ω be an F-invariant measurable set of H/Γ_H. Assume $\mathrm{m}_{H.x_0}(\Omega) > 0$, we need to show that its complement has zero measure. Since F is normal, we see that the preimage $\widetilde{\Omega}$ of Ω in H right invariant under the group $F\cdot\Gamma_H$. Let m_H be a right H-invariant locally finite measure on H. Then $\mu := 1_{\widetilde{\Omega}}\cdot\mathrm{m}_H$ is right $F\cdot\Gamma_H$-invariant. Since μ is a locally finite measure, by continuity, the stabilizer of μ in H (w.r.t. the action from the right) is a closed subgroup. Thus μ is right F'-invariant, hence H-invariant. By the uniqueness of invariant measures, $\mu = \mathrm{m}_H$ (up to a scalar, which has to be 1). In particular, the complement of $\widetilde{\Omega}$ has zero measure. This implies that the complement of Ω also has zero measure.

Step 4. It remains to show that $\mathrm{m}_{H.x_0}$ is a finite measure. In fact every U-ergodic locally finite measure ν is finite. Let us see why. By pointwise ergodic theorem (see [**Wal82**, Theorem 1.14, Section 1.6]), for every $f \in L^1(\nu)$, for ν-almost every X,

$$f^*(x) := \lim_{S\to+\infty}\frac{1}{S}\int_0^S f(\mathbf{u}_s.x)\,\mathrm{d}s \text{ exists.}$$

Moreover $f^* \in L^1(\nu)$ and is U-invariant. By ergodicity, f^* is a constant, which has to be 0 if ν is an infinite measure.

On the other hand, by non-divergence of unipotent flow, there exists a compact set C such that if f is the indicator function of C, then $f^* \neq 0$.

Thus ν has to be finite. This finishes the proof of 1 and 2 of Thm.2.3.

Step 5. To save notation, we assume $g_0 = \mathrm{id}$ here.

Let $\mathbf{L}$ be the smallest $\mathbb{Q}$-algebraic subgroup of SL_n containing U. Let $\pi_1 : \mathbf{L} \to \mathbf{T}$ be the maximal quotient (algebraic) torus of $\mathbf{L}$. π_1 is defined over $\mathbb{Q}$. Since U is unipotent and π_1 preserves this property, the image of $\pi_1(\mathrm{U})$ consists of unipotent elements. But torus $\mathbf{T}$ only contains semisimple elements. Thus U is contained in the kernel of π_1, which is in the form of a semisimple (algebraic) group semidirect product with a unipotent (algebraic) group. In particular, $\mathbf{L}$ admits no nontrivial characters (:=algebraic group morphisms to $\mathbb{C}^\times = \mathrm{GL}_1$). By a theorem of Borel–Harish-Chandra (see for instance [Bor19, Corollary 13.2]), $\mathrm{L} \cap \Gamma$ ($\mathrm{L} := \mathbf{L}(\mathbb{R})^\circ$) is a lattice in L and in particular, $\mathrm{L}\Gamma/\Gamma$ is closed. See exercises attached to Chapter 8. By minimality of H, $H \subset \mathrm{L}$. Our goal is to show $H = \mathrm{L}$ (this is what we mean by saying H is "algebraic"). Note that the Zariski closure of H is equal to $\mathbf{L}$.

Step 6. Let $\mathfrak{h}$ be the Lie algebra of H. By Levi's decomposition (reference? probably Bourbaki's book?), there exists a semisimple sub Lie algebra $\mathfrak{m}$ and a solvable ideal $\mathfrak{r}$ of $\mathfrak{h}$ such that $\mathfrak{h} = \mathfrak{m} \ltimes \mathfrak{r}$. By [Bor91, ChII, Corollary 7.9], since $\mathfrak{m} = [\mathfrak{m}, \mathfrak{m}]$, $\mathfrak{m}$ is already algebraic (i.e., the corresponding Lie subgroup M is algebraic). Let $\mathbf{M}$ be the corresponding $\mathbb{R}$-algebraic subgroup. We seek to show that $\mathfrak{r}$ consists of nilpotent matrices and hence is algebraic. By [Bor91, ChII, Corollary 7.7], this shows that H is algebraic. Since H normalizes $\mathfrak{r}$ and "normalizing $\mathfrak{r}$" is an algebraic condition, we have that $\mathbf{L}$ normalizes $\mathfrak{r}$. Thus $\mathfrak{r}$ is an ideal of $\mathfrak{l}$. Let $\pi_2 : \mathfrak{l} \to \mathfrak{l}/\mathfrak{u} \cong \mathfrak{m}$ (here $\mathfrak{u}$ is the Lie algebra of the unipotent radical of $\mathbf{L}$), then $\pi_2(\mathfrak{r})$ is an ideal of $\mathfrak{l}/\mathfrak{u}$. But every non-zero ideal of a semisimple Lie algebra is semisimple and can not be solvable. Thus $\pi_2(\mathfrak{r}) = 0$, or $\mathfrak{r} \subset \mathfrak{u}$, which consists of nilpotent matrices. The rest of the claim in 3. of Thm.2.3 now follows from Borel density theorem.

3. Tubes and ergodic components

Assume U acts on $g_*\mathrm{m}_{[H]}$ ergodically, it is still possible for some $h \in H$, $\mathrm{U}[gh]_\Gamma$ is trapped in a closed homogeneous set of smaller dimension.

DEFINITION 3.1. *For $H \in \mathscr{H}$, define*

$$\begin{aligned} \mathrm{Sing}(H,\mathrm{U}) &:= \bigcup_{L \in \mathscr{H}, L \subsetneq H} N(L,\mathrm{U}); \\ \mathrm{NS}(H,\mathrm{U}) &:= N(H,\mathrm{U}) \setminus \mathrm{Sing}(H,\mathrm{U}); \\ T(H,\mathrm{U}) &:= \mathrm{NS}(H,\mathrm{U})\Gamma/\Gamma. \end{aligned}$$

LEMMA 3.2. *Let $H_1, H_2 \in \mathscr{H}$. If $\mathrm{NS}(H_1,\mathrm{U})\Gamma \cap \mathrm{NS}(H_2,\mathrm{U})\Gamma \neq \emptyset$, then H_1 is Γ-conjugate to H_2 and $\mathrm{NS}(H_1,\mathrm{U})\Gamma = \mathrm{NS}(H_2,\mathrm{U})\Gamma$.*

PROOF. So assume $\mathrm{NS}(H_1,\mathrm{U})\Gamma \cap \mathrm{NS}(H_2,\mathrm{U})\Gamma \neq \emptyset$, which means that there exist $g_1 \in \mathrm{NS}(H_1,\mathrm{U})$ and $\gamma_1 \in \Gamma$ such that $g_1\gamma_1 \in \mathrm{NS}(H_2,\mathrm{U})$. By definition, we have

$$g_1^{-1}\mathrm{U}g_1 \subset H_1 \cap \gamma_1 H_2 \gamma_1^{-1}.$$

We know (the connected component of) $H' := H_1 \cap \gamma_1 H_2 \gamma_2^{-1}$ has a closed orbit based at $[\mathrm{id}]_\Gamma$. But we do not know whether it supports a finite H'-invariant measure. This is where we apply Thm.2.3 (to the unipotent group $g_1^{-1}\mathrm{U}g_1$ and $x_0 = [\mathrm{id}]_\Gamma$) to conclude that there exists $L \subset H'$, $L \in \mathscr{H}$ such that $g_1^{-1}\mathrm{U}g_1 \subset L$. So $g_1 \in N(L,\mathrm{U})$.

Therefore $H_1 = \gamma_1 H_2 \gamma_1^{-1}$ for otherwise L will be strictly contained in at least one of H_1 or $\gamma_1 H_2 \gamma_1^{-1}$ and this would imply $g_1 \notin \mathrm{NS}(H_1,\mathrm{U})$ or $g_1\gamma_1 \notin \mathrm{NS}(H_2,\mathrm{U})$, contradicting against our assumption. $\mathrm{NS}(H_1,\mathrm{U})\Gamma = \mathrm{NS}(H_2,\mathrm{U})\Gamma$ follows immediately. $\square$

Since U acts ergodically on m_{X} (it is even mixing!), we have $\mathrm{G} \in \mathscr{H}$ and

$$\mathrm{X} = \bigsqcup_{[H] \in \mathscr{H}/\sim_{\Gamma}} T(H, \mathrm{U})$$

thanks to the Lem.3.2.

DEFINITION 3.3. *For $[H] \in \mathscr{H}/\sim_{\Gamma}$, let*

$$\mu^{[H]} := \mu|_{T(H,\mathrm{U})}.$$

EXAMPLE 3.4. *If $\mathrm{X} = \mathrm{SL}_2(\mathbb{R})/\mathrm{SL}_2(\mathbb{Z})$ and $\mathrm{U} = \left\{ \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle| s \in \mathbb{R} \right\}$, then $\mathscr{H}/\sim_{\Gamma} = \{\mathrm{U}, \mathrm{SL}_2(\mathbb{R})\}$ (if you pass to a smaller subgroup of $\mathrm{SL}_2(\mathbb{Z})$ then this set has other $\mathrm{SL}_2(\mathbb{Q})$-conjugates of U that are not conjugate over Γ). Here $T(\mathrm{U},\mathrm{U})$ consists of compact orbits of U, $T(\mathrm{G},\mathrm{U})$ is the complement of $T(\mathrm{U},\mathrm{U})$.*

See Sec.4 for more examples.

Assuming Lem.3.6, we have proved

THEOREM 3.5. *For a U-invariant probability measure μ,*

$$\mu = \sum_{[H] \in \mathscr{H}/\sim_{\Gamma}} \mu^{[H]}$$

and each $\mu^{[H]}$ is U-invariant.

LEMMA 3.6. *$\mathscr{H}$ is countable.*

PROOF. For every $H \in \mathscr{H}$, $H \cap \Gamma$ is a lattice in H. Thus $H \cap \Gamma$ is finitely generated. Note that this seems not obvious unless $H \cap \Gamma$ is cocompact. In the case at hand, H is algebraic by Thm.2.3. Thus $H \cap \Gamma$ is an arithmetic lattice of H and finite generation follows from the theory of Siegel sets, see [**Bor19**]. See also [**Gel14**, Lecture 3, Section 5] for another possibly more geometric proof. Hence the set $\{H \cap \Gamma,\ H \in \mathscr{H}\}$ is countable.

Since H can be recovered from $H \cap \Gamma$ by

$$H = \left(\overline{\overline{H \cap \Gamma}} \cap \mathrm{SL}_n(\mathbb{R}) \right)^{\circ},$$

we are done. Here $\overline{\overline{H \cap \Gamma}}$ means the closure of $H \cap \Gamma$ in $\mathrm{SL}_n(\mathbb{C})$ with respect to the topology defined by polynomials. □

We have not used Thm.1.1 yet. For a finite positive measure μ on X, let $\widehat{\mu} := \mu/\mu(X)$ be the unique probability measure proportional to μ.

THEOREM 3.7. *Assume $\mu^{[H]} \neq 0$. For almost every U-ergodic component ν of $\widehat{\mu^{[H]}}$, there exists $g_\nu \in N(H,\mathrm{U})$ such that $\nu = (g_\nu)_* \mathrm{m}_{[H]_\Gamma}$.*

PROOF. First we have the (abstract) ergodic decomposition

$$\widehat{\mu^{[H]}} = \int_{\mathrm{Prob(X)}^{\mathrm{U,Erg}}} \nu \, \lambda(\nu).$$

Thus for almost every ν, $\nu(T(H,\mathrm{U})) = 1$. Take such a ν, by Thm.1.2, there exists $H_1 \in \mathscr{H}$ and $g_1 \in N(H_1,\mathrm{U})$ such that $\nu = (g_1)_* \widehat{\mathrm{m}}_{[H_1]_\Gamma}$. By pointwise ergodic theorem, we can find a full measure set of $h_1 \in H_1$ such that

$$\lim_{S \to +\infty} \frac{1}{S} \int_0^S (\mathbf{u}_s)_* \delta_{[g_1 h_1]_\Gamma} \, \mathrm{d}s = (g_1)_* \widehat{\mathrm{m}}_{[H_1]_\Gamma}.$$

In particular, $\overline{\mathrm{U}.[g_1 h_1]_\Gamma} = g_1 [H_1]_\Gamma$. One sees that $g_1 h_1 \in N(H_1,\mathrm{U})$ and we claim that $g_1 h_1 \in \mathrm{NS}(H_1,\mathrm{U})$. Otherwise, there exists $L \subsetneqq H_1$ with $L \in \mathscr{H}$ such that $g_1 h_1 \in N(L,\mathrm{U})$. This implies that $\overline{\mathrm{U}.[g_1 h_1]_\Gamma} \subset g_1 h_1 [L]_\Gamma$. Since $\dim L$ is strictly smaller than $\dim H_1$, we have a contradiction.

So now $[g_1 h_1]_\Gamma \in T(H_1, \mathrm{U})$, moreover, all such h_1's are of full measure in H_1 and consequently $\nu(T(H_1,\mathrm{U})) = (g_1)_* \widehat{\mathrm{m}}_{[H_1]_\Gamma}(T(H_1,\mathrm{U})) = 1$. But $\nu(T(H,\mathrm{U})) = 1$. Thus $T(H_1,\mathrm{U})$ and $T(H,\mathrm{U})$ have nontrivial intersection. By Lem.3.2, for some $\gamma_1 \in \Gamma$, $H_1 = \gamma_1 H \gamma_1^{-1}$ and $T(H_1,\mathrm{U}) = T(H,\mathrm{U})$. Hence $[g_1 H_1]_\Gamma = g_1\gamma_1[H]_\Gamma$. Let $g_\nu := g_1\gamma_1$. One can check that $g_\nu \in N(H,\mathrm{U})$ and $\nu = (g_\nu)_* \widehat{\mathrm{m}}_{[H]_\Gamma}$ □

4. Two examples

Here we include two examples, a little bit beyond $\mathrm{SL}_2(\mathbb{R})$, to illustrate what kind of objects we are dealing with. You are welcome to test the general theory using these (still rather special) examples!

In both examples, set

- $G = \mathrm{SL}_2(\mathbb{C})$, $\Gamma = \mathrm{SL}_2(\mathbb{Z}[i])$, $U = \left\{ \mathbf{u}_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle| s \in \mathbb{R} \right\}$;
- $\mathbf{U}(\mathbb{C}) := \left\{ \mathbf{u}_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle| s \in \mathbb{C} \right\}$
- for $t \in \mathbb{C}^\times$, let $\mathbf{a}_t := \begin{bmatrix} t & 0 \\ 0 & t^{-1} \end{bmatrix}$.

One can show that Γ is a lattice in G (using non-divergence of unipotent flows, for instance). And G/Γ can be embedded in $\mathrm{SL}_4(\mathbb{R})/\mathrm{SL}_4(\mathbb{Z})$ (so this example does not escape away from the setting in this chapter).

4.1. Example 1.

- $H := \left\{ \mathbf{u}_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle| s \in \mathbb{C} \right\}$, $\mathfrak{h}$ is the Lie algebra of H;
- $K_H := \{\mathbf{a}_z,\ z \in \mathbb{C},\ |z| = 1\}$.

LEMMA 4.1. $N(H,U) = N_G(H) = \left\{ \mathbf{a}_t \cdot \mathbf{u}_s,\ t \in \mathbb{C}^\times,\ s \in \mathbb{C} \right\} =: B$.

PROOF. Let $g \in G$. Indeed, g belongs to $N(H,U)$ iff $\mathrm{Ad}(g)\cdot\mathfrak{h}$ contains $\mathfrak{u}$. By Bruhat decomposition (ref??),

$$G = BwB \sqcup B$$

where $w = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. If $g \in B$, then $\mathrm{Ad}(g)\cdot\mathfrak{h} = \mathfrak{h} \supset \mathfrak{u}$. On the other hand, if $g = b_1 w b_2$ for $b_i \in B$ then

$$\mathrm{Ad}(g)\cdot\mathfrak{h} = Ad(b_1)Ad(w)\cdot\mathfrak{h} = Ad(b_1)\cdot\begin{bmatrix} 0 & 0 \\ * & 0 \end{bmatrix} \Longrightarrow \mathrm{Ad}(g)\cdot\mathfrak{h}\cap\mathfrak{h} = \{0\}.$$

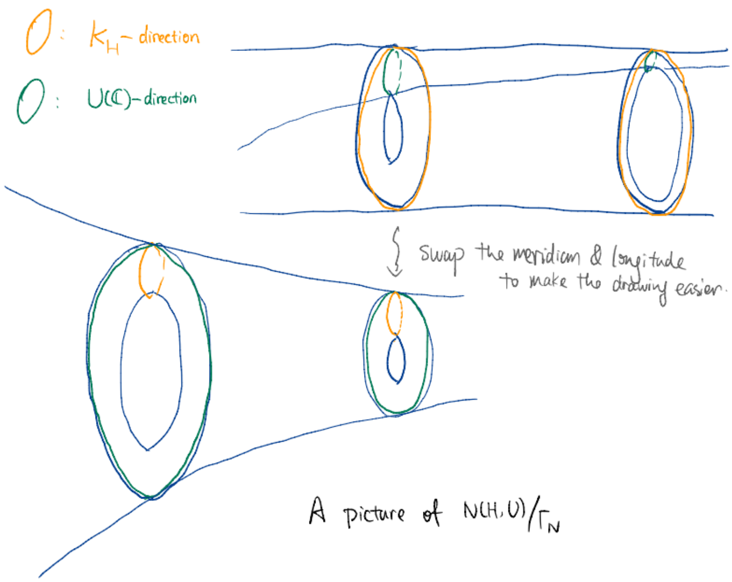
So we are done. □

The orbits of B on G/Γ are all dense, and hence not easy to draw. Since $N(H,U)$ (= B here) is stable under right translation by $N_G(H)$ and therefore $N_G(H)\cap\Gamma$ (call it Γ_N for simplicity). Thus $N(H,U)$ being closed implies that $N(H,U)/\Gamma_N \subset G/\Gamma_N$ is closed. We will draw pictures for $N(H,U)/\Gamma_N$. (warning! pictures are just for illustration, they may be wrong in many aspects!)

By the way, a quick computations show that

$$\Gamma_N = \left\{ \begin{bmatrix} 1 & \mathbb{Z}[i] \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} -1 & \mathbb{Z}[i] \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} i & \mathbb{Z}[i] \\ 0 & -i \end{bmatrix} \right\}.$$

Here is a picture for $N(H,U)/\Gamma_N$ with $U\Gamma_N/\Gamma_N$ contained in here:

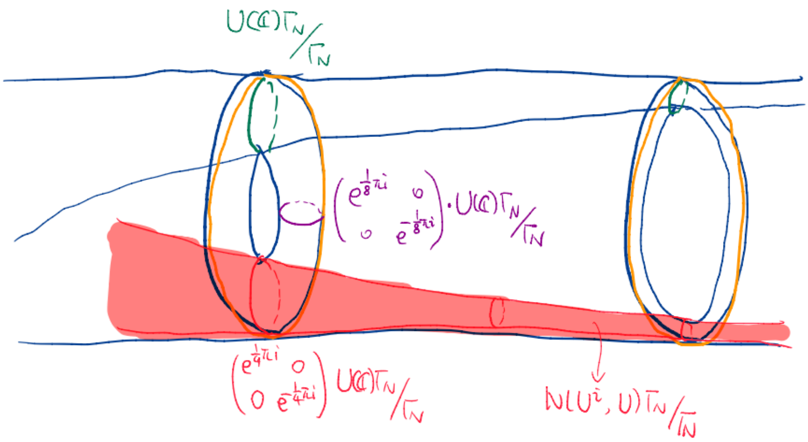


What about $\mathrm{Sing}(H,U)$? The possible $L \in \mathscr{H}$ and $L \subsetneq H$ are given as follows. For $z \in \mathbb{C}$, let $U^z := \left\{ \mathbf{u}_{s\cdot z} = \begin{bmatrix} 1 & sz \\ 0 & 1 \end{bmatrix} \middle| s \in \mathbb{R} \right\}$. Then every proper nontrivial connected subgroup of H is of this form. And $U^z \in \mathscr{H}$ iff $U^z \cap \Gamma$ is a lattice in U^z iff $\mathbb{R}.z \cap \mathbb{Z}[i] \le \mathbb{R}.z$ is a lattice.

