NUMBER THEORY: LITTLEWOOD CONJECTURE

RUNLIN ZHANG[†]

Abstract. This is the notes for a course I taught in 2024SP. I thank everyone who attended the lectures and gave me feedbacks. These notes are still evolving and the notes for the last two/three lectures are still missing.

Contents

† Email: zhangrunlinmath@outlook.com.

Notation

The set of positive integers is denoted by $\mathbb{Z}^+$. For a real number x, let $\langle x \rangle$ denote the distance to the nearest integer, namely, $\langle x \rangle = \inf_{n\in\mathbb{Z}} |x-n|$. Leb denotes the standard Lebesgue measure on $\mathbb{R}^n$ where the n is understood from the context.

For $x, y \in \mathbb{Z}$ non-zero, we let $\gcd(x,y) \in \mathbb{Z}^+$ to be the greatest common divisor of $|x|$ and $|y|$. If one of them is zero but the other is not, we set $\gcd(x,y)$ to be the absolute value of the non-zero one. Also, $\gcd(0,0) := 0$. Two integers are said to be coprime iff $\gcd(x,y) = 1$.

Abbreviate "infinitely many" as "i.m."; " almost every " as "a.e.".

1. Lecture 1, Dirichlet's Theorem, Badly Approximable Numbers and Khintchin's Zero-one Law

References: I am mostly following [Zaf17, Cas50]. One may also consult the survey [BRV16] (available on arxiv).

1.1. **Prelude.** Number theory provides a huge amount of interesting problems. Besides "elementary" methods, tools from different branches of math are introduced to solve them. This course focuses on so-called "Diophantine approximation" problems, which are concerned with approximating real numbers by rational numbers. The second half of this course will be devoted to a specific (still unsolved!) problem: Littlewood conjecture. We will present the work of Einsiedler–Katok–Lindenstrauss on this conjecture, showing the exception set has dimension zero. They use tools coming from dynamics, which will be introduced later.

1.2. **The beginning.** The starting point of Diophantine approximation is the following:

Fact 1.1. *The set of rational numbers $\mathbb{Q}$ is dense in real numbers $\mathbb{R}$. In other words, for every $x \in \mathbb{R}$ and $\varepsilon > 0$, there exist two integers (p,q) with $q > 0$ such that $\left|x - \frac{p}{q}\right| < \varepsilon$.*

A more refined statement goes as:

Theorem 1.2 (Dirichlet)**.** *For every $x \in \mathbb{R}$ and $N \in \mathbb{Z}^+$, there exists $(p,q) \in \mathbb{Z}^2$ with $0 < q \le N$ such that*

$$\left|x - \frac{p}{q}\right| < \frac{1}{Nq}.$$

The proof is based on "drawer's principle" (or pigeon-hole principle).

Proof. For $k = 1, 2, ..., N$ find $n_k \in \mathbb{Z}$ such that $kx - n_k \in [0,1)$. Write

$$\{kx - n_k,\ k = 1, ..., N\} = \{x_1 \le x_2 \le ... \le x_N\}.$$

Thus, one of the $N+1$ numbers

$$\{x_1, x_2 - x_1, x_3 - x_2, ..., x_N - x_{N-1}, 1 - x_N\}$$

has to be strictly smaller than $\frac{1}{N}$ since they sum up to one. Say $x_{i_0} - x_{i_0-1} < \frac{1}{N}$. By convention, $x_0 := 0 = 0 \cdot x$ and $x_{N+1} := 1 = 0 \cdot x + 1$. Therefore, for some integers $\{k_1 < k_2\} \subset \{0, ..., N\}$, one has for some $p \in \mathbb{Z}$,

$$|k_2 x - k_1 x - p| < \frac{1}{N}$$

Let $q := k_2 - k_1$, then
$$\left|x - \frac{p}{q}\right| < \frac{1}{Nq},$$
proving the assertion. $\square$

1.3. **Badly approximable numbers.** As a corollary of the above theorem, one gets

Corollary 1.3. *For every $x \in \mathbb{R}$, there exist infinitely many pairs of integers (p, q) with $q > 0$ such that*
$$\left|x - \frac{p}{q}\right| < \frac{1}{q^2}.$$
In other words, $\langle qx \rangle < \frac{1}{q}$ for infinitely many $q \in \mathbb{Z}^+$.