Note that

$$N(U^z, U) = \begin{bmatrix} \sqrt{z}^{-1} & 0 \\ 0 & \sqrt{z} \end{bmatrix} \cdot N_G(U^z) = \begin{bmatrix} \sqrt{z}^{-1} & 0 \\ 0 & \sqrt{z} \end{bmatrix} \cdot \{\mathbf{a}_t,\ t \in \mathbb{R}^\times\} \cdot \mathbf{U}(\mathbb{C})$$

And $\mathrm{Sing}(H,U)$ is the union of these $N(U^z,U)$ as z varies over $\mathbb{Z}[i]$.



4.2. Example 2.

- $H := \mathrm{SL}_2(\mathbb{R})$, $U^i := \{\mathbf{u}_{is},\ s \in \mathbb{R}\}$.

LEMMA 4.2. $N(H,U) = U^i \cdot \mathrm{SL}_2(\mathbb{R}) \sqcup U^i \cdot \mathrm{SL}_2(\mathbb{R}) \cdot \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$.

Note that $U^i \cdot \mathrm{SL}_2(\mathbb{R}) \cdot \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix} \cdot U^i \cdot \mathrm{SL}_2(\mathbb{R}) = \mathbf{U}(\mathbb{C}) \cdot \mathrm{SL}_2(\mathbb{R}) \cdot \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$.

PROOF. As before, $N(H,U) = \{g \in G, \mathrm{Ad}(g)\cdot\mathfrak{h} \supset \mathfrak{u}\}$. It is direct to observe that RHS is a subset of LHS. It remains to do the converse.

Recall Bruhat decomposition again: $G = BwB \sqcup B$. If $g \in B$, then we are done since B is contained in the right hand side.

Now assume $g \in BwB$. Every element b of B can be written as $\mathbf{a}_t\mathbf{u}_s$ with $t \in \mathbb{C}^\times, s \in \mathbb{C}$. Since w normalizes $\{\mathbf{a}_t, t \in \mathbb{C}^\times\}$, we can write

$$g^{-1} = u_2\mathbf{a}_{t_1}wu_1, \quad \exists u_1, u_2 \in \mathbf{U}(\mathbb{C}), t_1 \in \mathbb{C}^\times.$$

Thus (to save notation we omit Ad in the following)

$$\begin{aligned} g^{-1}\cdot\mathfrak{u} =& (u_2\mathbf{a}_{t_1}wu_1)\cdot\mathfrak{u} = (u_2\mathbf{a}_{t_1})\cdot\begin{bmatrix} 0 & 0 \\ \mathbb{R} & 0 \end{bmatrix} \\ =& u_2\cdot\begin{bmatrix} 0 & 0 \\ t_1^{-2}\mathbb{R} & 0 \end{bmatrix} = \begin{bmatrix} * & * \\ t_1^{-2}\mathbb{R} & * \end{bmatrix} \subset \mathfrak{h} = \mathfrak{sl}_2(\mathbb{R}). \end{aligned}$$

Thus $t_1^{-2}\mathbb{R} \subset \mathbb{R} \implies t_1 \in \mathbb{R} \cup i\mathbb{R}$. In either case (write $u_1 = \mathbf{u}_{z_1}$ for some $z_1 \in \mathbb{C}$),

$$g^{-1}\cdot\mathfrak{u} = u_2\cdot\begin{bmatrix} 0 & 0 \\ \mathbb{R} & 0 \end{bmatrix} = \begin{bmatrix} 1 & z_1 \\ 0 & 1 \end{bmatrix}\cdot\begin{bmatrix} 0 & 0 \\ \mathbb{R} & 0 \end{bmatrix}\cdot\begin{bmatrix} 1 & -z_1 \\ 0 & 1 \end{bmatrix} = \mathbb{R}\cdot\begin{bmatrix} z_1 & -z_1^2 \\ 1 & -z_1 \end{bmatrix} \subset \mathfrak{sl}_2(\mathbb{R}).$$

Thus $z_1 \in \mathbb{R}$. And the proof completes. □

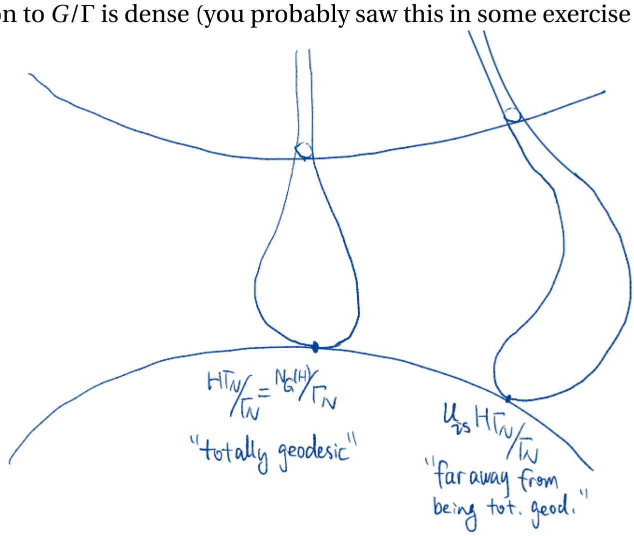
The above proof also shows that

LEMMA 4.3. $N_G(H) = \mathrm{SL}_2(\mathbb{R}) \sqcup \mathrm{SL}_2(\mathbb{R})\cdot\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$.

And hence one can check that

LEMMA 4.4. $\Gamma_N := N_G(H) \cap \Gamma = \mathrm{SL}_2(\mathbb{Z}) \sqcup \mathrm{SL}_2(\mathbb{Z})\cdot\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$.

Below is a picture of $N(H,U)/\Gamma_N$ sitting inside G/Γ_N as a closed subset. Note that its projection to G/Γ is dense (you probably saw this in some exercise section).

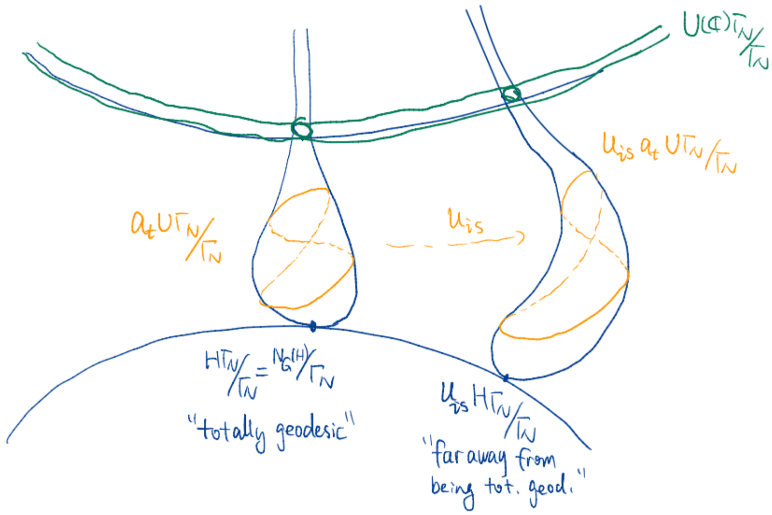


A picture of $N(H,U)/\Gamma_N$

Up to Γ_N-conjugacy, the only proper nontrivial connected subgroup of H containing U is just U itself. Thus $\mathrm{Sing}(H,U) = N(U,U)\Gamma_N$.

LEMMA 4.5. *$N(U,U) = N_G(U)$ and is generated by* $\{\mathbf{a}_t \cdot \mathbf{u}_s,\ t \in \mathbb{R}^\times,\ s \in \mathbb{R}\} \cup \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$.

Thus the picture is not so new. Note that since U^i commutes with U, U^i translates of $U\Gamma_N/\Gamma_N$ does not "twist" the appearance of $U\Gamma_N/\Gamma_N$ (unlike U^i translates of $H\Gamma_N/\Gamma_N$).



Remark/Question: if you let $s,t \to +\infty$ at the same time, it is unclear to me (asymptotically) how $U_{is} a_t U\Gamma_N/\Gamma_N$ look like relative to $U_{is} H\Gamma_N/\Gamma_N$

5. Exercises

CHAPTER 12

Linearization technique

Back to the Top.

In this chapter we are going to discuss the linearization method due to Dani–Margulis [DM93]. We will illustrate the method by proving an equidistribution statement, which on the one hand implies Oppenheim conjecture, on the other hand forms one ingredient of quantitative Oppenheim. Same method can be used to deduce equidistribution of unipotent flows, and hence to classify orbit closures of orbits of unipotent flows from the measure classification theorem. This will be left as a difficult exercise.

The main reference of this chapter is [MS95, Section 3]. Other related resources include Shah [Sha91b, Sha91a, Sha09], Ratner [Ra91a, Ra91b], Eskin–Mozes–Shah [EMS96], Eskin–Margulis–Mozes [EMM98, Section 4]. An effective treatment appears in [LMMS19].

Recall the notations when we discuss Oppenheim conjecture.

- $\mathrm{G} = \mathrm{SL}_3(\mathbb{R})$, $\Gamma = \mathrm{SL}_3(\mathbb{Z})$, $\mathrm{X} := \mathrm{G}/\Gamma$;
- $\widehat{\mathrm{m}}_{\mathrm{X}}$ is the unique G-invariant probability measure on X;
- $\mathrm{H}_0 := \mathrm{SO}_{Q_0}(\mathbb{R})$ with $Q_0(x_1, x_2, x_3) = 2x_1x_3 - x_2^2$;
- $\mathrm{U} = \left\{ \mathbf{u}_s := \exp\left(s \cdot \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \right) \middle| \, s \in \mathbb{R} \right\} \subset \mathrm{H}_0$;
- $\mathrm{A} = \left\{ \mathbf{a}_t := \exp\left(t \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \right) \middle| \, t \in \mathbb{R} \right\} \subset \mathrm{H}_0$;
- also we fix some $g_0 \in \mathrm{G}$ and $x_0 := [g_0]_\Gamma := g_0\Gamma/\Gamma$.

1. Statement

As we explained, Oppenheim conjecture follows once we prove

THEOREM 1.1. *If $Q := Q_0 \circ g_0$ is irrational, then $\mathrm{H}_0 g_0 \Gamma/\Gamma$ is dense in X.*

Actually something weaker is proved in Ch.5, Thm.1.2, which is sufficient. Now we would like to explain how to use Ratner's description of ergodic U-invariant probability measures to prove this stronger claim.

The idea is as follows. Take

$$K_0 := (\mathrm{H}_0 \cap \mathrm{SO}_3(\mathbb{R}))^\circ,$$

a maximal compact subgroup of H_0°. Then $\mathbf{a}_t K_0.x_0 \subset \mathrm{H}_0.x_0$ and we seek to show that as $t \to +\infty$ ($-\infty$ is also ok), $\mathbf{a}_t K_0.x_0$ becomes dense in X. And this is achieved by the following equidistribution theorem

THEOREM 1.2. *Let $\widehat{\mathrm{m}}_{K_0.x_0}$ be the unique K_0-invariant probability measure on $K_0.x_0$. Then*

$$\lim_{t \to +\infty} (\mathbf{a}_t)_* \widehat{\mathrm{m}}_{K_0.x_0} = \widehat{\mathrm{m}}_{\mathrm{X}}$$

in weak$*$ topology.

REMARK 1.3. *From the proof, you will see that $\widehat{\mathrm{m}}_{K_0.x_0}$ can be replaced by any other probability measure that is absolutely continuous with respect to this one without affecting the conclusion.*

REMARK 1.4. *Instead of K_0, you can also use other subgroups of H_0 and prove analogues of the theorem above. Actually it would be easier if we replace K_0 by a bounded open subset of H_0. However, I prefer to do this in preparation for our later discussion on quantitative Oppenheim. This change only has an effect on Sec.11.*

2. Step 1, nondivergence

Let μ be a limit of $(\mu_t) := ((\mathbf{a}_t)_* \widehat{\mathrm{m}}_{K_0.x_0})$ as $t \to +\infty$.

LEMMA 2.1. $\mu \in \mathrm{Prob}(\mathrm{X})$.

In other words, there is no escape of mass. This is a consequence of (C,α)-good property and a lemma in representation theory/linear algebra.

3. Step 2, unipotent invariance

LEMMA 3.1. *μ is U-invariant.*

PROOF. Since $\widehat{\mathrm{m}}_{K_0.x_0}$ is K_0-invariant, μ_t is $\mathbf{a}_t K_0 \mathbf{a}_t^{-1}$-invariant. Hence μ is invariant under the limit group, which turns out to be U.

More details: Let $\mathfrak{k}_0$ be the Lie algebra of K_0. Take $v_t \in \mathrm{Ad}(\mathbf{a}_t).\mathfrak{k}_0$, if $\lim v_t = v$, then by continuity of the induced map $\mathrm{G} \times \mathrm{LFM}(X) \to \mathrm{LFM}(X)$, μ is $\exp(v)$-invariant.

Recall that the Lie algebra of H_0 is

$$\mathfrak{so}_{Q_0} = \left\{ \begin{bmatrix} x_{11} & x_{12} & 0 \\ x_{21} & 0 & x_{12} \\ 0 & x_{21} & -x_{11} \end{bmatrix} \right\}.$$

And the Lie algebra of $\mathrm{SO}_3(\mathbb{R})$ is given by anti-symmetric matrices. Thus by taking their intersection:

$$\mathfrak{k}_0 = \left\{ \begin{bmatrix} 0 & x_{12} & 0 \\ -x_{12} & 0 & x_{12} \\ 0 & -x_{12} & 0 \end{bmatrix} \right\}.$$

And

$$\mathrm{Ad}(\mathbf{a}_t)\mathfrak{k}_0 = \left\{ \begin{bmatrix} 0 & e^t x_{12} & 0 \\ -e^{-t} x_{12} & 0 & e^t x_{12} \\ 0 & -e^{-t} x_{12} & 0 \end{bmatrix} \right\}.$$

So depending on $s \in \mathbb{R}$, we take

$$v_t := \begin{bmatrix} 0 & s & 0 \\ -e^{-2t} s & 0 & s \\ 0 & -e^{-2t} s & 0 \end{bmatrix} \in \mathrm{Ad}(\mathbf{a}_t)\mathfrak{k}_0.$$

Then as s varies, $\lim v_t$ fills $\mathfrak{u}$, the Lie algebra of U. □

Thus by the first two steps we get (by passing to a subsequence)

$$\lim \mu_t = \mu \in \mathrm{Prob}(\mathrm{X})^{\mathrm{U}}.$$

4. Step 3, ergodic components and tubes

By Thm.3.5 and 3.7 from Chapter 11, to show $\mu = \widehat{\mathrm{m}}_{\mathrm{X}}$, it suffices to show that for every $H \in \mathscr{H}$, $H \neq \mathrm{G}$, $\mu(T(H,\mathrm{U})) = 0$. Note that Ratner's theorem Ch.11, Thm.1.1 is only used to go from $\mu(T(H,\mathrm{U})) = 0$ to $\mu = \widehat{\mathrm{m}}_{\mathrm{X}}$. To show $\mu(T(H,\mathrm{U})) = 0$, we do not need it. The way to achieve this is via:

LEMMA 4.1. *For every compact subset E of $T(H,\mathrm{U})$ and $\varepsilon > 0$, there exists a neighborhood $\mathscr{N}_\varepsilon$ of E such that*

$$\limsup_{t\to+\infty} \mu_t(\mathscr{N}_\varepsilon) \leq \varepsilon.$$

In view of Ch.10 (cf. [DS84]), we hope to find a bigger $\mathscr{N}_\varepsilon'$ such that

$$\mu_t(\mathscr{N}_\varepsilon) \leq \varepsilon \mu_t(\mathscr{N}_\varepsilon').$$

Since μ_t is a probability measure, this finishes the proof.

5. Step 4, a lemma on linear representations

Though we do not know how to find $\mathscr{N}_\varepsilon \subset \mathscr{N}_\varepsilon'$ at the moment, we do have something like this happening in a representation (rather than the complicated G/Γ) due to the (C,α)-good property. To give us more freedom (see below, the choice of Φ) for things to come, we need a slightly more flexible statement.

DEFINITION 5.1. *Fix a non-empty connected bounded open set $\mathrm{D} \subset \mathfrak{k}_0$, let*

$$\psi_t : \mathrm{D} \to \mathrm{G}, \quad x \mapsto \psi_t(x) := \mathbf{a}_t \exp(x).$$

LEMMA 5.2. *Let V be a representation of G. Let W be a linear subspace of V. For every compact subset E of W and every $\varepsilon > 0$, there exists another compact set $F \subset W$ such that the following is true. For every open neighborhood Φ of F, there exists an open neighborhood Ψ of E such that for every $t \in \mathbb{R}$, $v \in V$, every ball $\mathrm{B} \subset \mathrm{D}$, at least one of the following is true*

1. $\overline{\psi_t(\mathrm{B}).v} \subset \Phi$;
2. $\mathrm{Leb}\left\{x \in \mathrm{B} \,\middle|\, \psi_t(x).v \in \Psi\right\} \leq \varepsilon \mathrm{Leb}\left\{x \in \mathrm{B} \,\middle|\, \psi_t(x).v \in \Phi\right\}$.

REMARK 5.3. *The first possibility can often be excluded due to "algebraic" reasons (for instance, see Sec.11). And the second option is what we want.*

REMARK 5.4. *You can replace* Leb *by any other measure equivalent to* Leb*, it is just that the choice of Ψ may depend on this measure. Actually in application we have in mind,* Leb *should be replaced by some measure which maps to $\widehat{\mathrm{m}}_{K_0 x_0}$ under the exponential and the orbit map. We are going to ignore this issue in the following.*

6. Step 5, representation and dynamics, naive ideas

Let $\Gamma_N := \Gamma \cap N_{\mathrm{G}}(H)$.

DEFINITION 6.1. *Let*

$$N_{\mathrm{G}}(H)^{(1)} := \left\{g \in N_{\mathrm{G}}(H) \,\middle|\, \det\left(\mathrm{Ad}(g), \mathfrak{h}\right) = \pm 1\right\}$$

LEMMA 6.2. $\Gamma_N = \Gamma \cap N_{\mathrm{G}}(H)^{(1)}$.

Take a representation V_H of G and a vector $v_H \in V_H$ such that the stabilizer of v_H (or just $\pm v_H$) in G is equal to $N_{\mathrm{G}}(H)^{(1)}$. Moreover, we want V_H to be equipped with a $\mathbb{Q}$-structure (i.e., fix a copy of $\mathbb{Q}^{\dim V_H}$ in V_H, call it $V_H(\mathbb{Q})$) and $v_H \in V_H(\mathbb{Q})$. A priori, $N_{\mathrm{G}}(H)^{(1)}$ is not known to be "observable" , the existence of such a pair (V_H, v_H) is not obvious. But one can take $V_H := \bigwedge^{\dim H} \mathfrak{sl}_n$ and $v_H := v_1 \wedge \ldots \wedge v_{\dim H}$ where $(v_1, \ldots, v_{\dim H})$ is a basis of $\mathfrak{h}$. For this specific choice of v_H, the stabilizer of $\pm v_H$ in G is equal to $N_{\mathrm{G}}(H)^{(1)}$. You may also have other choices. For instance when $\mathrm{G} = \mathrm{SL}_2(\mathbb{R})$ and H is equal to the upper triangular unipotent group, then V_H can be taken to be the standard representation $\mathbb{R}^2$ and $v_H = e_1$.

To go from the representation V_H to G/Γ, the following diagram is very natural.

$$\begin{tikzcd}
& \mathrm{G}/\Gamma_N \arrow[dl, "p"'] \arrow[dr, "q"] & & \\
\mathrm{G}/\Gamma & & \mathrm{G}/N_{\mathrm{G}}(H)^{(1)} \arrow[r, hook, "\phi"] & V_H.
\end{tikzcd}$$

Here p and q are natural projections and $\phi([g]) := g.v_H$. Strictly speaking ϕ may only be injective replacing V_H by $V_H/\pm 1$, but we will ignore this minor issue.

Here is something naive one can do at this stage. Recall $E \subset T(H,\mathrm{U})$ is a compact set.

1. Take a compact subset $\widetilde{E} \subset N(H,\mathrm{U})$ such that $E = [\widetilde{E}]_\Gamma$, the image of $\widetilde{E}$ in G/Γ;
2. Let $E^\vee := \phi \circ q(\widetilde{E}) = \widetilde{E}.v_H$;
3. Apply Lem.5.2 above to $E = E^\vee$ and W to be determined (you may take $W = V$ and see why it does NOT work). Then we get F (depending also on ε) by Lem.5.2, which asserts that for every open neighborhood Φ (we do not have a favorite Φ yet, so just fix some) there exists an open neighborhood Ψ of $E^\vee$ such that something holds.
4. We simply take $\mathcal{N}_\varepsilon := p\left((\phi\circ q)^{-1}\Psi\right)$ and $\mathcal{N}'_\varepsilon := p\left((\phi\circ q)^{-1}\Phi\right)$.

To simplify notations,

Definition 6.3. *For $t \in \mathbb{R}$ and $[\gamma]_{\Gamma_N} \in \Gamma/\Gamma_N$,*

$$\mathrm{D}_t(\mathcal{N}_\varepsilon) := \left\{ y \in \mathrm{D} \,\middle|\, \psi_t(y).x_0 \in \mathcal{N}_\varepsilon \right\}$$
$$\mathrm{D}_t(\Psi, [\gamma]_{\Gamma_N}) := \left\{ y \in \mathrm{D} \,\middle|\, \psi_t(y) g_0 \gamma.v_H \in \Psi \right\}.$$

And define $\mathrm{D}_t(\Psi) := \bigcup_{[\gamma]\in\Gamma/\Gamma_N} \mathrm{D}_t(\Psi,[\gamma])$. Similarly define $\mathrm{D}_t(\mathcal{N}'_\varepsilon)$, $\mathrm{D}_t(\Phi,[\gamma])$ and $\mathrm{D}_t(\Phi)$.

Thus from the definition (the naive definition of $\mathcal{N}_\varepsilon$ above, we will work with a different $\mathcal{N}_\varepsilon$ later in Sec.8)

$$\mathrm{D}_t(\mathcal{N}_\varepsilon) = \bigcup \mathrm{D}_t(\Psi, [\gamma]_{\Gamma_N}).$$

7. Step 6, self-intersection

Now we seek to refine the rather crude strategy proposed in Step 5 so that it would actually work.

First of all in general, unlike the $\mathrm{SL}_2(\mathbb{R})$-case, the projection $N(H,\mathrm{U})/\Gamma_N \to \mathrm{G}/\Gamma$ is not injective.

Lemma 7.1. *If $g \in \mathrm{G}$ is such that for two different $[\gamma_1]_{\Gamma_N} \neq [\gamma_2]_{\Gamma_N} \in \Gamma/\Gamma_N$ we have $g\gamma_i \in N(H,\mathrm{U})$ for $i = 1,2$, then $g \in \mathrm{Sing}(H,\mathrm{U})\Gamma$.*

So ideally we would like to avoid $\mathrm{Sing}(H,\mathrm{U})$ (or its projection to G/Γ_N, or G/Γ) from our discussion. But this is impossible! Since usually $\mathrm{Sing}(H,\mathrm{U})$ is dense in $N(H,\mathrm{U})$ modulo Γ_N, every non-empty open set intersects non-trivially with it. Lucky for us, each time we only work with certain compact set F (to be found) in V_H (and we have the freedom of choosing its neighborhood). And the subset of $\mathrm{Sing}(H,\mathrm{U})$ that is "relevant to F" is indeed closed, see Lem.7.5.

To detect $N(H,\mathrm{U})$ inside V_H, it is convenient (though maybe not necessary) to have:

Definition 7.2. *Let W_H be the $\mathbb{R}$-linear subspace of V_H spanned by $N(H,\mathrm{U}).v_H$.*

This W_H would be the W when we apply Lem.5.2 above.

LEMMA 7.3. *We have*

$$(\phi \circ q)^{-1}(W_H) = N(H, \mathrm{U})/\Gamma_N.$$

The reader is reminded that being compact in V_H is not the same as being compact in $\mathrm{G}/N_{\mathrm{G}}(H)^{(1)}$, as the $\mathrm{SL}_2(\mathbb{R})$-case already told us, unless $\mathrm{G}.v_H$ is closed in V_H, which is true if H is reductive by [**Kem78**] or if Γ is arithmetic and cocompact in G, for other reductive G's.

Here is the important observation

DEFINITION 7.4. *Let F be a compact subset of W_H, let*

$$\mathrm{Sing}(F) := \left\{ g \in \mathrm{G} \,\middle|\, g.v_H \in F,\ g\gamma.v_H \in F\ \exists \gamma \in \Gamma \setminus \Gamma_N \right\}$$

Thus $\mathrm{Sing}(F) \subset \mathrm{Sing}(H, \mathrm{U})$. The fact we need is that

LEMMA 7.5. $\mathrm{Sing}(F)\Gamma$ *is closed.*

SKETCH OF PROOF. First note that $\Gamma.v_H$ is discrete in V_H. This is rather straight-forward since v_H is a rational vector and the image of Γ in $\mathrm{SL}(V_H)$ is commensurable with $\mathrm{SL}_N(\mathbb{Z})$ for $N = \dim V_H$. For non-arithmetic lattices, see [**DM93**] for a proof.

Therefore if (g_n) is bounded mod Γ and $(g_n.v_H)$ is bounded in V_H, then (g_n) is bounded modulo Γ_N. The conclusion follows quickly from here. $\square$

Note that

$$\mathrm{Sing}(F)\Gamma := \left\{ g \in \mathrm{G} \,\middle|\, g\gamma_1.v_H \in F,\ g\gamma_2.v_H \in F,\ \exists [\gamma_1] \neq [\gamma_2] \in \Gamma/\Gamma_N \right\}$$

Consequently, by a continuity argument and the discreteness of $\Gamma.v_H$,

LEMMA 7.6. *Let E' be a compact set in $X \setminus [\mathrm{Sing}(F)]_\Gamma$. Then there exists an open neighborhood Φ of F such that for every $[g]_\Gamma \in E'$,*

$$\#\left\{ [\gamma] \in \Gamma/\Gamma_N \,\middle|\, g\gamma.v_H \in \Phi \right\} \leq 1.$$

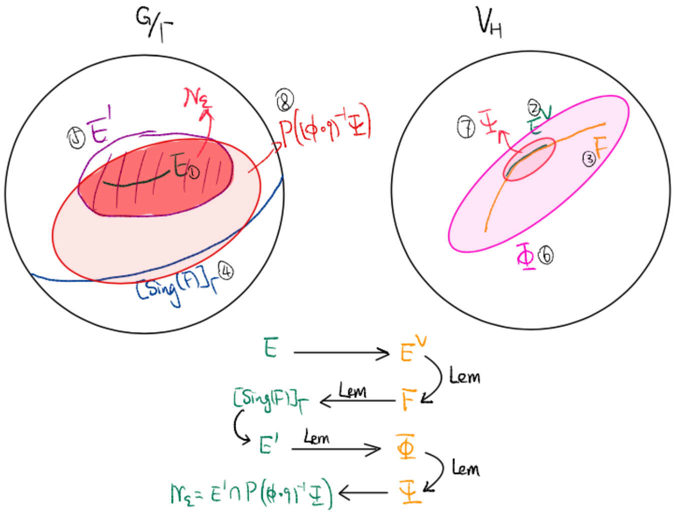
8. Step 7, define the neighborhood

Let us explain how to find $\mathcal{N}_\varepsilon$. Fix $E \subset T(H, \mathrm{U})$ and $\varepsilon > 0$.

Define $E^\vee$ as in Sec.6. By taking $W = W_H$ (see Def.7.2), Lem.5.2 offers some compact set F of W_H. By Lem.7.5, $[\mathrm{Sing}(F)]_\Gamma$ is closed and is contained $[\mathrm{Sing}(H, \mathrm{U})]_\Gamma$ by Lem.7.1.

Now we take E' to be any compact set away from $[\mathrm{Sing}(H, \mathrm{U})]_\Gamma$ whose interior contains E. We find an open neighborhood Φ of F such that the conclusion of Lem.7.6 holds. Then Ψ, an open neighborhood of E, is chosen according to Lem.5.2.

Just in case one gets confused, here is a diagram summarizing the logical dependence:



Now

$$\mathscr{N}_\varepsilon := \mathrm{Int}(E') \cap p\left((\phi \circ q)^{-1}\Psi\right)$$

9. Step 8, a covering argument

The proof will be concluded with the help of a covering argument, something we encountered when discussing nondivergence of unipotent flow on X_N. The argument here seems to differ from that of [**EMS96**].

Without loss of generality, assume D itself is a ball (the general case can be reduced to this one). The $\mathrm{D}^{(3)}(\bullet)$ is almost the same as $\mathrm{D}(\bullet)$ except that in Def.6.3, we replace D by the disk with the same center but whose radius is 3 times the radius of D (this is in order to apply Besicovitch's covering lemma, see Stein's book on real analysis, Chapter 3, Problem 3).

We further assume (this will be explained later in Sec.11)

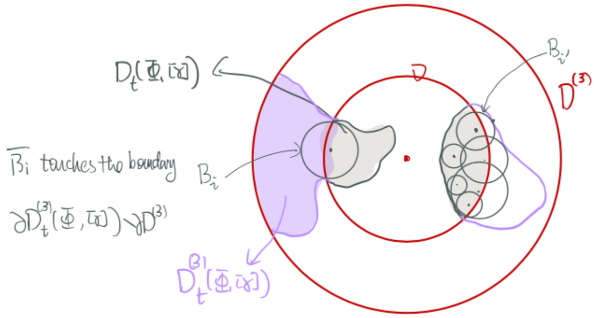
$$\text{for } t \text{ large enough,} \quad \psi_t(\mathrm{D})g_0\gamma.v_H \not\subseteq \Phi,\ \forall \gamma \in \Gamma. \tag{41}$$

Recall the definition $\mathrm{D}_t(\Phi)$ and $\mathrm{D}_t(\Phi,[\gamma])$ from Def.6.3.

For each $[\gamma]$ such that $\mathrm{D}_t(\Phi,[\gamma])$ is non-empty. Find balls $\{B_i\}_{i\in\mathscr{I}_{t,[\gamma]}} \subset \mathrm{D}_t^{(3)}(\Phi,[\gamma])$ whose **centers** cover $\mathrm{D}_t(\Phi,[\gamma])$. Here $\mathscr{I}_{t,[\gamma]}$ is some index set. We claim that we can find a covering such that for every $i \in \mathscr{I}_{t,[\gamma]}$, there exists $y \in \overline{B_i}$ such that

$$\psi_t(y)g_0\gamma.v_H \notin \Phi. \tag{42}$$

Indeed, for each $y \in \mathrm{D}_t(\Phi,[\gamma])$, take B_y to be the largest open ball centered at y. Then this collection would satisfy Equa.(42) by Equa.(41).



Let $\mathscr{I}_t := \bigsqcup_{[\gamma]\in\Gamma/\Gamma_N} \mathscr{I}_{t,[\gamma]}$. Then $\mathrm{D}_t(\Phi)$ is covered by the centers of $\{B_i\}_{i\in\mathscr{I}_t}$. By Besicovitch covering lemma, there exists a constant $C_0>0$, depending only on the dimension of $\mathfrak{k}_0$, and a subset $\mathscr{J}_t \subset \mathscr{I}_t$ such that $\{B_i\}_{i\in\mathscr{J}_t}$ is a covering of $\mathrm{D}_t(\Phi)$ of multiplicity (i.e., the maximal number of possible overlaps among B_i's) bounded by C_0.

Let $\mathscr{J}_{t,[\gamma]} := \mathscr{J}_t \cap \mathscr{I}_{t,[\gamma]}$.

Let me summarize the discussion in this subsection by the following lemma:

LEMMA 9.1. *Take t such that Equa.(41) holds. There exists a covering of $\mathrm{D}_t(\Phi)$ by open balls $(B_j)_{j\in\mathscr{J}_t}$ together with a partition of the index set $\mathscr{J}_t = \sqcup_{[\gamma]\in\Gamma/\Gamma_N}\mathscr{J}_{t,[\gamma]}$ satisfying the following:*

1. *For $j\in\mathscr{J}_{t,[\gamma]}$, $B_j \subset \mathrm{D}^{(3)}(\Phi,[\gamma])$.*
2. *For $j\in\mathscr{J}_{t,[\gamma]}$, there exists $y\in\overline{B_j}$ such that*
$$\mathbf{a}_t\exp(y)g_0\gamma.v_H \notin \Phi.$$
3. *The multiplicity of the covering is at most C_0 for a constant $C_0>0$ depending only on the dimension of $\mathfrak{k}_0$. Or more formally,*
$$\sum_{j\in\mathscr{J}_t} 1_{B_j} \le C_0.$$

10. Step 9, finish the proof under some assumption

LEMMA 10.1. *There is a constant $C_1>0$ such that for t satisfying Equa.(41)*
$$\mathrm{Leb}(\mathrm{D}_t(\mathscr{N}_\varepsilon)) \le C_1\varepsilon\,\mathrm{Leb}(\mathrm{D}).$$

Thus Lem.4.1 follows from this lemma provided Equa.(41) is verified.

PROOF. Take $C_1 := 3^{\dim\mathfrak{k}_0}C_0$.

Take $y\in\mathrm{D}_t(\mathscr{N}_\varepsilon)$, then by Lem.7.6, there exists a unique $[\gamma_y]\in\Gamma/\Gamma_N$ such that
$$\psi_t(y)g_0\gamma_y.v_H\in\Phi.$$

On the other hand, since $y\in\mathrm{D}_t(\Psi)\subset\mathrm{D}_t(\Phi)$, there exists $[\gamma]\in\Gamma/\Gamma_N$ and $j\in\mathscr{J}_{t,[\gamma]}$ such that
$$\psi_t(y)g_0\gamma.v_H\in\Phi,\ \text{ and } y\in B_j.$$

By uniqueness, $[\gamma]=[\gamma_y]$. Let $\mathrm{D}_t(\mathscr{N}_\varepsilon,[\gamma]) := \mathrm{D}_t(\mathscr{N}_\varepsilon)\cap\mathrm{D}_t(\Psi,[\gamma])$.

We have proved that for every $[\gamma]\in\Gamma/\Gamma_N$,
$$\mathrm{D}_t(\mathscr{N}_\varepsilon,[\gamma]) = \bigcup_{j\in\mathscr{J}_{t,[\gamma]}} B_j\cap\mathrm{D}_t(\mathscr{N}_\varepsilon,[\gamma]). \tag{43}$$

By comparison, it may not be true that (even if you replace Φ by the smaller Ψ)
$$\mathrm{D}_t(\Phi,[\gamma]) = \bigcup_{j\in\mathscr{J}_{t,[\gamma]}} B_j\cap\mathrm{D}_t(\Phi,[\gamma]).$$

Now everything follows from this, the linear algebra lemma Lem.5.2 and the covering argument Lem.9.1. More details:

$$\begin{aligned}
\mathrm{Leb}(\mathrm{D}_t(\mathcal{N}_\varepsilon)) &= \sum_{[\gamma]\in\Gamma/\Gamma_N} \mathrm{Leb}(\mathrm{D}_t(\mathcal{N}_\varepsilon,[\gamma])) \\
(\text{ Equa.(43) }) &\leq \sum_{[\gamma]\in\Gamma/\Gamma_N}\sum_{j\in\mathcal{J}_{t,[\gamma]}} \mathrm{Leb}(\mathrm{D}_t(\mathcal{N}_\varepsilon,[\gamma])\cap B_j) \\
&\leq \sum_{[\gamma]\in\Gamma/\Gamma_N}\sum_{j\in\mathcal{J}_{t,[\gamma]}} \mathrm{Leb}(\mathrm{D}_t(\Psi,[\gamma])\cap B_j) \\
(\text{ Lem.5.2 and Equa.(42) }) &\leq \sum_{[\gamma]\in\Gamma/\Gamma_N}\sum_{j\in\mathcal{J}_{t,[\gamma]}} \varepsilon\,\mathrm{Leb}(B_j) \\
(\text{ Lem.9.1 }) &\leq \varepsilon C_0\,\mathrm{Leb}(\mathrm{D}^{(3)}_t(\Phi)) \leq C_0\varepsilon\,\mathrm{Leb}(\mathrm{D}^{(3)}) = C_1\varepsilon\,\mathrm{Leb}(\mathrm{D}).
\end{aligned}$$

□

The promised $\mathcal{N}'_\varepsilon$ did not show up explicitly. You may take it to be $p\left((\phi\circ q)^{-1}\Phi\right)$ in light of the discussion above.

11. Step 10, linear expansion

Note that the discussion so far only uses

- the limit measure μ is unipotent-invariant;
- (C,α)-good properties.

In particular, as long as μ can be shown to be unipotent invariant, the discussion above applies equally well if you replace $\mathbf{a}_{t_n}$ by any other sequences (g_n) in G and exp(D) by any other bounded smooth curve/manifold in G equipped with a smooth measure.

Now we explain why Equa.(41) holds, for our particular choice of $\mathbf{a}_t$ and exp(D).

Recall that we may think of (the connected component of) H_0 as the image of $\mathrm{SL}_2(\mathbb{R})$ under the Adjoint representation. And K_0 may be thought of as the image of $\mathrm{SO}_2(\mathbb{R})$, $\{\mathbf{a}_t\}$ the image of $\mathbf{b}_t := \mathrm{diag}(e^t, e^{-t})$.