Definition 1.4. *A real number x is said to be **badly approximable** iff there exists $c > 0$ such that*
$$\left|x - \frac{p}{q}\right| > \frac{c}{q^2}, \quad \forall\, (p, q) \in \mathbb{Z}^2,\ q > 0.$$
Or in other words, $\langle qx \rangle q > c$ for all $q \in \mathbb{Z}^+$. Let **BAD** *denote the set of badly approximable numbers. If an irrational number $x \in \mathbb{R} \setminus \mathbb{Q}$ is not badly approximable, we say that x is **well-approximable**.*

This definition is non-trivial in the sense that there are badly approximable numbers as well as well-approximable numbers.

Example 1.5. *$\sqrt{2}$ is badly approximable.*

Proof. Take $\varepsilon \in (0, 1)$. Assume that there are integers p, q with $q > 0$ such that
$$q\left|q\sqrt{2} - p\right| < \varepsilon.$$
Thus
$$\left|q\sqrt{2} + p\right| < \varepsilon/q + 2q\sqrt{2} < 4q.$$
Multiplying the above two together gives
$$q\left|2q^2 - p^2\right| < 4q\varepsilon \implies \left|2q^2 - p^2\right| < 4\varepsilon.$$
But $2q^2 - p^2$ is a non-zero integer, so $\left|2q^2 - p^2\right| \geq 1$. Thus $1 < 4\varepsilon$. This finishes the proof, showing $q\langle q\sqrt{2}\rangle \geq \frac{1}{4}$ for every $q \in \mathbb{Z}^+$. $\square$

Conjecture 1.6. *Algebraic numbers that are not contained in a quadratic number field are not bad.*

So far no single example seems known about this conjecture. For instance, it is unknown whether $\sqrt[3]{2}$ is badly approximable or not.

Remark 1.7. *However, for every $\varepsilon > 0$ and irrational algebraic number x, there exists $c = c(x, \varepsilon) > 0$ such that $q^{1+\varepsilon}\langle qx \rangle > c(x, \varepsilon)$ for every $q \in \mathbb{Z}^+$. This is a theorem of Roth. It has been generalized by Schmidt under the name of subspace theorem.*

Example 1.8. 0.10100001000000001... *(the n-th group of 0's consists of $n + m$ consecutive zeros if there are m digits in front of it) is well-approximable.*

1.4. **Littlewood conjecture.**

Conjecture 1.9 (Littlewood). *For every pair (x, y) of real numbers, for every $\varepsilon > 0$, there exists $q \in \mathbb{Z}^+$ such that*
$$q\langle qx \rangle\langle qy \rangle < \varepsilon.$$
Equivalently,
$$\inf_{q \in \mathbb{Z}^+} q\langle qx \rangle\langle qy \rangle = 0. \tag{1}$$

Remark 1.10. *Note that if one of x or y does not belong to* **BAD***, then Equa.(1) holds.*

Theorem 1.11 (Einsiedler–Katok–Lindenstrauss [EKL06])**.**
$$\dim\{(x, y) \in \mathbb{R}^2,\ \textit{Equa.}(1)\ \textit{fails}\ \} = 0.$$

This will be proved in later lectures. We will soon prove that

Theorem 1.12. $\mathrm{Leb}(\mathbf{BAD}) = 0$. *Consequently,*

$$\mathrm{Leb}\{(x,y) \in \mathbb{R}^2,\ \textit{Equa.}(1) \textit{ fails }\} = 0.$$

The following two theorems will not be proved, but I find it healthy to compare them with Littlewood conjecture and EKL's work.