LEMMA 11.1. *Let V be an irreducible nontrivial representation of $\mathrm{SL}_2(\mathbb{R})$. Let Ω be a nonempty open subset of $\mathrm{SO}_2(\mathbb{R})$. Then for every constant $C>0$, there exists $T_0>0$ (depending on C, Ω, the choice of metric on V) such that for every $t>T_0$, every $v_{\neq}0\in V$*

$$\sup_{\omega\in\Omega}\|\mathbf{b}_t\omega.v\| \geq C\,\|v\|.$$

REMARK 11.2. *After the proof is given, it should be clear that $\mathrm{SL}_2(\mathbb{R})$ can be replaced by any other simple Lie group, $\mathrm{SO}_2(\mathbb{R})$ replaced by a maximal compact subgroup, $\mathbf{b}_t$ replaced by any one-parameter diagonalizable subgroup that is stable under Cartan involution associated with this maximal compact subgroup. Moreover, once V is fixed, C can be taken to be $\kappa_1 e^{\kappa_2|t|}$ for some $\kappa_1,\kappa_2>0$ and the condition $t>T_0$ can be removed.*

REMARK 11.3. *A weaker statement, with "for every $C>0$" replaced by "there exists some $c>0$" (and ignore the $t>T_0$ condition) holds in much greater generality, see* [**RS18**]. *And this condition is sufficient to conclude the limit measure supports on a unique tube (see* [**RZ16**]*).*

PROOF OF EQUA.(41) ASSUMING LEM.11.1. Assume otherwise, find some $\gamma_t\in\Gamma$ such that

$$\mathbf{a}_t\exp(\mathrm{D})g_0\gamma_t.v_H\subset\Phi$$

for t inside certain sequence tending to $+\infty$.

Decompose $V = V_1 \bigoplus V_2$ in a H_0-equivariant way such that $V_1 = V^{\mathrm{H}_0}$, the vectors fixed by H_0. Write π_i for the projection $V \to V_i$ w.r.t. this decomposition. Without loss of generality we assume $V_1 \perp V_2$ by changing the Euclidean metric. Thus for $t \in R$, $y \in \mathrm{D}$.

$$\begin{aligned}\mathbf{a}_t \exp(y)(g_0\gamma_t.v_H) &= \mathbf{a}_t \exp(y)(\pi_1(g_0\gamma_t.v_H) + \pi_2(g_0\gamma_t.v_H))\\ &= \pi_1(g_0\gamma_t.v_H) + \mathbf{a}_t \exp(y)\pi_2(g_0\gamma_t.v_H)\\ \implies \left\|\mathbf{a}_t \exp(y) g_0\gamma_t.v_H\right\| &= \left\|\pi_1(g_0\gamma_t.v_H)\right\| + \left\|\mathbf{a}_t \exp(y)\pi_2(g_0\gamma_t.v_H)\right\|.\end{aligned}$$

For the 2nd term, the above Lem.11.1 implies that for t large enough, for suitable choice of y_t,

$$\left\|\mathbf{a}_t \exp(y_t)\pi_2(g_0\gamma_t.v_H)\right\| \geq \left\|\pi_2(g_0\gamma_t.v_H)\right\|.$$

So $\mathbf{a}_t \exp(y_t)$ action does not decrease the norm of $g_0\gamma_t.v_H$. Since Φ is bounded, this implies that

$$(g_0\gamma_t.v_H) \text{ is bounded.}$$

But $\Gamma.v_H$, and hence $g_0\Gamma.v_H$ is discrete in V_H. A discrete, bounded set has no choice but being finite. After passing to a subsequence, we assume $\gamma_t = \gamma_1$ for all t (in some infinite subsequence tending to $+\infty$).

Now if $g_0\gamma_1.v_H \notin V_1$, then $\pi_2(g_0\gamma_1.v_H) \neq 0$. Take $C_2 > 0$ such that every element in Φ has norm at most C_2. Apply Lem.11.1 to $C = 1.1C_2\left\|\pi_2(g_0\gamma_1.v_H)\right\|^{-1}$, then we find y'_t, for t large enough, such that

$$\left\|\mathbf{a}_t \exp(y'_t)\pi_2(g_0\gamma_t.v_H)\right\| \geq 1.1C_2\left\|\pi_2(g_0\gamma_1.v_H)\right\|^{-1}\left\|\pi_2(g_0\gamma_t.v_H)\right\| = 1.1C_2.$$

So $\mathbf{a}_t \exp(y'_t) g_0\gamma_1.v_H$ can not live in Φ, a contradiction.

Thus $g_0\gamma_1.v_H \in V_1$, or in other words, $g_0\gamma_1.v_H$ is fixed by H_0. Recall the stabilizer of v_H in G is $N_{\mathrm{G}}(H)^{(1)}$, thus, $g_0^{-1}\mathrm{H}_0 g_0 \subset \gamma_1 N_{\mathrm{G}}(H)^{(1)}\gamma_1^{-1} \subset \gamma_1 N_{\mathrm{G}}(H)\gamma_1^{-1}$.

A Lie algebra computation shows that $\mathrm{Ad}(g_0)^{-1}\mathfrak{h}_0$ is a maximal proper Lie subalgebra. Actually, the only non-zero and non-full $\mathrm{Ad}(\mathrm{H}_0)$-stable Lie subalgebra of $\mathfrak{sl}_3$ is $\mathfrak{h}_0$. Thus $\mathrm{Ad}(g_0^{-1})\mathfrak{h}_0 = \mathrm{Ad}(\gamma_1)\mathfrak{h}$ and $g_0^{-1}\mathrm{H}_0^\circ g_0 = \gamma_1 H\gamma_1^{-1}$. In particular $g_0^{-1}\mathrm{H}_0^\circ g_0 \cap \Gamma$ is a lattice in $g_0^{-1}\mathrm{H}_0^\circ g_0$. This implies that $Q_0 \circ g_0$ is proportional to a rational quadratic form, a contradiction.

□

Proof of Lem.11.1. Decompose V w.r.t. the $\mathbf{b}_t$ action

$$V = V^- \oplus V^0 \oplus V^+$$

into contracting/fixed/expanding subspaces. Namely, this decomposition is stable under $\mathbf{b}_t$ action. Moreover $V^0 = V^{\{\mathbf{b}_t\}}$ and for some $c_1, \kappa_1 > 0$,

$$\begin{aligned}\|\mathbf{b}_t.v\| &\geq c_1 e^{\kappa_1 t}\|v\|, \forall v \in V^+;\\ \|\mathbf{b}_t.v\| &\leq c_1^{-1} e^{-\kappa_1 t}\|v\|, \forall v \in V^-.\end{aligned}$$

Let π^-, π^0, π^+ be the corresponding projections. We claim that there exists $c_2 > 0$ such that

$$\sup_{\omega\in\Omega}\left\|\pi^+(\omega.v)\right\| \geq c_2\|v\|, \forall v \in V. \tag{44}$$

Once this is done, the proof completes. It suffices to verify Equa.(44) under the assumption $\|v\| = 1$. If not true, then we can find a sequence of unit vectors (v_n) such that

$$\sup_{\omega\in\Omega}\left\|\pi^+(\omega.v_n)\right\| \to 0.$$

Let v_∞ be any limit of (v_n). Since Ω is bounded, we have

$$\pi^+(\omega.v_\infty) = 0, \forall \omega \in \Omega.$$

In other words,

$$\Omega.v_\infty \subset V^- \oplus V^0.$$

Since this is a condition defined by vanishing of some polynomials and Ω is Zariski dense in $\mathrm{SO}_2(\mathbb{R})$, we have

$$\mathrm{SO}_2(\mathbb{R}).v_\infty \subset V^- \oplus V^0.$$

Since $w_0 := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \in \mathrm{SO}_2(\mathbb{R})$ and $w_0 \mathbf{b}_t w_0^{-1} = \mathbf{b}_{-t}$, we see that $w_0 V^- = V^+$, $w_0 V^0 = V^0$ and $w_0 V^+ = V^-$. So

$$\mathrm{SO}_2(\mathbb{R}).v_\infty \subset V^0.$$

So every vector in V^0 is fixed by $\mathrm{SO}_2(\mathbb{R})$ and $\{\mathbf{b}_t\}$, which generate the full $\mathrm{SL}_2(\mathbb{R})$. This is a contradiction. □

12. Exercises

12.1. An example of equidistribution of unipotent flows. Notations

- $G = \mathrm{SL}_2(\mathbb{C})$, $\Gamma = \mathrm{SL}_2(\mathbb{Z}[i])$ and $X := G/\Gamma$;
- $U = \left\{ \mathbf{u}_s = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \,\middle|\, s \in \mathbb{R} \right\}$ and $x_0 = [g_0] \in G/\Gamma$.

Let (S_n) be a sequence of positive real numbers tending to $+\infty$ such that the following limit exists:

$$\mu := \lim_{S_n \to +\infty} \frac{1}{S_n} \int_0^{S_n} (\mathbf{u}_s)_* \delta_{[g_0]} \,\mathrm{ds}.$$

Assume the fact that such a μ belongs to $\mathrm{Prob}(X)^U$.

Recall the definitions of $\mathscr{H}$, $T(H,U)$,... (see Lec.11, Def.1.6, Def.3.1). And V_H, v_H same as in Lec.12.

EXERCISE 12.1. *Let $H \in \mathscr{H}$, $H \neq G$. Show that if $\mu(T(H,U)) > 0$, then there exists a bounded set $\Phi \subset V_H$ and a sequence $(\gamma_n) \subset \Gamma$ such that*

$$\mathbf{u}_{[0,S_n]} g_0 \gamma_n . v_H \subset \Phi.$$

EXERCISE 12.2. *Same notations as the exercise above. Conclude that there exists $\gamma \in \Gamma$ such that*

$$\mathbf{u}_{[0,+\infty)} g_0 \gamma . v_H \subset \Phi.$$

EXERCISE 12.3. *Same notations as the exercise above. Conclude that $g_0^{-1} U g_0 \subset N_G(\gamma H \gamma^{-1})^{(1)}$.*

EXERCISE 12.4. *Use exercises above to show that if $x_0 = [g_0] \notin [\mathrm{Sing}(G,U)]_\Gamma$, then*

$$\lim_{S_n \to +\infty} \frac{1}{S_n} \int_0^{S_n} (\mathbf{u}_s)_* \delta_{[g_0]} \,\mathrm{ds} = \widehat{\mathrm{m}}_{G/\Gamma}.$$

[Hint: use Lec.11, Thm.2.3 if it helps.]

EXERCISE 12.5. *Conclude that if $x_0 = [g_0] \notin [\mathrm{Sing}(G,U)]_\Gamma$, then $U.x_0$ is dense in G/Γ.*

12.2. Homogeneous sets of bounded volume. Notations

- $G := \mathrm{SL}_N(\mathbb{R})$ and $\Gamma := \mathrm{SL}_N(\mathbb{Z})$.
- Fix a right G-invariant Riemannian metric on G, which induces Riemannian metrics on G/Γ and also on immersed submanifolds. Volumes below are all induced from this.

For $C > 0$, let

$$\mathscr{A} := \{ H \leq G \mid H \text{ is a closed connected subgroup of } G, \mathrm{Vol}(H/H \cap \Gamma) < \infty. \}$$

$$\mathscr{A}_C := \{ H \leq G \mid H \text{ is a closed connected subgroup of } G, \mathrm{Vol}(H/H \cap \Gamma) < C. \}$$

DEFINITION 12.1. *Given a sequence (H_n) of closed subgroups of G, we say that (H_n)* ***converges*** *iff for every (infinite) subsequence (n_k) and $h_{n_k} \in H_{n_k}$ such that $\lim_k h_{n_k}$ exists, there exists $h'_n \in H_n$ for each n, such that*

$$\lim_k h_{n_k} = \lim_n h'_n.$$

EXERCISE 12.6. *Given a sequence (H_n) of closed subgroups of G, there exists a subsequence that converges.*

From now on we fix a convergent sequence (H_n). And assume each H_n is connected. Let

$$L := \left\{ g \in G \,\middle|\, g = \lim_n h_n,\ \exists h_n \in H_n \right\}$$

EXERCISE 12.7. *Show that L is a closed subgroup.*

EXERCISE 12.8. *There exists a subsequence n_k such that $(\mathfrak{h}_{n_k})$ (the Lie algebra of H_{n_k}) converges.*

From now on we assume $(\mathfrak{h}_n)$ converges to $\mathfrak{h}_\infty$.

EXERCISE 12.9. *Find an example of (H_n) such that $\mathfrak{h}_\infty$ is not the Lie algebra of L.*

Now we further assume that $\{H_n\} \subset \mathscr{A}_{C_0}$ for some $C_0 > 0$.

EXERCISE 12.10. *Show that under the assumption above, $\mathfrak{h}_\infty = \mathrm{Lie}(L)$.*

EXERCISE 12.11. *Show that $(H_n \cap \Gamma)$ converges and its limit is given by*

$$\Gamma_\infty := \left\{ \gamma \in \Gamma \,\middle|\, \exists n_0,\ \forall n > n_0,\ \gamma \in H_n \cap \Gamma \right\}.$$

EXERCISE 12.12. *Show that Vol_{H_n} converges to Vol_L in the weak$*$ topology.*

EXERCISE 12.13. *Show that Γ_∞ is a lattice in L. Indeed show that*

$$\mathrm{Vol}(L/\Gamma_\infty) \le \limsup \mathrm{Vol}(H_n/H_n \cap \Gamma).$$

[Hint, consider compact parts of a fundamental domain]
It is a fact that once you know Γ_∞ is a lattice in L, then it is finitely generated.

EXERCISE 12.14. *Assume the fact above. Show that there exists n_0 such that for all $n > n_0$, $\Gamma \cap H_n \supset \Gamma_\infty$.*

Continuing this way, using more inputs from the theory of algebraic groups, one can show that

THEOREM 12.2 (Dani–Margulis). *We have that*

$$\#\left\{ H \cap \Gamma \,\middle|\, H \in \mathscr{A}_{C_0} \right\} < \infty.$$

EXERCISE 12.15. *For $H \in \mathscr{A}$ and $g \in G$, show that*

$$\mathrm{Vol}(gH\Gamma/\Gamma) = \frac{\left\| \mathrm{Ad}(g).v_H \right\|}{\| v_H \|} \mathrm{Vol}(H\Gamma/\Gamma).$$

Here v_H is a vector in $\wedge^{\dim H} \mathfrak{sl}_n$ defined by $v_1 \wedge \ldots \wedge v_{\dim H}$ where $(v_1, \ldots, v_{\dim H})$ is a basis for $\mathfrak{h}$, the Lie algebra of H.

EXERCISE 12.16. *Assume the theorem above, show that $\Gamma.v_H$ is a discrete subset of $\wedge^{\dim H} \mathfrak{sl}_n$.*

12.3. Orbit counting and equidistribution. Notations

- $G = \mathrm{SL}_2(\mathbb{R})$, $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, $H = \left\{ \begin{bmatrix} x & 2y \\ y & x \end{bmatrix} \,\middle|\, x^2 - 2y^2 = 1 \right\}$;
- $V := \{\text{2-by-2 real matrices with trace } 0\}$;
- $V(\mathbb{Z}) := \{\text{2-by-2 integer matrices with trace } 0\}$
- $M_0 := \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$ and $p_0(x) := x^2 - 2$;
- for a matrix M, its characteristic polynomial is denoted by $\mathrm{char}_M(x) := \det(xI - M) = x^2 - \mathrm{Tr}(M)x + \det(M)$;
- $X_{p_0}(\mathbb{R}) := \{M \in V, \mathrm{char}_M(x) = p_0(x)\}$, $X_{p_0}(\mathbb{Z}) := \{M \in V(\mathbb{Z}), \mathrm{char}_M(x) = p_0(x)\}$;
- for a 2-by-2 matrix $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, define $\mathrm{ht}(M) := \sqrt{a^2 + b^2 + c^2 + d^2}$;
- $B_R := \{M \in X_{p_0}(\mathbb{R}) \mid \mathrm{ht}(M) \le R\}$.

EXERCISE 12.17. *Show that every pair of matrices $M_1, M_2 \in X_{p_0}(\mathbb{R})$, there exists $g \in G$ such that $gM_1g^{-1} = M_2$.*

Let G acts on $X_{p_0}(\mathbb{R})$ by $g.M := gMg^{-1}$. The above exercise shows that this action is transitive.

EXERCISE 12.18. *The stabilizer of M_0 in G is equal to H.*

EXERCISE 12.19. *$H \cap \Gamma$ is a lattice in H.*

EXERCISE 12.20. *Show that the action of Γ on $X_{p_0}(\mathbb{Z})$ is transitive.*

[Hint: $\mathbb{Z}[\sqrt{2}]$ is a PID]
Further notations

- $\mathrm{m}_{G/H}$ is a G-invariant locally finite measure on G/H;
- similarly, m_G and m_H denote Haar measures on G and H respectively.

Note that G and H are unimodular: left Haar measures are the same as right Haar measures.

DEFINITION 12.3. *We say that a triple $(\mathrm{m}_G, \mathrm{m}_H, \mathrm{m}_{G/H})$ is compatible iff for every compactly supported function $f \in C_c(G)$, we have*

$$\int_{G/H} \int_H f(gh)\mathrm{m}_H(h)\mathrm{m}_{G/H}([g]) = \int_G f(g)\mathrm{m}_G([g]). \tag{45}$$

EXERCISE 12.21. *Show that for every triple of Haar measures $(\mathrm{m}_G, \mathrm{m}_H, \mathrm{m}_{G/H})$, there exists a constant $c > 0$ such that for every $f \in C_c(G)$,*

$$\int_{G/H} \int_H f(gh)\mathrm{m}_H(h)\mathrm{m}_{G/H}([g]) = c \cdot \int_G f(g)\mathrm{m}_G([g]).$$

From now on we fix the unique triple $(\mathrm{m}_G, \mathrm{m}_H, \mathrm{m}_{G/H})$ satisfying

1. $(\mathrm{m}_G, \delta_\Gamma, \widehat{\mathrm{m}}_{G/\Gamma})$ and $(\mathrm{m}_H, \delta_{H\cap\Gamma}, \widehat{\mathrm{m}}_{H/H\cap\Gamma})$ are compatible. Here δ_Γ (resp. $\delta_{H\cap\Gamma}$) denotes the counting measure on Γ (resp. $H \cap \Gamma$).
2. $(\mathrm{m}_G, \mathrm{m}_H, \mathrm{m}_{G/H})$ is compatible.

Its existence is guaranteed by the Exer.12.21 above.

EXERCISE 12.22. *Find the asymptotics of*

$$\mathrm{m}_{G/H}(B_R) := \mathrm{m}_{G/H}\left(\{[g] \in G/H \mid \mathrm{ht}(g.M_0) \le R\}\right).$$

DEFINITION 12.4. *Define $\varphi_R : G/\Gamma \to \mathbb{R}$ by*

$$\varphi_R([g]) := \#\left(g\Gamma.M_0 \cap B_R\right).$$

We say that $\frac{1}{\mathrm{m}_{G/H}(B_R)}\varphi_R$ *converges to* 1 *weakly iff for all* $\psi \in C_c(G/\Gamma)$,

$$\lim_{R\to+\infty} \frac{1}{\mathrm{m}_{G/H}(B_R)} \int_{G/\Gamma} \varphi_R([g])\psi([g])\widehat{\mathrm{m}}_{G/\Gamma}([g]) = \int \psi([g])\widehat{\mathrm{m}}_{G/\Gamma}([g]). \tag{46}$$

EXERCISE 12.23. *Show that if* $\frac{1}{\mathrm{m}_{G/H}(B_R)}\varphi_R$ *converges to* 1 *weakly then for every* $[g] \in G/\Gamma$,

$$\lim_{R\to+\infty} \frac{1}{\mathrm{m}_{G/H}(B_R)}\varphi_R([g]) = 1.$$

In particular, in light of Exer.12.20,

$$\# X_{p_0}(\mathbb{Z}) \cap B_R \sim \mathrm{m}_{G/H}(B_R).$$

[Hint: use Exer.12.22].