Theorem 1.13 (Gallagher)**.**

$$\mathrm{Leb}\{(x,y) \in \mathbb{R}^2,\ \inf_{q \in \mathbb{Z}^+} q\langle qx\rangle\langle qy\rangle \cdot (\log q)^2 = 0\} = 0.$$

Theorem 1.14 (Badziahin [Bad13])**.**

$$\dim\{(x,y) \in \mathbb{R}^2,\ \inf_{q \in \mathbb{Z}^+} q\langle qx\rangle\langle qy\rangle \cdot \log q \log\log q = 0\} = 2.$$

We will sometimes restrict our attention to numbers in the interval $[0,1)$ without loss of generality.

1.5. **Khintchine's zero-one law.** Let $\psi : \mathbb{Z}^+ \to \mathbb{R}^+$ be a sequence of positive real numbers (for instance $\psi(q) := q^{-1}$). Define

$$\begin{aligned} W(\psi) &:= \left\{x \in [0,1),\ \left|x - \frac{p}{q}\right| < q^{-1}\psi(q) \text{ for i.m. } q \in \mathbb{Z}^+,\, p \in \mathbb{Z}\right\} \\ &= \left\{x \in [0,1),\ \langle qx\rangle < \psi(q) \text{ for i.m. } q \in \mathbb{Z}^+\right\} \end{aligned} \tag{2}$$

Theorem 1.15 (Khintchin)**.** *Assume ψ is non-increasing. Then,*

$$\mathrm{Leb}(W(\psi)) = \begin{cases} 0 & \textit{if } \sum \psi(n) < +\infty; \\ 1 & \textit{if } \sum \psi(n) = +\infty. \end{cases}$$

Remark 1.16. *The assumption that ψ is non-increasing cannot be dropped.*

That $\mathrm{Leb}(\mathbf{BAD}) = 0$ follows directly from this theorem.

Proof of Theorem 1.12 assuming Theorem 1.15. For every $c > 0$ and $q \in \mathbb{Z}^+$, let $\psi_c(q) := cq^{-1}$. Then, **BAD** is the complement in $[0,1)$ of the union of $W(\psi_{n^{-1}})$ as n ranges over positive integers. Thus it suffices to show that $\mathrm{Leb}(W(\psi_c)) = 1$ for every $c > 0$. By Theorem 1.15, this follows from the fact that $\sum_{n \in \mathbb{Z}^+} cn^{-1} = +\infty$. □

1.6. **Proof of the convergence part.** In this subsection, we explain the convergence part of Theorem 1.15. Namely, we assume $\sum \psi(n) < +\infty$ and prove $\mathrm{Leb}(W(\psi)) = 0$. We use the Borel–Cantelli lemma:

Lemma 1.17. *Let $(E_n)_{n \in \mathbb{Z}^+}$ be a sequence of measurable subsets of $[0,1)$ such that $\sum \mathrm{Leb}(E_n) < +\infty$. Then*

$$\mathrm{Leb}\left(\{x \in E_n \textit{ for i.m. } n\}\right) = 0.$$

Let

$$W_n(\psi) := \{x \in [0,1) \mid \langle nx\rangle < \psi(n)\}.$$

In light of Lemma 1.17, it suffices to show that $\sum \mathrm{Leb}(W_n(\psi)) < \infty$. Indeed, for n large enough (such that $\psi(n) < 0.5$),

$$\begin{aligned} W_n(\psi) &= \bigsqcup_{i=0,1,\dots,n-1} \{x \in W_n(\psi) \mid nx \in i + [0,1)\} \\ &= \bigsqcup_{i=0,1,\dots,n-1} \left\{x \in [\frac{i}{n}, \frac{i+1}{n}),\ nx \in [i, i+\psi(n)) \cup ((i+1) - \psi(n), i+1)\right\} \\ &= \bigsqcup_{i=0,1,\dots,n-1} \left[\frac{i}{n}, \frac{i}{n} + \frac{\psi(n)}{n}\right) \sqcup \left(\frac{i+1}{n} - \frac{\psi(n)}{n}, \frac{i+1}{n}\right). \end{aligned}$$

Hence,

$$\mathrm{Leb}(W_n(\psi)) = \sum_{i=0,1,\dots,n-1} \frac{2\psi(n)}{n} = 2\psi(n).$$

Thus the convergence of $\sum \mathrm{Leb}(W_n(\psi))$ follows.

1.7. **Proof of the divergence part.** From now on assume $\sum \psi(n) = +\infty$ and we wish to show $\mathrm{Leb}(W(\psi)) = 1$. The proof will consist of two steps: $\mathrm{Leb}(W(\psi)) > 0$ and $\mathrm{Leb}(W(\psi)) > 0 \implies \mathrm{Leb}(W(\psi)) = 1$.

1.8. **Cassels' zero-one law.** In this subsection we prove

Theorem 1.18. $\mathrm{Leb}(W(\psi)) = 0$ *or* 1.

Though we use the non-increasing feature of ψ below, this assumption can be removed without much effort.

Choose a bijection $n \mapsto \lambda_n$ from $\mathbb{Z}^+$ to $\mathbb{Q}$. For a rational number x, find coprime integers p, q with $q > 0$ (if $p = 0$, we set $q := 1$) such that $x = \frac{p}{q}$ and define $\Psi_{\mathrm{red}}(x) := q^{-1}\psi(q)$. Let

$$W_{\mathrm{red}}(\psi) := \{x \in [0,1) \mid |x - \lambda_n| < \Psi_{\mathrm{red}}(\lambda_n) \text{ for i.m. } n\}.$$

Then $W(\psi) = W_{\mathrm{red}}(\psi) \cup (\mathbb{Q} \cap [0,1))$.

For $k, N \in \mathbb{Z}^+$, let

$$E_k := \left\{ x \in [0,1) \;\middle|\; |x - \lambda_n| < \frac{1}{k}\Psi_{\mathrm{red}}(\lambda_n) \text{ for i.m. } n \right\},$$

$$E_k^N := \left\{ x \in [0,1) \;\middle|\; |x - \lambda_n| < \frac{1}{k}\Psi_{\mathrm{red}}(\lambda_n) \text{ for some } n > N \right\}.$$

Also let $E_\infty := \bigcap E_k$. So $E_1 = W_{\mathrm{red}}(\psi)$ and $E_k = \bigcap_{N=1}^{\infty} E_k^N$. Theorem 1.18 would follow from the following three lemmas.

For a positive integer n and $x \in [0,1)$, define $T_n(x)$ to be the unique element in $[0,1)$ such that $nx + \mathbb{Z} = T_n(x) + \mathbb{Z}$.

Lemma 1.19. *For every* $k \in \mathbb{Z}^+$, $T_k(E_\infty) \subset E_1$.

Proof. For $x \in E_k$ with $|x - \lambda_n| < k^{-1}\Psi_{\mathrm{red}}(\lambda_n)$, then $|kx - k\lambda_n| < \Psi_{\mathrm{red}}(\lambda_n) \le \Psi_{\mathrm{red}}(k\lambda_n)$. Thus $kx \in E_1$. □

Lemma 1.20. *For every measurable set* $E \subset [0,1)$ *with* $\mathrm{Leb}(E) > 0$ *and every* $\varepsilon > 0$, *there exists* $N \in \mathbb{Z}^+$ *such that*

$$\mathrm{Leb}\left(T_N(E)\right) > 1 - \varepsilon.$$

Let $\mathrm{I}(x, \delta) := (x - \delta, x + \delta)$. We will need Lebesgue's density theorem (see e.g. [SS05, Chapter 3, Theorem 1.4]) for characteristic functions of Borel measurable subsets.