EXERCISE 12.24. *Show that the left hand side of Equa.(46) (excluding the limit) is equal to*

$$\frac{1}{\mathrm{m}_{G/H}(B_R)} \int_{\{g.M_0 \in B_R\}} \left(\int \psi(x)\, g_* \widehat{\mathrm{m}}_{H\Gamma/\Gamma}(x) \right) \mathrm{m}_{G/H}([g])$$

EXERCISE 12.25. *Use "linearization technique" to show that for every sequence* (g_n) *such that* $([g_n])$ *diverges in* G/H, *we have*

$$\lim_{n\to+\infty} (g_n)_* \widehat{\mathrm{m}}_{H\Gamma/\Gamma} = \widehat{\mathrm{m}}_{G/\Gamma}.$$

EXERCISE 12.26. *Use Exer.12.25 to conclude that* $\frac{1}{\mathrm{m}_{G/H}(B_R)}\varphi_R$ *converges to* 1 *weakly.*

Quantitative Oppenheim I, reducing to dynamics

Back to the Top.
Main reference: [EMM98, Section 3].
Notations

- Let $Q_0(x_1, x_2, x_3, x_4) := 2x_1x_4 + x_2^2 + x_3^2$, a real quadratic form of signature $(3,1)$ on $\mathbb{R}^4$.
- Let $(\mathbf{e}_1, ..., \mathbf{e}_4)$ be the standard basis of $\mathbb{R}^4$; and for a vector $\mathbf{v}$, define its coefficients by $v = \sum (v)_i \mathbf{e}_i$ and we also write $v = ((v)_1, ..., (v)_4)$.
- Let $(\mathbf{f}_1, ..., \mathbf{f}_4)$ be another ONB(=orthogonal normal basis) defined by $\mathbf{f}_2 = \mathbf{e}_2, \mathbf{f}_3 = \mathbf{e}_3$ and $\mathbf{f}_1 = \frac{\mathbf{e}_1+\mathbf{e}_4}{\sqrt{2}}, \mathbf{f}_4 = \frac{\mathbf{e}_1-\mathbf{e}_4}{\sqrt{2}}$. If $v = \sum a_i \mathbf{f}_i$, we also write $v = (a_1, ..., a_4)_{\mathbf{f}}$.
- One can verify that $Q_0((x_1, ..., x_4)_{\mathbf{f}}) = x_1^2 + x_2^2 + x_3^2 - x_4^2$.
- $\mathrm{K} := \mathrm{SO}_{Q_0}(\mathbb{R}) \cap \mathrm{SO}_4(\mathbb{R})$.
- $\mathbf{a}_t := \mathrm{diag}(e^{-t}, 1, 1, e^t)$, contained in $\mathrm{SO}_{Q_0}(\mathbb{R})$.

1. Detect points by probabilistic methods

Assume $Q_0 \circ g_0$ is irrational. Define

$$V_{(a,b)}(\mathbb{Z}) := \left\{\mathbf{v} \in g_0.\mathbb{Z}^4 \;\middle|\; Q_0(\mathbf{v}) \in (a,b)\right\},$$

$$N_T := \# V_{a,b}(\mathbb{Z}, T),\ V_{a,b}(\mathbb{Z}, T) := \left\{\mathbf{v} \in V_{(a,b)}(\mathbb{Z}) \;\middle|\; \|\mathbf{v}\| \leq T\right\}.$$

Consider the function

$$1_{\square}(x,y) := 1_{(1,2]}(x) \cdot 1_{(a,b)}(y).$$

Hence

$$N_{2T} - N_T = \sum_{\mathbf{v} \in g_0.\mathbb{Z}^4} 1_{\square}\left(\frac{\|\mathbf{v}\|}{T}, Q_0(\mathbf{v})\right).$$

Find a compactly supported continuous function h approximating $1_{\square}$ from above. Then one can find some (non-negative) $f \in C_c(\mathbb{R}_{>0} \times \mathbb{R}^3)$ such that

$$h(x,y) = \frac{1}{x^2} \int f(x, w_2, w_3, y')\, |\mathrm{dw}_2 \wedge \mathrm{dw}_3| \tag{47}$$

where $y' := \frac{y - w_2^2 - w_3^2}{2x}$.

1.1. A coarse upper bound.

By abbreviating $V_{a,b}(\mathbb{Z}, 2T - T) := V_{a,b}(\mathbb{Z}, 2T) \setminus V_{a,b}(\mathbb{Z}, T)$, we have

$$\begin{aligned} N_{2T} - N_T &\leq \sum_{\mathbf{v} \in g_0.\mathbb{Z}^4, \mathbf{v} \in V_{a,b}(\mathbb{Z}, 2T-T)} h\left(\frac{\|\mathbf{v}\|}{T}, Q_0(\mathbf{v})\right) \\ &= \sum_{\mathbf{v} \in g_0.\mathbb{Z}^4, \mathbf{v} \in V_{a,b}(\mathbb{Z}, 2T-T)} \frac{T^2}{\|\mathbf{v}\|^2} \int f\left(\frac{\|\mathbf{v}\|}{T}, w_2, w_3, \frac{Q_0(\mathbf{v}) - w_2^2 - w_3^2}{2\,\|\mathbf{v}\|\, T^{-1}}\right) |\mathrm{dw}_2 \wedge \mathrm{dw}_3| \end{aligned} \tag{48}$$

Each summand here is either 0 or ≥ 1 since we are keeping the index $\mathbf{v} \in V_{a,b}(\mathbb{Z}, 2T - T)$.

Now we need the following lemma, to be proved later (see Lem.2.10 where this is proved).

LEMMA 1.1. *Given $f \in C_c(\mathbb{R}_{>0} \times \mathbb{R}^3)$ and $\varepsilon \in (0,1)$, there exists $T_0 = T_0(f,\varepsilon) > 0$ such that for every $T > T_0$, for every $\mathbf{v} \in \mathbb{R}^4$ we have*

$$\left| \frac{1}{2C_4} T^2 \int f(\mathbf{a}_{\ln T} k.\mathbf{v}) \widehat{\mathrm{m}}_{\mathrm{K}}(k) - \frac{T^2}{\|\mathbf{v}\|^2} \int f\left(\frac{\|\mathbf{v}\|}{T}, w_2, w_3, w_4\right) |\mathrm{dw}_2 \wedge \mathrm{dw}_3| \right| < \varepsilon$$

where

$$w_4 := \frac{Q_0(\mathbf{v}) - w_2^2 - w_3^2}{2\,\|\mathbf{v}\|\, T^{-1}}$$

is a function in (w_2, w_3), for every fixed $\mathbf{v}$ and T.

Apply Lem.1.1 with some $\varepsilon < 0.5$, then for T sufficiently large, each $\mathbf{v} \in V_{a,b}(\mathbb{Z}, 2T - T)$, either

$$\frac{T^2}{\|\mathbf{v}\|^2} \int f\left(\frac{\|\mathbf{v}\|}{T}, w_2, w_3, \frac{Q_0(\mathbf{v}) - w_2^2 - w_3^2}{2\,\|\mathbf{v}\|\, T^{-1}}\right) = \frac{1}{2C_4} T^2 \int 2 f(\mathbf{a}_t k.\mathbf{v}) \widehat{\mathrm{m}}_{\mathrm{K}}(k) = 0$$

or ≥ 0.5 .

Therefore

$$\begin{aligned} N_{2T} - N_T &\leq \sum_{\mathbf{v} \in g_0.\mathbb{Z}^4, \mathbf{v} \in V_{a,b}(\mathbb{Z}, 2T-T)} 2 \cdot \frac{1}{2C_4} T^2 \int f(\mathbf{a}_t k.\mathbf{v}) \widehat{\mathrm{m}}_{\mathrm{K}}(k) \\ &\leq \sum_{\mathbf{v} \in g_0.\mathbb{Z}^4} 2 \frac{1}{2C_4} T^2 \int f(\mathbf{a}_t k.\mathbf{v}) \widehat{\mathrm{m}}_{\mathrm{K}}(k) = 2 \frac{1}{2C_4} T^2 \int \widetilde{f}(\mathbf{a}_t k g_0.\mathbb{Z}^4) \widehat{\mathrm{m}}_{\mathrm{K}}(k). \end{aligned} \tag{49}$$

where

$$\widetilde{f} : \mathrm{X}_4 \to \mathbb{R} \text{ defined by } \widetilde{f}(\Lambda) := \sum_{\mathbf{v} \in \Lambda} f(\mathbf{v}).$$

If $\widetilde{f}$ were a bounded function, then immediately we see that for some constant $C = C(f) > 0$,

$$N_{2T} - N_T \leq T^2 C \implies N_{2^n T_0} \leq T_0^2 C(1 + 4^1 + \dots + 4^{n-1}) + N_{T_0} = \frac{1-4^n}{1-4} T_0^2 C + N_{T_0} \leq (2^n T_0)^2 C + N_{T_0}.$$

This shows that for T large,

$$N_T \leq 2CT^2.$$

Unfortunately our $\widetilde{f}$ is not bounded. Nevertheless we still have

THEOREM 1.2. *There exists a constant $C = C(f) > 0$ such that*

$$\int \widetilde{f}(\mathbf{a}_t k g_0.\mathbb{Z}^4) \widehat{\mathrm{m}}_{\mathrm{K}}(k) \leq C$$

for all $t > 0$.

By arguments outlined above and Thm.1.2 we get

THEOREM 1.3. *There exists a constant $C > 0$ such that $N_T \leq CT^2$ for T sufficiently large.*

1.2. The exact upper/lower bound. Equipped with Thm.1.3, let us revisit Equa.(48):

$$\begin{aligned} \frac{N_{2T} - N_T}{T^2} &\leq T^{-2} \sum_{\mathbf{v} \in g_0.\mathbb{Z}^4, \mathbf{v} \in V_{a,b}(\mathbb{Z}, 2T)} h\left(\frac{\|\mathbf{v}\|}{T}, Q_0(\mathbf{v})\right) \\ &= T^{-2} \sum_{\mathbf{v} \in g_0.\mathbb{Z}^4, \mathbf{v} \in V_{a,b}(\mathbb{Z}, 2T)} \frac{T^2}{\|\mathbf{v}\|^2} \int f\left(\frac{\|\mathbf{v}\|}{T}, w_2, w_3, Q_0(\mathbf{v})\right) |\mathrm{dw}_2 \wedge \mathrm{dw}_3|. \end{aligned} \tag{50}$$

Fix an $\varepsilon > 0$, the range of T such that Lem.1.1 is not applicable is bounded. Thus

$$\frac{N_{2T} - N_T}{T^2} \leq T^{-2} \sum_{\mathbf{v} \in g_0.\mathbb{Z}^4, \mathbf{v} \in V_{a,b}(\mathbb{Z}, 2T)} \left(\frac{1}{2C_4} T^2 \int f(\mathbf{a}_t k.\mathbf{v}) \widehat{\mathrm{m}}_{\mathrm{K}}(k) + O(\varepsilon) \right) + O_\varepsilon(T^{-2}). \tag{51}$$

By Thm.1.3, the number of indices is bounded by $C(2T)^2$, hence

$$\begin{aligned}
\frac{N_{2T}-N_T}{T^2} &\le T^{-2}\left(\sum_{\mathbf{v}\in g_0.\mathbb{Z}^4,\mathbf{v}\in V_{a,b}(\mathbb{Z},2T)} \frac{1}{2C_4}T^2\int f(\mathbf{a}_t k.\mathbf{v})\widehat{\mathrm{m}}_{\mathrm{K}}(k)\right)+O_\varepsilon(T^{-2})+O(\varepsilon)\\
&\le \frac{1}{2C_4}\left(\sum_{\mathbf{v}\in g_0.\mathbb{Z}^4}\int f(\mathbf{a}_t k.\mathbf{v})\widehat{\mathrm{m}}_{\mathrm{K}}(k)\right)+O_\varepsilon(T^{-2})+O(\varepsilon)\\
&= \frac{1}{2C_4}\int \widetilde{f}(\mathbf{a}_t k g_0\mathbb{Z}^4)\widehat{\mathrm{m}}_{\mathrm{K}}(k)+O_\varepsilon(T^{-2})+O(\varepsilon)
\end{aligned} \tag{52}$$

Hence (let $\varepsilon\to 0$ after taking the limit $\lim_T$)

$$\limsup_{T\to+\infty}\frac{N_{2T}-N_T}{T^2}\le \lim_{t\to+\infty}\int\frac{1}{2C_4}\widetilde{f}(\mathbf{a}_t k g_0\mathbb{Z}^4)\widehat{\mathrm{m}}_{\mathrm{K}}(k).$$

That the RHS is a true limit is justified below.
The exact lower bound is proved similarly.

THEOREM 1.4. *Assume $Q_0\circ g_0$ is not rational, then for every $f\in C_c(\mathbb{R}^4)$,*

$$\lim_{t\to+\infty}\int \widetilde{f}(\mathbf{a}_t k g_0\mathbb{Z}^4)\widehat{\mathrm{m}}_{\mathrm{K}}(k)=\int_{\mathrm{X}_4}\widetilde{f}(x)\widehat{\mathrm{m}}_{\mathrm{X}_4}(x)=C_6\int_{\mathbb{R}^4} f(\mathbf{v})\,\mathrm{dv}$$

where $C_6>0$ depending only on the dimension.

Let us evaluate $\int_{\mathbb{R}^4} f(\mathbf{v})\,\mathrm{dv}$ for our f. By change of variables $y' =: \frac{y-w_2^2-w_3^2}{2x}$,

$$\int f(\mathbf{v})\,\mathrm{dv}=\int f(x,w_2,w_3,y')\,\mathrm{dx}\,\mathrm{dy}'\,\mathrm{dw}_2\,\mathrm{dw}_3=\int\frac{1}{2x}f(x,w_2,w_3,\frac{y-w_2^2-w_3^2}{2x})\,\mathrm{dx}\,\mathrm{dy}\,\mathrm{dw}_2\,\mathrm{dw}_3\,.$$

where we have used

$$\mathrm{dy}'=\frac{\mathrm{dy}-2w_2\,\mathrm{dw}_2-2w_3\,\mathrm{dw}_3}{2x}-\frac{\mathrm{dx}}{2x^2}(y-w_2^2-w_3^2).$$

Recall Equa.(47), we have

$$\int f(\mathbf{v})\,\mathrm{dv}=\int\frac{x}{2}h(x,y)\,\mathrm{dx}\,\mathrm{dy}.$$

As $h(x,y)$ approximates $1_\square$ we get

$$\int\frac{x}{2}h(x,y)\,\mathrm{dx}\,\mathrm{dy}\to\int_{y=a}^{b}\int_{x=1}^{2}\frac{x}{2}\,\mathrm{dx}\,\mathrm{dy}=\frac{2^2-1}{4}(b-a).$$

Thus, by collecting the constants $C_7:=\frac{1}{2C_4}C_6\frac{2^2-1}{4}$,

$$\lim_{T\to+\infty}\frac{N_{2T}-N_T}{T^2}=C_7(b-a).$$

Now a geometric series argument shows that

COROLLARY 1.5.

$$\lim_{T\to+\infty}\frac{N_T}{T^2}=\frac{1}{3}C_7(b-a).$$

2. Proof of the Lemma

2.1. Nontrivial contribution to the integral.

DEFINITION 2.1. *For $(x, y, z) \in \mathbb{R}^3$ with $x \neq 0$ and $a \in \mathbb{R}$, we let*

$$\phi_a(x, y, z) := \frac{a - y^2 - z^2}{2x},$$

in other words, $\phi_a(x, y, z)$ is the unique real number such that

$$Q_0\left(x, y, z, \phi_a(x, y, z)\right) = a.$$

DEFINITION 2.2. *Given $f \in C_c(\mathbb{R}_{>0} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R})$, we fix $C_1 = C_1(f) > 1$ such that*

$$\mathrm{Supp}(f) \subset (C_1^{-1}, C_1) \times (-C_1, C_1)^3.$$

Also fix $C_2 > |a_0|, |b_0|$.

The following two directly follow from the definition.

LEMMA 2.3. *Let $\mathbf{v}_{\neq 0} \in \mathbb{R}^4$ and $T > 1$. Let $(w_2, w_3) \in \mathbb{R}^2$ be such that*

$$f\left(\frac{\|\mathbf{v}\|}{T}, w_2, w_3, \phi_{Q_0(\mathbf{v})}(\frac{\|\mathbf{v}\|}{T}, w_2, w_3)\right) \neq 0,$$

then

1. $C_1^{-1}T \leq \|\mathbf{v}\| \leq C_1 T$ *and* $|w_2|, |w_3| \leq C_1$*;*
2. $\left|\phi_{Q_0(\mathbf{v})}(\frac{\|\mathbf{v}\|}{T}, w_2, w_3)\right| \leq C_1.$*;*
3. $|Q_0(\mathbf{v})| \leq 4C_1^2$.

For a vector $\mathbf{w}$, $\mathbf{w}(i) \in \mathbb{R}$ is defined by $\mathbf{w} = \sum \mathbf{w}(i)\mathbf{e}_i$.

LEMMA 2.4. *Let $\mathbf{v}_{\neq 0} \in \mathbb{R}^4$ and $T > 1$. Let $\mathbf{w} \in K.\mathbf{v}$. If $f(\mathbf{a}_{\ln T}.\mathbf{w}) \neq 0$, then*

1. $C_1^{-1}T \leq \mathbf{w}(1) \leq C_1 T$, $|\mathbf{w}(2)|, |\mathbf{w}(3)| \leq C_1$ *and* $|\mathbf{w}(4)| \leq C_1 T^{-1}$*;*
2. $\|\mathbf{v}\| \leq 2C_1 T$*;*
3. $\|\mathbf{v}\| \geq C_1^{-1} T$*;*
4. $|Q_0(\mathbf{v})| \leq 4C_1^2$.