Theorem 1.21 (Lebesgue density theorem)**.** *Let f be an integrable function on* $[0,1)$. *Then for Lebesgue almost every* $x \in [0,1)$, *one has*

$$\lim_{\varepsilon \to 0^+} \frac{1}{2\varepsilon} \int_{x-\varepsilon}^{x+\varepsilon} f(t)\,\mathrm{Leb}(t) = f(x).$$

Proof of Lemma 1.20. By Lebesgue density theorem, find $\theta^* \in E$ such that for any $\varepsilon > 0$, there exists $\eta^* > 0$ such that for every $0 < \eta < \eta^*$,

$$\frac{\mathrm{Leb}(E \cap \mathrm{I}(\theta^*, \eta))}{2\eta} > 1 - \varepsilon.$$

Choose $\eta = \frac{1}{2N}$ for $N \in \mathbb{Z}^+$ sufficiently large such that the above inequality holds. Then

$$\mathrm{Leb}\left(NE \cap (N\theta^* - \frac{1}{2}, N\theta^* + \frac{1}{2}) \right) = N\,\mathrm{Leb}(E \cap \mathrm{I}(\theta^*, \eta)) > 1 - \varepsilon.$$

By reducing modulo $\mathbb{Z}$, we get $\mathrm{Leb}(T_N(E)) > 1 - \varepsilon$. □

Lemma 1.22. *For every* $k \in \mathbb{Z}^+$, $\mathrm{Leb}(E_1 \setminus E_k) = 0$. *Consequently,* $\mathrm{Leb}(E_\infty) > 0$ *if* $\mathrm{Leb}(E_1) > 0$.

Proof. To save notation write $\alpha_n := \Phi_{\mathrm{red}}(\lambda_n)$ in the proof. Assume $\mathrm{Leb}(E_1 \setminus E_k) > 0$, find $N \in \mathbb{Z}^+$ large enough such that $\mathrm{Leb}(E_1 \setminus E_k^N) > 0$.

By Lebesgue density theorem again, we find $\theta^*_{\neq 0} \in E_1$ such that for every $\varepsilon > 0$, there exists $\eta^*(\varepsilon) > 0$ such that for every $0 < \eta < \eta^*(\varepsilon)$, one has

$$\mathrm{Leb}\left(\mathrm{I}(\theta^*, \eta) \cap E_k^N \right) < \varepsilon\,\mathrm{Leb}\left(\mathrm{I}(\theta^*, \eta) \right). \tag{3}$$

We take $\varepsilon := \frac{1}{2(k+1)}$ and write $\eta^* := \eta^*(\varepsilon)$. Take n sufficiently large such that $n > N$, $2\alpha_n < \eta^*$ and, since $\theta^* \in E_1$, $|\theta^* - \lambda_n| < \alpha_n$.

By definition, one has $\mathrm{I}\left(\lambda_n, \frac{1}{k}\alpha_n\right) \subset E_k^N$. Let $\eta := |\theta^* - \lambda_n| + \frac{1}{k}\alpha_n$, which is smaller than η^*. Also, $\mathrm{I}(\lambda_n, \frac{1}{k}\alpha_n) \subset \mathrm{I}(\theta^*, \eta)$. Hence,

$$\begin{aligned} &\frac{\mathrm{Leb}(\,\mathrm{I}(\lambda_n, \frac{\alpha_n}{k})\,)}{\mathrm{Leb}(\,\mathrm{I}(\theta^*, \eta)\,)} = \frac{\alpha_n/k}{\eta} > \frac{\alpha_n/k}{\alpha_n + \alpha_n/k} = \frac{1}{1+k} \\ &\implies \mathrm{Leb}\left(\,\mathrm{I}(\theta^*, \eta) \cap E_k^N\,\right) > \frac{1}{1+k}\,\mathrm{Leb}(\,\mathrm{I}(\theta^*, \eta)\,), \end{aligned}$$

which is a contradiction against Equa.(3). □

Proof of Theorem 1.18. Assume $\mathrm{Leb}(E_1) > 0$ and want to show $\mathrm{Leb}(E_1) = 1$. By Lemma 1.22, $\mathrm{Leb}(E_\infty) > 0$. Apply Lemma 1.20 to $E = E_\infty$, we get $\mathrm{Leb}(\bigcup_{n\in\mathbb{Z}^+} T_n(E_\infty)) = 1$. But this set is contained in E_1 by Lemma 1.19. So we obtain $\mathrm{Leb}(E_1) = 1$ and we are done. □

1.9. **Partial converse to Borel–Cantelli.** The proof of the divergence part is more difficult partly because the converse to the Borel–Cantelli lemma is not true. However, we do have a partial converse assuming certain independence properties for the sequence of sets (E_n).