PROOF. For item 3, $\|\mathbf{v}\| = \|\mathbf{w}\| \geq \mathbf{w}(1) \geq C_1^{-1}T$.
For item 4, $Q_0(\mathbf{v}) = Q_0(\mathbf{w}) = \mathbf{w}(1)\mathbf{w}(4) + \mathbf{w}(2)^2 + \mathbf{w}(3)^2 \leq 2C_1^2 + C_1^2 + C_1^2 = 4C_1^2$. □

2.2. Representative in a K-orbit. By working with the basis $\mathbf{f}$, one sees that for every $\mathbf{v} \in \mathbb{R}^4$, there exists $k_{\mathbf{v}} \in K$ such that

$$k_{\mathbf{v}}.\mathbf{v} = (u_1, 0, 0, u_4)_{\mathbf{f}} \text{ for some } u_1, u_4 \geq 0.$$

Indeed, if we set

$$r_1(\mathbf{v}) := \frac{\|\mathbf{v}\| + Q_0(\mathbf{v})}{2},\ r_2(\mathbf{v}) := \frac{\|\mathbf{v}\| - Q_0(\mathbf{v})}{2}$$

or equivalently,

$$r_1(\mathbf{v}) := \mathbf{v}_{\mathbf{f}}(1)^2 + \mathbf{v}_{\mathbf{f}}(2)^2 + \mathbf{v}_{\mathbf{f}}(3)^2,\ r_2(\mathbf{v}) := \mathbf{v}_{\mathbf{f}}(4)^2$$

where we assume $\mathbf{v} = (\mathbf{v}_{\mathbf{f}}(1), ..., \mathbf{v}_{\mathbf{f}}(4))_{\mathbf{f}}$. Then there exists $k \in K$ such that

$$k.\mathbf{v} = (\sqrt{r_1}, 0, 0, \sqrt{r_2})_{\mathbf{f}} =: \mathbf{v}^*.$$

To summarize the discussion in the basis $\mathbf{e}$:

LEMMA 2.5. *For every $\mathbf{v} \in R^4$ there exists a unique $\mathbf{v}^* \in \mathbb{R}^4$ satisfying*

1. $Q_0(\mathbf{v}^*) = Q_0(\mathbf{v})$*;*
2. $\|\mathbf{v}^*\| = \|\mathbf{v}\|$*;*

3. $\mathbf{v}^*(1) \geq |\mathbf{v}^*(4)|$ *and* $\mathbf{v}^*(2) = \mathbf{v}^*(3) = 0$.

Also $\mathbf{v}^* \in K.\mathbf{v}$.

What we are going to need is the following slightly perturbed version.

LEMMA 2.6. *Let* $\mathbf{v} \in \mathbb{R}^4$ *and* $(w_2, w_3) \in \mathbb{R}$ *satisfying* $|w_2|, |w_3| \leq C_1$. *Assume* $\|\mathbf{v}\|^2 \geq Q_0(\mathbf{v}) + 4C_1^2$. *Then there exists a unique* $\mathbf{v}^*(w_2, w_3) = \mathbf{w} \in \mathbb{R}^4$ *such that*

1. $Q_0(\mathbf{w}) = Q_0(\mathbf{v})$;
2. $\|\mathbf{w}\| = \|\mathbf{v}\|$;
3. $\mathbf{w}(1) \geq |\mathbf{w}(4)|$ *and* $\mathbf{w}(2) = w_2, \mathbf{w}(3) = w_3$.

Also $\mathbf{w} \in K.\mathbf{v}$.

SKETCH OF PROOF. Indeed under the assumption above

$$\left|\frac{Q_0(\mathbf{v}) - w_2^2 - w_3^2}{2}\right| \leq \frac{1}{2}(Q_0(\mathbf{v}) + 2C_1^2)$$

and

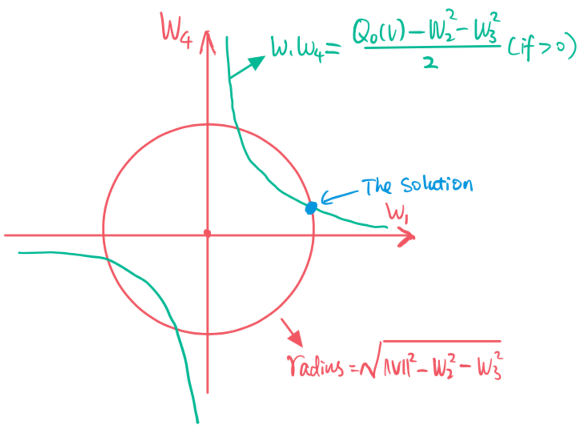
$$\|\mathbf{v}\|^2 - w_2^2 - w_3^2 \geq Q_0(\mathbf{v}) + 4C_1^2 - C_1^2 - C_1^2 = Q_0(\mathbf{v}) + 2C_1^2.$$

Hence the equation

$$\begin{cases} xy = \frac{Q_0(\mathbf{v}) - w_2^2 - w_3^2}{2} \\ x^2 + y^2 = \|\mathbf{v}\|^2 - w_2^2 - w_3^2 \end{cases}$$

admits a unique solution with $x \geq |y|$.

Here is a picture ($x = w_1$, $y = w_4$)



□

2.3. Approximates I, the points.

LEMMA 2.7. *Assumption as in Lem.2.3. Further assume* $T \geq 8C_1^3$ *and* $T^2 \geq 16C_1^4$. *Define* $\mathbf{w} = \mathbf{v}^*(w_2, w_3)$ *as in Lem.2.6. Then for* $C_3 = 46C_1^7$,

$$\operatorname{dist}_\infty\left(\left(\frac{\|\mathbf{v}\|}{T}, w_2, w_3, \phi_{Q_0(\mathbf{v})}(\frac{\|\mathbf{v}\|}{T}, w_2, w_3)\right), \mathbf{a}_{\ln T}.\mathbf{w}\right) \leq C_3 T^{-2}.$$

Note that $T \geq 8C_1^3 \implies \|\mathbf{v}\| \geq 4C_1^2 + 4C_1^2 \geq Q_0(\mathbf{v}) + 4C_1^2$ by Lem.2.3. Thus Lem.2.6 is applicable.

PROOF. First we have

$$|\mathbf{w}(4)|^2 \le |\mathbf{w}(1)|\,|\mathbf{w}(4)| = \left|Q_0(\mathbf{v}) - w_2^2 - w_3^2\right| \le 4C_1^2 + 2C_1^2 = 6C_1^2.$$

Hence the difference of the first coordinate:

$$\left|\|\mathbf{v}\|^2 - \mathbf{w}(1)^2\right| = \mathbf{w}(2)^2 + \mathbf{w}(3)^2 + \mathbf{w}(4)^2 \le 8C_1^2$$

$$\implies \left|T^{-1}\|\mathbf{v}\| - T^{-1}\mathbf{w}(1)\right| \le T^{-1}\frac{8C_1^2}{\|\mathbf{v}\| + \mathbf{w}(1)} \le T^{-1}\frac{8C_1^2}{\|\mathbf{v}\|} \le 8C_1^3 T^{-2} \le C_3 T^{-2}.$$

From here we also see that

$$|\mathbf{w}(1)| \ge \|\mathbf{v}\| - 8C_1^3 T^{-1} \ge \frac{1}{2}C_1^{-1}T + (\frac{1}{2}C_1^{-1}T - 8C_1^3 T^{-1}) \ge \frac{1}{2}C_1^{-1}T.$$

Here we are using the assumption $T^2 \ge 16C_1^4 \implies \frac{1}{2}C_1^{-1}T - 8C_1^3T^{-1} \ge 0$.

Now the difference of the last coordinate (note that $w_2 = \mathbf{w}(2)$ and $w_3 = \mathbf{w}(3)$ from Lem.2.6)

$$\begin{aligned}
&\left|\frac{Q_0(\mathbf{v}) - \mathbf{w}(2)^2 - \mathbf{w}(3)^2}{2\|\mathbf{v}\|\,T^{-1}} - \frac{Q_0(\mathbf{v}) - \mathbf{w}(2)^2 - \mathbf{w}(3)^2}{2\mathbf{w}(1)\,T^{-1}}\right| \\
\le& \frac{1}{2}(6C_1^2)T\left|\frac{1}{\|\mathbf{v}\|} - \frac{1}{\mathbf{w}(1)}\right| = \frac{(6C_1^2)T}{2}\frac{|\|\mathbf{v}\| - \mathbf{w}(1)|}{\|\mathbf{v}\|\,\mathbf{w}(1)} \\
\le& \frac{(6C_1^2)T}{2}\frac{8C_1^3T^{-1}}{1/2C_1^{-2}T^2} = 48C_1^7T^{-2} \le C_3T^{-2}.
\end{aligned}$$

□

LEMMA 2.8. *Assumption as in Lem.2.4. Define $w_2 := \mathbf{w}(2)$ and $w_3 := \mathbf{w}(3)$. Then for $C_3 = 48C_1^7$,*

$$\operatorname{dist}_\infty\left(\left(\frac{\|\mathbf{v}\|}{T}, w_2, w_3, \phi_{Q_0(\mathbf{v})}(\frac{\|\mathbf{v}\|}{T}, w_2, w_3)\right), \mathbf{a}_{\ln T}.\mathbf{w}\right) \le C_3T^{-2}.$$

PROOF. The difference of the first coordinate:

$$\left|\|\mathbf{v}\|^2 - \mathbf{w}(1)^2\right| = \mathbf{w}(2)^2 + \mathbf{w}(3)^2 + \mathbf{w}(4)^2 \le 3C_1^2$$

$$\implies \left|T^{-1}\|\mathbf{v}\| - T^{-1}\mathbf{w}(1)\right| \le T^{-1}\frac{3C_1^2}{\mathbf{w}(1)} \le 3C_1^3T^{-2} \le C_3T^{-2}.$$

And the difference of the last coordinate

$$\begin{aligned}
&\left|\frac{Q_0(\mathbf{v}) - \mathbf{w}(2)^2 - \mathbf{w}(3)^2}{2\|\mathbf{v}\|\,T^{-1}} - \frac{Q_0(\mathbf{v}) - \mathbf{w}(2)^2 - \mathbf{w}(3)^2}{2\mathbf{w}(1)\,T^{-1}}\right| \\
\le& \frac{T}{2}(6C_1^2)\left|\frac{1}{\|\mathbf{v}\|} - \frac{1}{\mathbf{w}(1)}\right| = \frac{(6C_1^2)T}{2}\frac{|\|\mathbf{v}\| - \mathbf{w}(1)|}{\|\mathbf{v}\|\,\mathbf{w}(1)} \\
\le& \frac{(6C_1^2)T}{2}\frac{3C_1^3T^{-1}}{C_1^{-2}T^2} = 9C_1^7T^{-2} \le C_3T^{-2}.
\end{aligned}$$

□

2.4. Approximates II, the measures. Let $S(r)$ be the sphere of radius r in $\mathbb{R}^3$ centered at the origin. Let $\widehat{\mathrm{m}}_{S(r)}$ be the normalized (to be a probability measure) volume measure on $S(r)$.

Assume $r_1(\mathbf{v}) \ge 2C_1^2$. For $(x_2, x_3) \in \mathbb{R}^2$ with $|x_2|, |x_3| \le C_1$, there exists a unique $x_1 > 0$ such that

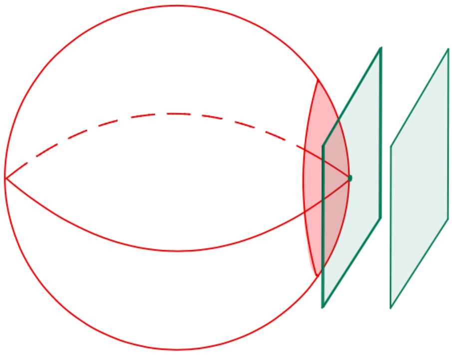
$$x_1^2 + x_2^2 + x_3^2 = r_1(\mathbf{v}).$$

Let $\mathrm{D}(C_1)$ be the image of $\{(x_2, x_3),\ |x_2|, |x_3| \le C_1\}$ in $S(\sqrt{r_1})$ thus defined. And identify $|\mathrm{d}x_2 \wedge \mathrm{d}x_3|\,|_{|x_i| \le C_1}$ as a measure on $\mathrm{D}(C_1) \subset S(\sqrt{r_1})$ by this. Equivalently one may first restrict the differential form $\mathrm{d}x_2 \wedge \mathrm{d}x_3$ to $\mathrm{D}(C_1)$ and then take the measure associated with it.

LEMMA 2.9. *Let $\mathbf{v}\in\mathbb{R}^4$ be satisfying $\frac{\|\mathbf{v}\|^2}{2}\geq|Q_0(\mathbf{v})|$ and $\|\mathbf{v}\|^2\geq 16C_1^2$.*

$$\left\|\|\mathbf{v}\|^2\,\widehat{\mathrm{m}}_{S(\sqrt{r_1})}-2C_4\,|\mathrm{d}w_2\wedge\mathrm{d}w_3|\right\|_{D(C_1)}\leq\frac{1}{\|\mathbf{v}\|^2}(C_5Q_0(\mathbf{v})+C_5)$$

where $C_4>0$ is a constant depending only on the dimension and $C_5>1$ depends on C_1. See Equa.(53), (54) below.



Note that our assumption implies that $r_1(\mathbf{v})=1/2(\|\mathbf{v}\|^2+Q_0(\mathbf{v}))\geq 4C_1^2$. Thus the paragraph above the proposition makes sense.

PROOF. First let us write $\widehat{\mathrm{m}}_{S(\sqrt{r_1})}$ in terms of differential forms. By taking the differential

$$x_1^2+x_2^2+x_3^2=r^2\implies 2x_1\,\mathrm{d}x_1+2x_2\,\mathrm{d}x_2+2x_3\,\mathrm{d}x_3=2r\,\mathrm{d}r.$$

Thus

$$\mathrm{d}x_1\wedge\mathrm{d}x_2\wedge\mathrm{d}x_3=\frac{r\,\mathrm{d}x_2\wedge\mathrm{d}x_3}{x_1}\wedge\mathrm{d}r$$

So up to constant (depending possibly on r), the spherical measure can be induced from $\frac{r\,\mathrm{d}x_2\wedge\mathrm{d}x_3}{x_1}$. To make it have total mass independent of r, we consider

$$\mathrm{d}x_1\wedge\mathrm{d}x_2\wedge\mathrm{d}x_3=\frac{\mathrm{d}x_2\wedge\mathrm{d}x_3}{rx_1}\wedge r^2\,\mathrm{d}r.$$

Since the volume of ball of radius R is some constant multiple of $R^3/3=\int_0^R r^2\,\mathrm{d}r$, there exists some constant $C_4>0$ depending only on the dimension such that

$$\widehat{\mathrm{m}}_{S(\sqrt{r_1})}=C_4\frac{\mathrm{d}x_2\wedge\mathrm{d}x_3}{\sqrt{r_1}x_1}.\tag{53}$$

By assumption,

$$2r_1=\|\mathbf{v}\|^2+Q_0(\mathbf{v})\geq\|\mathbf{v}\|^2-|Q_0(\mathbf{v})|\geq\frac{1}{2}\|\mathbf{v}\|^2\implies r_1\geq 4C_1^2.$$

Thus for $(x_1,x_2,x_3)\in S(\sqrt{r_1})$,

$$2\sqrt{r_1}x_1=2\sqrt{r_1}\sqrt{r_1-x_2^2-x_3^2}\geq\|\mathbf{v}\|\sqrt{r_1-2C_1^2}\geq\|\mathbf{v}\|\sqrt{\frac{1}{2}r_1}\geq\frac{\|\mathbf{v}\|^2}{8}.$$

On the other hand

$$\left|\|\mathbf{v}\|^2-2r_1\right|=|Q_0(\mathbf{v})|$$

and

$$\left|2r_1-2\sqrt{r_1}x_1\right|=2\sqrt{r_1}\left|\frac{r_1-(r_1-x_2^2-x_3^2)}{\sqrt{r_1}+\sqrt{r_1-x_2^2-x_3^2}}\right|\leq 2\left|x_2^2+x_3^2\right|\leq 4C_1^2.$$

Therefore, when restricted to $\mathrm{D}(C_1)$, we have

$$\begin{aligned}\left|\|\mathbf{v}\|^2\,\widehat{\mathrm{m}}_{S(\sqrt{r_1})} - 2C_4\,|\mathrm{dx}_2\wedge\mathrm{dx}_3|\right| &= 2C_4\left|\frac{\|\mathbf{v}\|^2}{2\sqrt{r_1}x_1}-1\right||\mathrm{dx}_2\wedge\mathrm{dx}_3| \\ &= 2C_4\left|\frac{\|\mathbf{v}\|^2-2\sqrt{r_1}x_1}{2\sqrt{r_1}x_1}\right||\mathrm{dx}_2\wedge\mathrm{dx}_3| \\ &\le 2C_4\left|\frac{|Q_0(\mathbf{v})|+4C_1^2}{\frac{1}{4}\,\|\mathbf{v}\|^2}\right||\mathrm{dx}_2\wedge\mathrm{dx}_3|\end{aligned}$$

Thus if integrating a function taking value in $[-M,M]$, the difference is at most

$$2C_4\left|\frac{Q_0(\mathbf{v})+4C_1^2}{\frac{1}{8}\,\|\mathbf{v}\|^2}\right|(2C_1)^2\cdot M = \|\mathbf{v}\|^{-2}\cdot\left|64C_4C_1^2(|Q_0(\mathbf{v})|+4C_1^2)\right|\cdot M.$$

Taking

$$C_5 := 256C_4C_1^4 \tag{54}$$

completes the proof. □

2.5. Proof. Fix $\mathbf{v}\in\mathbb{R}^4$, we identify $S(r)$ with a subset of $\mathbb{R}^4$ by embedding

$$(x_1,x_2,x_3)\mapsto(x_1,x_2,x_3,\sqrt{r_2(\mathbf{v})})_{\mathbf{f}}.$$

Let us state Lem.1.1 again:

LEMMA 2.10. *Given $f\in C_c(\mathbb{R}_{>0}\times\mathbb{R}^3)$ and $\varepsilon\in(0,1)$, there exists $T_0=T_0(f,\varepsilon)>0$ such that for every $T>T_0$, for every $\mathbf{v}\in\mathbb{R}^4$ we have*

$$\left|\frac{1}{2C_4}T^2\int f(\mathbf{a}_{\ln T}k.\mathbf{v})\widehat{\mathrm{m}}_{\mathrm{K}}(k)-\frac{T^2}{\|\mathbf{v}\|^2}\int f\left(\frac{\|\mathbf{v}\|}{T},w_2,w_3,w_4\right)|\mathrm{dw}_2\wedge\mathrm{dw}_3|\right|<\varepsilon$$

where

$$w_4 := \frac{Q_0(\mathbf{v})-w_2^2-w_3^2}{2\,\|\mathbf{v}\|\,T^{-1}}$$

is a function in (w_2,w_3), for every fixed $\mathbf{v}$ and T.

PROOF. We are going to choose some $T_0\ge 10C_1^3$.
Rewrite

$$\frac{1}{2C_4}T^2\int f(\mathbf{a}_{\ln T}k.\mathbf{v})\widehat{\mathrm{m}}_{\mathrm{K}}(k)=\frac{1}{2C_4}T^2\int f(\mathbf{a}_{\ln T}.\mathbf{w})\widehat{\mathrm{m}}_{\mathrm{K}.\mathbf{v}}(\mathbf{w})$$

By Lem.2.4, 2.6, if $T\ge T_0$, by change of variable $\mathbf{w}\mapsto(w_2,w_3):=(\mathbf{w}(2),\mathbf{w}(3))$:

$$\begin{aligned}\frac{1}{2C_4}T^2\int f(\mathbf{a}_{\ln T}.\mathbf{w})\widehat{\mathrm{m}}_{\mathrm{K}.\mathbf{v}}(\mathbf{w}) &= \frac{1}{2C_4}T^2\int_{f(\mathbf{a}_{\ln T}.\mathbf{w})\neq 0} f(\mathbf{a}_{\ln T}.\mathbf{w})\widehat{\mathrm{m}}_{\mathrm{K}.\mathbf{v}}(\mathbf{w}) \\ &= \frac{1}{2C_4}T^2\int_{\mathrm{D}(C_1)} f(\mathbf{a}_{\ln T}.\mathbf{v}^*(w_2,w_3))\widehat{\mathrm{m}}_{S(\sqrt{r_1})}(w_2,w_3).\end{aligned}$$

Note that when $f(\mathbf{a}_{\ln T}k.\mathbf{v})\neq 0$ for some $k\in K$, $T\ge 10C_1^3 \implies \|\mathbf{v}\|^2\ge Q_0(\mathbf{v})+4C_1^2$ by Lem.2.4. So Lem.2.6 is applicable to $\mathbf{v}$ and $(w_2,w_3):=(\mathbf{w}(2),\mathbf{w}(3))$. Moreover, Lem.2.6 implies that $\mathbf{w}=\mathbf{v}^*(\mathbf{w}(2),\mathbf{w}(3))$.

By Lem.2.3, the RHS is equal to

$$\frac{T^2}{\|\mathbf{v}\|^2}\int f\left(\frac{\|\mathbf{v}\|}{T},w_2,w_3,w_4\right)|\mathrm{dw}_2\wedge\mathrm{dw}_3| = \frac{T^2}{\|\mathbf{v}\|^2}\int_{\mathrm{D}(C_1)} f\left(\frac{\|\mathbf{v}\|}{T},w_2,w_3,w_4\right)|\mathrm{dw}_2\wedge\mathrm{dw}_3|$$

Recall from Lem.2.3 and 2.4 that when $f(\mathbf{a}_{\ln T}.\mathbf{w})\neq 0$ or when $f\left(\frac{\|\mathbf{v}\|}{T},w_2,w_3,w_4\right)\neq 0$, we always have

$$\frac{1}{C_1}T\le\|\mathbf{v}\|\le 2C_1T$$

and

$$|Q_0(\mathbf{v})| \leq 4C_1^2. \tag{55}$$

Now it suffices to show that

$$\left| \|\mathbf{v}\|^2 \int_{\mathrm{D}(C_1)} f(\mathbf{a}_{\ln T}.\mathbf{w}) \widehat{\mathrm{m}}_{S(\sqrt{r_1})}(w_2, w_3) - 2C_4 \int_{\mathrm{D}(C_1)} f\left(\frac{\|\mathbf{v}\|}{T}, w_2, w_3, w_4\right) |\mathrm{dw}_2 \wedge \mathrm{dw}_3| \right| < \varepsilon.$$

By Lem.2.7 and 2.8, for T large enough,

$$\left| 2C_4 \int_{\mathrm{D}(C_1)} f(\mathbf{a}_{\ln T}.\mathbf{w}) |\mathrm{dw}_2 \wedge \mathrm{dw}_3| - 2C_4 \int_{\mathrm{D}(C_1)} f\left(\frac{\|\mathbf{v}\|}{T}, w_2, w_3, w_4\right) |\mathrm{dw}_2 \wedge \mathrm{dw}_3| \right| < 0.5\varepsilon.$$

By Lem.2.9 and Equa.(55), for T large enough,

$$\left| \|\mathbf{v}\|^2 \int_{\mathrm{D}(C_1)} f(\mathbf{a}_{\ln T}.\mathbf{w}) \widehat{\mathrm{m}}_{S(\sqrt{r_1})}(w_2, w_3) - 2C_4 \int_{\mathrm{D}(C_1)} f(\mathbf{a}_{\ln T}.\mathbf{w}) |\mathrm{dw}_2 \wedge \mathrm{dw}_3| \right| < 0.5\varepsilon.$$

Combining these two, we are done.

□

3. Exercises

CHAPTER 14

Quantitative Oppenheim II, height function and nondivergence

Back to the Top.

Main reference: [**EMM98**].

If you are new to this circle of ideas, a first example to keep in mind maybe : $\mathbf{a}_t := \operatorname{diag}(e^t, e^{-t})$, $K := \mathrm{SO}_2(\mathbb{R})$, $X = \mathrm{X}_2$. Most arguments are trivialized here, yet you could see the main idea.

Notations

- $Q_0(x_1, x_2, x_3, x_4) := 2x_1x_4 + x_2^2 + x_3^2$ real quadratic form of signature $(3,1)$ on $\mathbb{R}^4$.
- Let $(\mathbf{e}_1, ..., \mathbf{e}_4)$ be the standard basis of $\mathbb{R}^4$; and for a vector v, define its coefficients by $v = \sum (v)_i \mathbf{e}_i$ and we also write $v = ((v)_1, ..., (v)_4)$.
- Let $(\mathbf{f}_1, ..., \mathbf{f}_4)$ be another ONB(=orthogonal normal basis) defined by $\mathbf{f}_2 = \mathbf{e}_2, \mathbf{f}_3 = \mathbf{e}_3$ and $\mathbf{f}_1 = \frac{\mathbf{e}_1 + \mathbf{e}_4}{\sqrt{2}}$, $\mathbf{f}_4 = \frac{\mathbf{e}_1 - \mathbf{e}_4}{\sqrt{2}}$. If $v = \sum a_i \mathbf{f}_i$, we also write $v = (a_1, ..., a_4)_{\mathbf{f}}$.
- One can verify that $Q_0((x_1, ..., x_4)_{\mathbf{f}}) = x_1^2 + x_2^2 + x_3^2 - x_4^2$.
- $\mathrm{K} := \mathrm{SO}_{Q_0}(\mathbb{R}) \cap \mathrm{SO}_4(\mathbb{R})$.
- $\mathbf{a}_t := \operatorname{diag}(e^{-t}, 1, 1, e^t)$, contained in $\mathrm{SO}_{Q_0}(\mathbb{R})$.

1. Outline of the proof

Recall by last lecture, it remains to show the following

THEOREM 1.1. *Let f be a compactly supported continuous function on $\mathbb{R}^4$ and let $\widetilde{f} : \mathrm{X}_4 \to \mathbb{R}$ be its Siegel transform. Let $g_0 \in G$ be such that $Q_0 \circ g_0$ is irrational. Then*

$$\lim_{t \to +\infty} \int_K \widetilde{f}(\mathbf{a}_t k g_0 \mathbb{Z}^4) \widehat{\mathrm{m}}_K(k) = \int \widetilde{f}(x) \widehat{\mathrm{m}}_{\mathrm{X}_4}(x).$$

As we explained, the difficulty here is that $\widetilde{f}$ is usually an integrable but unbounded function. And it suffices to show that the contribution of the part outside a large compact set is small. The following observation reduces the general task to a rather special function.

DEFINITION 1.2. *For a lattice $\Lambda \leq \mathbb{R}^4$, let*

$$\mathrm{ht}_\infty(\Lambda) := \max_{i=1,...,3} \sup_{\Delta \in \mathrm{Prim}^i(\Lambda)} \frac{1}{\|\Delta\|} = \max_{i=1,...,3} (\mathrm{sys}^{(i)}(\Lambda))^{-1}.$$

LEMMA 1.3. *Let f be a bounded, non-negative function with compact support on $\mathbb{R}^4$. Then there exists a constant $C_1 = C_1(f) > 1$ such that*

$$\widetilde{f}(\Lambda) \leq C_1 \cdot \mathrm{ht}_\infty(\Lambda), \ \forall \Lambda \in \mathrm{X}_4.$$

Proof is left as an exercise.

THEOREM 1.4. *For every $\varepsilon > 0$, there exists a compact set $\mathscr{C}_\varepsilon$ of X_4 such that for all $t > 0$,*

$$\int \left(\mathrm{ht}_\infty \cdot 1_{\mathrm{X}_4 \setminus \mathscr{C}_\varepsilon}\right)(\mathbf{a}_t k g_0 \mathbb{Z}^4) \widehat{\mathrm{m}}_K(k) \leq \varepsilon.$$

PROOF OF THM.1.1 ASSUMING THM.1.4. Without loss of generality assume $f \geq 0$.

Fix $\varepsilon > 0$, choose $\mathscr{C}_\varepsilon \subset \mathrm{X}_4$ as in Thm.1.4. Choose a compactly supported continuous function $1 \geq \varphi_\varepsilon \geq 1_{\mathscr{C}_\varepsilon}$. Thus by equidistribution theorem obtained in Ch.12, Thm.1.2.

$$\lim_{t\to+\infty} \int \left(\widetilde{f}\cdot\varphi_\varepsilon\right)(\mathbf{a}_t k g_0 \mathbb{Z}^4)\widehat{\mathrm{m}}_K(k) = \int \left(\widetilde{f}\cdot\varphi_\varepsilon\right)(x)\widehat{\mathrm{m}}_{\mathrm{X}_4}(x).$$

On the other hand by Thm.1.4 and Lem.1.3

$$\begin{aligned}\limsup_{t\to+\infty} \int \left(\widetilde{f}\cdot(1-\varphi_\varepsilon)\right)(\mathbf{a}_t k g_0 \mathbb{Z}^4)\widehat{\mathrm{m}}_K(k) &\leq \limsup_{t\to+\infty} \int \left(C_1 \operatorname{ht}_\infty \cdot 1_{\mathrm{X}_4\setminus\mathscr{C}_\varepsilon}\right)(\mathbf{a}_t k g_0 \mathbb{Z}^4)\widehat{\mathrm{m}}_K(k)\\ &\leq C_1\varepsilon.\end{aligned}$$

Combining both and letting $\varepsilon \to 0$ we are done.

□

In fact, something stronger than Thm.1.4 will be proved.

PROPOSITION 1.5. *For $\delta \in (0,1)$ (we only need for some $\delta > 0$) and $\Lambda_0 \in \mathrm{X}_4$, there exists $C_2 = C_2(\delta, \Lambda_0) > 0$ such that for all $t > 0$*

$$\int \operatorname{ht}_\infty^{1+\delta}(\mathbf{a}_t k.\Lambda_0)\widehat{\mathrm{m}}_K(k) \leq C_2.$$

This will be deduced from the following two propositions.

PROPOSITION 1.6. *For every $\varepsilon > 0$, there exist $C_4(\varepsilon) > 1$ and $t_0(\varepsilon) > 0$ such that for all $\Lambda \in \mathrm{X}_4$ (this is important!), we have*

$$\int \operatorname{ht}_\delta^{\mathrm{new}}(\mathbf{a}_{t_0(\varepsilon)} k.\Lambda)\widehat{\mathrm{m}}_K(k) \leq \varepsilon \operatorname{ht}_\delta^{\mathrm{new}}(\Lambda) + C_4(\varepsilon)$$

where $\operatorname{ht}_\delta^{\mathrm{new}} : \mathrm{X}_4 \to \mathbb{R}_{>0}$ is some function satisfying

$$C_5^{-1} \operatorname{ht}_\infty^{1+\delta} \leq \operatorname{ht}_\delta^{\mathrm{new}} \leq C_5 \operatorname{ht}_\infty^{1+\delta}.$$

Actually, we will find constants $c_0 > 0$ and $\kappa_i > 0$ for $i = 0,1,2,3,4$ such that

$$\operatorname{ht}_\delta^{\mathrm{new}}(\Lambda) = \sum_{i=1,2,3} c_0^{\kappa_i} (\operatorname{sys}^{(i)}(\Lambda))^{-1-\delta}.$$

To yield the result by applying this operator repeatedly, we need the following:

PROPOSITION 1.7. *For every open neighborhood V of identity in H, there exists a neighborhood U of identity in K such that for all $t, s \geq 0$*

$$\mathbf{a}_t U \mathbf{a}_s \subset K \cdot V \cdot \mathbf{a}_{t+s} \cdot K.$$

PROOF OF PROP.1.5. From the description of $\operatorname{ht}_\delta^{\mathrm{new}}$ as in Prop.1.6, we can find V_0, an open neighborhood of identity in H, such that

$$\frac{1}{2}\operatorname{ht}_\delta^{\mathrm{new}}(\Lambda) \leq \operatorname{ht}_\delta^{\mathrm{new}}(v.\Lambda) \leq 2\operatorname{ht}_\delta^{\mathrm{new}}(\Lambda),\ \forall v \in V_0,\ \Lambda \in \mathrm{X}_4.$$

Find U_0 by Prop.1.7. Let $\varepsilon := \frac{1}{4}\widehat{\mathrm{m}}_K(U_0)$. Applying Prop.1.6 we get some C_4, t_0. Let $C_6 := \frac{C_4}{\widehat{\mathrm{m}}_K(U_0)}$.

Fix $\Lambda_0 \in \mathrm{X}_4$, define a continuous function $\phi : G \to \mathbb{R}_{>0}$ by

$$\phi(g) := \int \operatorname{ht}_\delta^{\mathrm{new}}(gk.\Lambda_0)\widehat{\mathrm{m}}_K(k).$$

Thus it suffices to show that $\phi(\mathbf{a}_t)$, as t varies in $(0,+\infty)$, is bounded by Prop. 1.6.

The function ϕ enjoys the following properties

1. ϕ is bi-K-invariant;
2. for every $v \in V_0$ and $g \in G$, $\frac{1}{2}\phi(g) \leq \phi(vg) \leq 2\phi(g)$.

Combined with Prop.1.7, we see that for all $t \geq t_0$,

$$\phi(\mathbf{a}_{t_0} k \mathbf{a}_{t-t_0}) \geq \frac{1}{2}\phi(\mathbf{a}_t).$$

Also observe that

$$\begin{aligned}
\frac{1}{\widehat{\mathrm{m}}_K(U_0)}\int_{U_0}\phi(\mathbf{a}_{t_0}kg)\widehat{\mathrm{m}}_K(k) &\leq \frac{1}{\widehat{\mathrm{m}}_K(U_0)}\int_K \phi(\mathbf{a}_{t_0}kg)\widehat{\mathrm{m}}_K(k)\\
&\leq \frac{1}{\widehat{\mathrm{m}}_K(U_0)}\cdot\left(\frac{1}{4}\widehat{\mathrm{m}}_K(U_0)\phi(g)+C_4\right)\\
&= \frac{1}{4}\phi(g)+C_6.
\end{aligned}$$

Therefore, for $t > t_0$,

$$\begin{aligned}
\phi(\mathbf{a}_t) &= \frac{1}{\widehat{\mathrm{m}}_K(U_0)}\int_{U_0}\phi(\mathbf{a}_t)\widehat{\mathrm{m}}_K(k)\\
&\leq 2\frac{1}{\widehat{\mathrm{m}}_K(U_0)}\int_{U_0}\phi(\mathbf{a}_{t_0}k\mathbf{a}_{t-t_0})\widehat{\mathrm{m}}_K(k)\\
&\leq \frac{1}{2}\phi(\mathbf{a}_{t-t_0})+C_6.
\end{aligned}$$

Now, for $t > 0$, choose the unique $n_t \in \mathbb{Z}_{\geq 0}$ such that $t' := t - n_t t_0 \in (0, t_0]$. By applying the above inequality n_t times we get

$$\phi(\mathbf{a}_t) \leq \frac{1}{2^{n_t}}\phi(\mathbf{a}_{t'}) + C_6(1+\frac{1}{2}+(\frac{1}{2})^2+\ldots)$$

Hence $\phi(\mathbf{a}_t)$, as t varies in $(0,+\infty)$, is bounded. □

2. Wavefront lemma

We explain how Prop.1.7 is proved.

PROOF. I am pretending $K = \mathrm{SO}_4(\mathbb{R})$ here. The justification of the arguments here without this false assumption is left to you.

Every matrix g of determinant one can be written as

$$g = k_1 d k_2,\ k_i \in \mathrm{SO}_n(\mathbb{R}),\ d \text{ is a diagonal matrix.}$$

The order of the diagonal entries of d can be permuted by changing k_1, k_2. The middle matrix is uniquely determined if we further assume

$$d = \mathrm{diag}(d_1,\ldots,d_n),\ \text{ with } d_1 \geq d_2 \geq \ldots \geq d_n > 0.$$

We let $\alpha_i(g) := d_1\cdot\ldots\cdot d_i$. It suffices to show that, when $k \in K$ is close to identity, for every i, $\alpha_i(\mathbf{a}_{t_1}k\mathbf{a}_{t_2})$ is closed to $\alpha_i(\mathbf{a}_{t_1+t_2})$ multiplicatively.

To do this, note that

$$\alpha_i(g) = \sup_{\mathbf{v}\in\wedge^i\mathbb{R}^n,\ \|\mathbf{v}\|=1}\left\|g.\mathbf{v}\right\| = \sup_{\mathbf{v},\mathbf{w}\in\wedge^i\mathbb{R}^n,\ \|\mathbf{v}\|=\|\mathbf{w}\|=1}\left|\langle g.\mathbf{v},\mathbf{w}\rangle\right|.$$

For $\varepsilon \in (0,1)$, choose $U = U(\varepsilon) \subset K$ such that for all i,

$$|\langle u.e_1\wedge\ldots\wedge e_i, e_1\wedge\ldots\wedge e_i\rangle| \geq \frac{1}{1+\varepsilon}.$$

Now take $u \in U$. On the one hand,

$$\begin{aligned}
\left|\langle \mathbf{a}_{t_1}u\mathbf{a}_{t_2}.\mathbf{v},\mathbf{w}\rangle\right| &= \left|\langle u\mathbf{a}_{t_2}.\mathbf{v},\mathbf{a}_{t_1}.\mathbf{w}\rangle\right|\\
&\leq \left\|\mathbf{a}_{t_2}.\mathbf{v}\right\|\cdot\left\|\mathbf{a}_{t_1}.\mathbf{w}\right\| \leq \alpha_i(\mathbf{a}_{t_1+t_2}).
\end{aligned}$$

On the other hand,

$$\begin{aligned}&\left|\langle \mathbf{a}_{t_1} u \mathbf{a}_{t_2}.e_1\wedge ...\wedge e_i, e_1\wedge ...\wedge e_i\rangle\right|\\ =&\alpha_i(\mathbf{a}_{t_1+t_2})\left|\langle u.e_1\wedge ...\wedge e_i, e_1\wedge ...\wedge e_i\rangle\right| \geq \frac{1}{1+\varepsilon}\alpha_i(\mathbf{a}_{t_1+t_2}).\end{aligned}$$

So we are done.

□

3. The height function

Prop.1.6 relies on the following proposition on representations. It is here that we are avoiding the case of signature $(2,1)$ and $(2,2)$.

PROPOSITION 3.1. *For every $\varepsilon > 0$ there exists $t_1 = t_1(\varepsilon) > 0$ such that for all $t \geq t_1$, $\delta \in (0,1)$ and for all pure wedges $\mathbf{v}_{\neq 0} \in \wedge^i \mathbb{R}^n$ ($n = 4$ here), we have*

$$\int \|\mathbf{a}_t k.\mathbf{v}\|^{-1-\delta}\, \widehat{\mathrm{m}}_K(k) \leq \varepsilon \|\mathbf{v}\|^{-1-\delta}.$$

PROOF. Omitted for now. □

A "pure wedge" (also called "decomposable vector") refers to a vector $\mathbf{v} \in \wedge^i \mathbb{R}^n$ that can be written as $v_1 \wedge ... \wedge v_k$ for some $v_i \in \mathbb{R}^n$.

3.1. Preparations. Fix $\varepsilon \in (0,1)$, find $t_1(\varepsilon)$ as in Prop.3.1. Find $C_7 = C_7(\varepsilon) > 1$ such that

$$C_7^{-1}\|\mathbf{v}\| \leq \left\|\mathbf{a}_{t_1}.\mathbf{v}\right\| \leq C_7\|\mathbf{v}\|,\ \forall \mathbf{v} \in \sqcup \wedge^i \mathbb{R}^4.$$

Fix a strictly convex function $\kappa > 0$ on $[0,4]$. And find $C_8 > 1$ such that

$$\kappa_j \geq \frac{\kappa_{j-i} + \kappa_{j+i}}{2} + C_8^{-1};\quad \kappa_0 = \kappa_4 = 1$$

for all $j \in \{1,2,3\}$ and $j \pm i \in \{0,1,2,3,4\}$.