Lemma 1.23. *Let (E_n) be a sequence of measurable subsets of $[0,1)$. Then for every pair of integers $0 < m < n$, we have*

$$\mathrm{Leb}(\bigcup_{i=m}^{n} E_i) \geq \frac{\left(\sum\limits_{i=m}^{n} \mathrm{Leb}(E_i)\right)^2}{\sum\limits_{i=m}^{n}\sum\limits_{j=m}^{n} \mathrm{Leb}(E_i \cap E_j)}. \tag{4}$$

Proof. This is a consequence of Cauchy-Schwarz.

$$\left(\int_0^1 \left(\sum_{i=m}^{n} \mathbf{1}_{E_i}\right) \cdot (\mathbf{1}_{\cup_{i=m}^n E_i})\,\mathrm{Leb}\right)^2 \leq \int_0^1 \left(\sum_{i=m}^{n} \mathbf{1}_{E_i}\right)^2 \mathrm{Leb} \cdot \int_0^1 \mathbf{1}^2_{\cup_{i=m}^n E_i}\,\mathrm{Leb}$$

For the left hand side one has:

$$\left(\int_0^1 \left(\sum_{i=m}^{n} \mathbf{1}_{E_i}\right) \cdot (\mathbf{1}_{\cup_{i=m}^n E_i})\,\mathrm{Leb}\right)^2 = \left(\sum_{i=m}^{n} \mathrm{Leb}(E_i)\right)^2,$$

and for the right hand side:

$$\int_0^1 \left(\sum_{i=m}^{n} \mathbf{1}_{E_i}\right)^2 \mathrm{Leb} \cdot \int_0^1 \mathbf{1}^2_{\cup_{i=m}^n E_i}\,\mathrm{Leb} = \sum_{i,j=m}^{n} \mathrm{Leb}(E_i \cap E_j) \cdot \mathrm{Leb}\left(\bigcup_{i=m}^{n} E_i\right).$$

Putting them together finishes the proof. □

From this lemma, one can easily prove the following converse to Borel–Cantelli.

Lemma 1.24. *Let $(E_n)_{n\in\mathbb{Z}^+}$ be a sequence of measurable subsets of $[0,1)$ such that $\sum \mathrm{Leb}(E_n) < \infty$. Assume furthermore that $\mathrm{Leb}(E_i \cap E_j) = \mathrm{Leb}(E_i)\,\mathrm{Leb}(E_j)$ for every $i \neq j$ (namely, E_n's are independent from each other). Then*

$$\mathrm{Leb}\left(\{x \in E_n \textit{ for i.m. } n\}\right) = 1.$$

However, our sets are not independent. Nevertheless, we will be able to find a lower bound for the RHS of Equa.(4), which shows that $W(\psi)$ has positive Lebesgue measure. Our proof is then complete by invoking Theorem 1.18.

1.10. **A reduction.** Let $\psi_1(n) := \min\{\psi(n), \frac{1}{n}\}$. As $W(\psi_1) \subset W(\psi)$, it suffices to show that $\mathrm{Leb}(W(\psi_1)) = 1$.

Lemma 1.25. $\sum \psi_1(n) = +\infty$.

Remark 1.26. *It is not true in general that for two non-increasing sequence (a_n) and (b_n) of positive real numbers, $\sum a_n = \sum b_n = +\infty$ would imply $\sum \min\{a_n, b_n\} = +\infty$.*

Proof. Assuming $\sum \psi_1(n) < +\infty$, we will show that $\sum \psi(n) < +\infty$, which is a contradiction.

Decompose $\mathbb{Z}^+ \setminus \{1\} = I \sqcup J$ such that

$$i \in I \iff \psi(n) \leq \frac{1}{n}, \quad i \in J \iff \psi(n) > \frac{1}{n}.$$

Thus $\sum_I \psi(n) < +\infty$ and $\sum_J \frac{1}{n} < +\infty$. Decompose $J = \bigsqcup_{i\in\mathbb{Z}^+} J_i$ where $J_i = \{a_i, a_i + 1, ..., b_i\}$ and $b_i + 1 < a_{i+1}$. Therefore

$$+\infty > \sum_{n\in J} \frac{1}{n} > \sum_{i\in\mathbb{Z}^+} \int_{a_i}^{b_i+1} \frac{1}{x} \mathrm{dx} = \log\left(\frac{b_i+1}{a_i}\right).$$

On the other hand,

$$\sum_{n\in J} \psi(n) = \sum_i \sum_{j\in J_i} \psi(n) \leq \sum_i \sum_{j\in J_i} \psi(a_i - 1) = \sum_i \sum_{j\in J_i} \frac{1}{a_i - 1} = \sum_i \frac{b_i - (a_i - 1)}{a_i - 1}.$$

Define $\lambda_i := \frac{b_i}{a_i-1} - 1 > 0$ for every $i \in \mathbb{Z}^+$, then $\sum_{n\in J} \psi(n) \leq \sum \lambda_i$, which will be shown to be convergent.

Note that (for $a_i > 1$)

$$\frac{b_i+1}{a_i} - 1 > \frac{1}{2}(\frac{b_i}{a_i - 1} - 1).$$

Indeed for $\frac{p}{q} > 1$ with $q > 1$, one has $\frac{p+1}{q+1} - 1 > \frac{1}{2}(\frac{p}{q} - 1)$. Since this, after the denominators are cleared, is equivalent to $(q-1)(p+q) > 0$.

So we have $\sum \log(1 + \frac{1}{2}\lambda_i)$ is convergent. This implies that $\sum \lambda_i$ is convergent by Lemma 1.27. ☐

Lemma 1.27. *Let (λ_n) be a sequence of non-negative real numbers, one has that*

$$\sum \lambda_n < +\infty \iff \sum \ln(1 + \lambda_n) < +\infty.$$

Proof. Note that we may assume that (λ_n) tends to 0 for otherwise both sides are divergent. For $x \geq 0$, $\ln(1+x) \leq x$. Conversely, for x sufficiently small, $\ln(1+x) > \frac{1}{2}x$. So we are done. ☐

In light of Lemma 1.25, we will assume $\psi(n) \leq \frac{1}{2n}$ in the next subsection.

1.11. **Quasi-independence.** For this subsection, define

$$E_n := \bigcup_{q=2^{n-1}}^{2^n-1} \bigcup_{p\in\{1,\dots,q\},\gcd(p,q)=1} \mathrm{I}\left(\frac{p}{q}, \frac{\psi(2^n)}{2^n}\right).$$

As $\psi(n) \leq \frac{1}{2n}$, one can check that for every two distinct indices $(p,q),(p',q')$ appearing above,

$$\mathrm{I}\left(\frac{p}{q}, \frac{\psi(2^n)}{2^n}\right) \cap \mathrm{I}\left(\frac{p'}{q'}, \frac{\psi(2^n)}{2^n}\right) = \emptyset.$$

Also, since $\frac{\psi(2^n)}{2^n} \leq \frac{\psi(q)}{q}$, the set E_n is contained in

$$W_n(\psi) := \left\{x \in [0,1) \;\middle|\; \left|x - \frac{p}{q}\right| < \Psi_{\mathrm{red}}(\frac{p}{q}) \text{ for some } 2^{n-1} \leq q \leq 2^n - 1\right\}.$$

Thus, if x belongs to E_n infinitely many n's, then x belongs to $W(\psi)$. Therefore, it suffices to prove that

- RHS of Equa.(4) for such (E_i) has a lower bound independent of M and for N large enough, i.e., there exists $C > 0$ such that for every M, for N large enough

$$\sum_{i,j=M}^{N} \mathrm{Leb}(E_i \cap E_j) \leq C\left(\sum_{i=M}^{N} \mathrm{Leb}(E_i)\right)^2;$$

- $\sum \mathrm{Leb}(E_n) = +\infty$.