Choose $c_0 \in (0,1)$ small enough, depending on ε,

$$c_0^{2C_8^{-1}} \leq C_7^2 c_0^{2C_8^{-1}} \leq (\varepsilon C_7^{-1})^{100}.$$

Define

$$\mathrm{ht}^{\mathrm{new}}_\delta(\Lambda) = \sum_{i=1,2,3} c_0^{\kappa_i}(\mathrm{sys}^{(i)}(\Lambda))^{-1-\delta}. \tag{56}$$

3.2. The proof. For each $l = 1,2,3$ find $\Delta_1^{(l)} \in \mathrm{Prim}^l(\Lambda)$ such that $\mathrm{sys}^{(l)}(\Lambda) = \left\|\Delta_1^{(l)}\right\|$.

3.2.1. *Good indices.* We define $\mathrm{Good}(\Lambda) \subset \{1,2,3\}$ by

$$l \in \mathrm{Good}(\Lambda) \iff \forall \Delta \in \mathrm{Prim}^l(\Lambda) \setminus \Delta_1^{(l)},\ C_7^2\|\Delta\|^{-1} < \mathrm{sys}^{(l)}(\Lambda)^{-1}. \tag{57}$$

Thus for $l \in \mathrm{Good}(\Lambda)$, $\Delta \in \mathrm{Prim}^l(\Lambda) \setminus \Delta_1^{(l)}$ and $k \in K$,

$$\begin{aligned}&\left\|\mathbf{a}_{t_1}k.\Delta\right\|^{-1-\delta} \leq C_7^{1+\delta}\|\Delta\|^{-1-\delta} < C_7^{-1-\delta}\mathrm{sys}^{(l)}(\Lambda)^{-1-\delta} = C_7^{-1-\delta}\left\|\Delta_1^{(l)}\right\|^{-1-\delta} \leq \left\|\mathbf{a}_{t_1}k.\Delta_1^{(l)}\right\|^{-1-\delta}\\ &\implies \forall k \in K,\ \mathrm{sys}^{(l)}(\mathbf{a}_{t_1}k.\Lambda)^{-1-\delta} = \left\|\mathbf{a}_{t_1}k.\Delta_1^{(l)}\right\|^{-1-\delta}.\end{aligned} \tag{58}$$

This implies that

$$\begin{aligned}\int c_0^{\kappa_l}\mathrm{sys}^{(l)}(\mathbf{a}_{t_1}k.\Lambda)^{-1-\delta}\widehat{\mathrm{m}}_K(k) &= \int c_0^{\kappa_l}\left\|\mathbf{a}_{t_1}k.\Delta_1^{(l)}\right\|^{-1-\delta}\widehat{\mathrm{m}}_K(k)\\ &\leq \varepsilon c_0^{\kappa_l}\left\|\mathbf{a}_{t_1}k.\Delta_1^{(l)}\right\|^{-1-\delta}\\ &= \varepsilon\cdot c_0^{\kappa_l}\,\mathrm{sys}^{(l)}(\mathbf{a}_{t_1}k.\Lambda)^{-1-\delta}.\end{aligned} \tag{59}$$

3.2.2. *Bad indices.* $\mathrm{Bad}(\Lambda) := \{1,2,3\}\backslash\mathrm{Good}(\Lambda)$. In other words, we can find $\Delta_2^{(l)} \in \mathrm{Prim}^l(\Lambda)\backslash \Delta_1^{(l)}$ such that

$$C_7^2 \left\|\Delta_2^{(l)}\right\|^{-1} \geq \mathrm{sys}^{(l)}(\mathbf{a}_{t_1} k.\Lambda)^{-1}.$$

Recall the following inequalities

$$\left\|\Delta_1^{(l)}\right\| \cdot \left\|\Delta_2^{(l)}\right\| \geq \left\|\Delta_1^{(l)} \cap \Delta_2^{(l)}\right\| \cdot \left\|\Delta_1^{(l)} + \Delta_2^{(l)}\right\|,$$

from which we deduce that (let $a := \mathrm{rank}\,\Delta_1^{(l)} - \mathrm{rank}\,\Delta_1^{(l)} \cap \Delta_2^{(l)}$)

$$c_0^{2\kappa_l} \left\|\Delta_2^{(l)}\right\|^{-1-\delta} \left\|\Delta_2^{(l)}\right\|^{-1-\delta} \leq \left(c_0^{\kappa_{l-a}} \left\|\Delta_1^{(l)} \cap \Delta_2^{(l)}\right\|^{-1-\delta}\right) \cdot \left(c_0^{\kappa_{l+a}} \left\|\Delta_1^{(l)} + \Delta_2^{(l)}\right\|^{-1-\delta}\right) \cdot c_0^{2\kappa_l - \kappa_{l-a} - \kappa_{l+a}}.$$

For the LHS we have

$$C_7^{-4}\left(c_0^{\kappa_l}\,\mathrm{sys}^{(l)}(\Lambda)^{-1-\delta}\right)^2 \leq C_7^{-2(1+\delta)}\left(c_0^{\kappa_l}\,\mathrm{sys}^{(l)}(\Lambda)^{-1-\delta}\right)^2 \leq \mathrm{LHS}$$

and for the RHS,

$$\mathrm{RHS} \leq \left(c_0^{\kappa_{l-a}}\,\mathrm{sys}^{(l-a)}(\Lambda)^{-1-\delta}\right) \cdot \left(c_0^{\kappa_{l+a}}\,\mathrm{sys}^{(l+a)}(\Lambda)^{-1-\delta}\right) \cdot c_0^{C_8^{-1}}.$$

Since $c_0^{C_8^{-1}} \leq \varepsilon^{50} C_7^{-50}$, by combining the above equations we get

$$\left(c_0^{\kappa_l}\,\mathrm{sys}^{(l)}(\Lambda)^{-1-\delta}\right)^2 \leq \varepsilon^{50} C_7^{-46}\left(c_0^{\kappa_{l-a}}\,\mathrm{sys}^{(l-a)}(\Lambda)^{-1-\delta}\right) \cdot \left(c_0^{\kappa_{l+a}}\,\mathrm{sys}^{(l+a)}(\Lambda)^{-1-\delta}\right).$$

Thus

$$c_0^{\kappa_l}\,\mathrm{sys}^{(l)}(\Lambda)^{-1-\delta} \leq \varepsilon^{20} C_7^{-23} \max_{l'=0,\dots,4}\left\{c_0^{\kappa_{l'}}\,\mathrm{sys}^{(l')}(\Lambda)^{-1-\delta}\right\}.$$

Now we choose $l_1 = l_1(\Lambda, \delta)$ such that the maximum of RHS is achieved. Then $l_1 \in \mathrm{Good}(\Lambda) \cup \{0,4\}$. Also take $l_0 \in \mathrm{Bad}(\Lambda)$. Then for every $k \in K$,

$$\begin{aligned} c_0^{\kappa_{l_0}}\,\mathrm{sys}^{(l_0)}(\mathbf{a}_{t_1} k.\Lambda)^{-1-\delta} \leq & C_7^{1+\delta} c_0^{\kappa_{l_0}}\,\mathrm{sys}^{(l_0)}(\Lambda)^{-1-\delta} \leq \varepsilon^{20} C_7^{-20} c_0^{\kappa_{l_1}}\,\mathrm{sys}^{(l_1)}(\Lambda)^{-1-\delta} \\ \leq & \varepsilon^{20} C_7^{-18} c_0^{\kappa_{l_1}}\,\mathrm{sys}^{(l_1)}(\mathbf{a}_{t_1} k.\Lambda)^{-1-\delta}. \end{aligned}$$

3.2.3. *Wrap-up.* To save notation define

$$\alpha_l(\Lambda) := c_0^{\kappa_l}\,\mathrm{sys}^{(l)}(\Lambda)^{-1-\delta}.$$

$$\pi_*(\alpha_l)(\Lambda) := \int \alpha_l(\mathbf{a}_{t_1} k.\Lambda)\widehat{\mathrm{m}}_K(k).$$

So for $l \in \mathrm{Good}(\Lambda)$, we have

$$\pi_*(\alpha_l)(\Lambda) \leq \varepsilon \alpha_l(\Lambda).$$

For $l \in \mathrm{Bad}(\Lambda)$, we have ($l_1 = l_1(\Lambda)$ as above)

$$\pi_*(\alpha_l)(\Lambda) \leq \varepsilon^{20} C_7^{-18} \pi_*(\alpha_l)(\Lambda).$$

There are two cases.

Case I, $l_1 \in \{0, n\}$. In this case, for all l, $\alpha_l(\Lambda) \leq \max\{c_0^{\kappa_0}, c_0^{\kappa_n}\} = c_0$. Thus $\mathrm{ht}_\delta^{\mathrm{new}}(\Lambda) \leq 3c_0$. And

$$\pi_*(\mathrm{ht}_\delta^{\mathrm{new}})(\Lambda) \leq 3c_0 C_7^2.$$

Case II, $l_1 \in \mathrm{Good}(\Lambda)$.

$$\begin{aligned}
\pi_*(\mathrm{ht}_\delta^{\mathrm{new}})(\Lambda) &= \sum \pi_*(\alpha_l)(\Lambda) \\
&\le \varepsilon \sum_{l\in \mathrm{Good}(\Lambda)} \alpha_l(\Lambda) + \varepsilon^{20} C_7^{-18} \pi_*(\alpha_{l_1})(\Lambda) \\
&\le \varepsilon \sum_{l\in \mathrm{Good}(\Lambda)} \alpha_l(\Lambda) + \varepsilon^{21} C_7^{-18} \alpha_{l_1}(\Lambda) \\
&\le 2\varepsilon \sum_{l\in \mathrm{Good}(\Lambda)} \alpha_l(\Lambda) \le 2\varepsilon\, \mathrm{ht}_\delta^{\mathrm{new}}(\Lambda).
\end{aligned}$$

In either case, the following holds

$$\pi_*(\mathrm{ht}_\delta^{\mathrm{new}})(\Lambda) \le 3c_0 C_7^2 + 2\varepsilon\, \mathrm{ht}_\delta^{\mathrm{new}}(\Lambda) \tag{60}$$

for all $\Lambda \in \mathrm{X}_4$. Recall c_0 and C_7 are only dependent on ε.

4. Exercises

Bibliography

[BGHM10] Paul Buterus, Friedrich Götze, Thomas Hille, and Gregory Margulis, *Distribution of Values of Quadratic Forms at Integral Points*, arXiv e-prints (2010), arXiv:1004.5123. page 16.

[BM00] M. Bachir Bekka and Matthias Mayer, *Ergodic theory and topological dynamics of group actions on homogeneous spaces*, London Mathematical Society Lecture Note Series, vol. 269, Cambridge University Press, Cambridge, 2000. MR 1781937 pages 5, 7, 16, and 44.

[Bor91] Armand Borel, *Linear algebraic groups*, second ed., Graduate Texts in Mathematics, vol. 126, Springer-Verlag, New York, 1991. MR 1102012 page 85.

[Bor19] ______, *Introduction to arithmetic groups*, University Lecture Series, vol. 73, American Mathematical Society, Providence, RI, 2019, Translated from the 1969 French original [MR0244260] by Lam Laurent Pham, Edited and with a preface by Dave Witte Morris. MR 3970984 pages 85 and 86.

[BP92] Riccardo Benedetti and Carlo Petronio, *Lectures on hyperbolic geometry*, Universitext, Springer-Verlag, Berlin, 1992. MR 1219310 page 18.

[Cou16] Yves Coudène, *Ergodic theory and dynamical systems*, Universitext, Springer-Verlag London, Ltd., London; EDP Sciences, [Les Ulis], 2016, Translated from the 2013 French original [MR3184308] by Reinie Erné. MR 3586310 page 70.

[DE14] Anton Deitmar and Siegfried Echterhoff, *Principles of harmonic analysis*, second ed., Universitext, Springer, Cham, 2014. MR 3289059 page 64.

[DM93] S. G. Dani and G. A. Margulis, *Limit distributions of orbits of unipotent flows and values of quadratic forms*, I. M. Gel´fand Seminar, Adv. Soviet Math., vol. 16, Amer. Math. Soc., Providence, RI, 1993, pp. 91–137. MR 1237827 pages 91 and 95.

[DS84] S. G. Dani and John Smillie, *Uniform distribution of horocycle orbits for Fuchsian groups*, Duke Math. J. **51** (1984), no. 1, 185–194. MR 744294 pages 74 and 93.

[EEE+10] Manfred Leopold Einsiedler, David Alexandre Ellwood, Alex Eskin, Dmitry Kleinbock, Elon Lindenstrauss, Gregory Margulis, Stefano Marmi, and Jean-Christophe Yoccoz (eds.), *Homogeneous flows, moduli spaces and arithmetic*, Clay Mathematics Proceedings, vol. 10, American Mathematical Society, Providence, RI; Clay Mathematics Institute, Cambridge, MA, 2010. MR 2572399 page 6.

[EMM98] Alex Eskin, Gregory Margulis, and Shahar Mozes, *Upper bounds and asymptotics in a quantitative version of the Oppenheim conjecture*, Ann. of Math. (2) **147** (1998), no. 1, 93–141. MR 1609447 pages 91, 104, and 113.

[EMS96] Alex Eskin, Shahar Mozes, and Nimish Shah, *Unipotent flows and counting lattice points on homogeneous varieties*, Ann. of Math. (2) **143** (1996), no. 2, 253–299. MR 1381987 pages 91 and 96.

[EMS97] A. Eskin, S. Mozes, and N. Shah, *Non-divergence of translates of certain algebraic measures*, Geom. Funct. Anal. **7** (1997), no. 1, 48–80. MR 1437473 page 48.

[EW11] Manfred Einsiedler and Thomas Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London, Ltd., London, 2011. MR 2723325 pages 56, 63, and 70.

[Fur73] Harry Furstenberg, *The unique ergodicity of the horocycle flow*, Recent advances in topological dynamics (Proc. Conf., Yale Univ., New Haven, Conn., 1972; in honor of Gustav Arnold Hedlund), 1973, pp. 95–115. Lecture Notes in Math., Vol. 318. MR 0393339 page 65.

[Gal62] P. Gallagher, *Metric simultaneous diophantine approximation*, J. London Math. Soc. **37** (1962), 387–390. MR 157939 page 3.

[Gel14] Tsachik Gelander, *Lectures on Lattices and locally symmetric spaces*, arXiv e-prints (2014), arXiv:1402.0962. page 86.

[Hed36] Gustav A. Hedlund, *Fuchsian groups and transitive horocycles*, Duke Math. J. **2** (1936), no. 3, 530–542. MR 1545946 pages 1 and 7.

[Kem78] George R. Kempf, *Instability in invariant theory*, Ann. of Math. (2) **108** (1978), no. 2, 299–316. MR 506989 page 95.

[Kle10] Dmitry Kleinbock, *Quantitative nondivergence and its Diophantine applications*, Homogeneous flows, moduli spaces and arithmetic, Clay Math. Proc., vol. 10, Amer. Math. Soc., Providence, RI, 2010, pp. 131–153. MR 2648694 pages 35, 48, and 49.

[LM14] Elon Lindenstrauss and Gregory Margulis, *Effective estimates on indefinite ternary forms*, Israel J. Math. **203** (2014), no. 1, 445–499. MR 3273448 page 16.

[LMMS19] Elon Lindenstrauss, Amir Mohammadi, Gregory Margulis, and Nimish Shah, *Quantitative behavior of unipotent flows and an effective avoidance principle*, arXiv e-prints (2019), arXiv:1904.00290. page 91.

[Mar97] G. A. Margulis, *Oppenheim conjecture*, Fields Medallists' lectures, World Sci. Ser. 20th Century Math., vol. 5, World Sci. Publ., River Edge, NJ, 1997, pp. 272–327. MR 1622909 page 16.

[Moo80] Calvin C. Moore, *The Mautner phenomenon for general unitary representations*, Pacific J. Math. **86** (1980), no. 1, 155–169. MR 586875 page 84.

[MS95] Shahar Mozes and Nimish Shah, *On the space of ergodic invariant measures of unipotent flows*, Ergodic Theory Dynam. Systems **15** (1995), no. 1, 149–159. MR 1314973 pages 82 and 91.

[MW02] Yair Minsky and Barak Weiss, *Nondivergence of horocyclic flows on moduli space*, J. Reine Angew. Math. **552** (2002), 131–177. MR 1940435 page 48.

[O'M63] O. T. O'Meara, *Introduction to quadratic forms*, Die Grundlehren der mathematischen Wissenschaften, Band 117, Academic Press, Inc., Publishers, New York; Springer-Verlag, Berlin-Göttingen-Heidelberg, 1963. MR 0152507 page 19.

[Rag72] M. S. Raghunathan, *Discrete subgroups of Lie groups*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 68, Springer-Verlag, New York-Heidelberg, 1972. MR 0507234 pages 4 and 64.

[Ra91a] Marina Ratner, *On Raghunathan's measure conjecture*, Ann. of Math. (2) **134** (1991), no. 3, 545–607. MR 1135878 pages 82 and 91.

[Ra91b] Marina Ratner, *Raghunathan's topological conjecture and distributions of unipotent flows*, Duke Math. J. **63** (1991), no. 1, 235–280. MR 1106945 page 91.

[Ra92] ———, *Raghunathan's conjectures for* $\mathrm{SL}(2,\mathbf{R})$, Israel J. Math. **80** (1992), no. 1-2, 1–31. MR 1248925 pages 65 and 72.

[RS18] Rodolphe Richard and Nimish A. Shah, *Geometric results on linear actions of reductive Lie groups for applications to homogeneous dynamics*, Ergodic Theory Dynam. Systems **38** (2018), no. 7, 2780–2800. MR 3846726 page 98.

[RZ16] Rodolphe Richard and Thomas Zamojski, *Limit distribution of Translated pieces of possibly irrational leaves in S-arithmetic homogeneous spaces*, arXiv e-prints (2016), arXiv:1604.08494. page 98.

[Sha91a] Nimish A. Shah, *Closures of totally geodesic immersions in manifolds of constant negative curvature*, Group theory from a geometrical viewpoint (Trieste, 1990), World Sci. Publ., River Edge, NJ, 1991, pp. 718–732. MR 1170382 page 91.

[Sha91b] ———, *Uniformly distributed orbits of certain flows on homogeneous spaces*, Math. Ann. **289** (1991), no. 2, 315–334. MR 1092178 pages 82, 84, and 91.

[Sha09] ———, *Unipotent flows on products of* $\mathrm{SL}(2,K)/\Gamma$*'s*, Dynamical systems and Diophantine approximation, Sémin. Congr., vol. 19, Soc. Math. France, Paris, 2009, pp. 69–104. MR 2808404 page 91.

[Spe42] D. C. Spencer, *The lattice points of tetrahedra*, J. Math. Phys. Mass. Inst. Tech. **21** (1942), 189–197. MR 7767 page 3.

[Wal82] Peter Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York-Berlin, 1982. MR 648108 pages 56 and 84.

[Zhe21] Cheng Zheng, *On the denseness of some sparse horocycles*, arXiv e-prints (2021), arXiv:2108.08567. page 7.