Let ϕ be Euler's totient function. Namely, for a positive integer N, $\phi(N) := |(\mathbb{Z}/N\mathbb{Z})^\times|$ is the number of integers in $\{1, ..., N\}$ that are coprime to N. Firstly we have

$$\mathrm{Leb}(E_n) = \left(2 \cdot \frac{\psi(2^n)}{2^n}\right) \cdot \sum_{q=2^{n-1}}^{2^n-1} \phi(q). \tag{5}$$

Then estimate the Lebesgue measure of $E_m \cap E_n$ for $m < n$. For (a, b) (resp. (c, d)) appearing in the index of E_m (resp. E_n), one has

$$\mathrm{Leb}\left(\mathrm{I}(\frac{a}{b}, \frac{\psi(2^m)}{2^m}) \cap \mathrm{I}(\frac{c}{d}, \frac{\psi(2^n)}{2^n})\right) \leq \mathrm{Leb}\left(\mathrm{I}(\frac{c}{d}, \frac{\psi(2^n)}{2^n})\right) = 2\frac{\psi(2^n)}{2^n}.$$

For distinct $(c_1, d_1), (c_2, d_2)$ appearing in the index of E_n, one has

$$\left|\frac{c_1}{d_1} - \frac{c_2}{d_2}\right| = \left|\frac{c_1 d_2 - c_2 d_1}{d_1 d_2}\right| \geq \frac{1}{d_1 d_2} \geq \frac{1}{2^{2n}}.$$

Thus, for every fixed (a, b) appearing in the index of E_m, the number of (c, d) appearing in the index of E_n such that $\mathrm{I}(\frac{a}{b}, \frac{\psi(2^m)}{2^m}) \cap \mathrm{I}(\frac{c}{d}, \frac{\psi(2^n)}{2^n}) \neq \emptyset$ is at most

$$\frac{2\frac{\psi(2^m)}{2^m}}{\frac{1}{2}\frac{1}{2^{2n}}} + 2 = 4 \cdot 2^{2n} \cdot \frac{\psi(2^m)}{2^m} + 2.$$

Therefore,

$$\mathrm{Leb}(E_m \cap E_n) \leq \left(2\frac{\psi(2^n)}{2^n}\right) \cdot \left(4 \cdot 2^{2n} \cdot \frac{\psi(2^m)}{2^m} + 2\right) \cdot \left(\sum_{q=2^{m-1}}^{2^m-1} \phi(q)\right)$$

Combining with Equa.(5), one has (for $m < n$)

$$\mathrm{Leb}(E_m \cap E_n) \leq 2 \cdot \frac{2^{2n}}{\sum_{q=2^{n-1}}^{2^n-1} \phi(q)} \cdot \mathrm{Leb}(E_n)\,\mathrm{Leb}(E_m) + \left(4\frac{\psi(2^n)}{2^n}\right) \cdot \left(\sum_{q=2^{m-1}}^{2^m-1} \phi(q)\right). \tag{6}$$

Before proceeding further, note two consequences of Lemma 1.28 to be presented in the next subsection. There exists a constant $C > 0$ such that for all positive integers k[1],

$$2^{2k} \leq C \cdot \sum_{q=2^{k-1}}^{2^k-1} \phi(q);$$

$$\sum_{q=1}^{2^{k-1}-1} \phi(q) \leq C \cdot \sum_{q=2^{k-1}}^{2^k-1} \phi(q).$$

The first inequality and Equa.(5) imply that

$$\sum_{n=1}^{N} \mathrm{Leb}(E_n) \geq \sum_{n=1}^{N} 2C^{-1} 2^n \psi(2^n) \geq 2C^{-1} \sum_{n=1}^{N} \sum_{q=2^n}^{2^{n+1}-1} \psi(q) = \sum_{q=2}^{2^{N+1}-1} \psi(q)$$

which diverges to $+\infty$.

Take two positive integers $M < N$.

Now we go back to Equa.(6) and sum over $m < n, m, n = M, ..., N$. The first summand in Equa.(6) is bounded from above by

$$\sum_{m<n, m,n=M,...,N} 2C\,\mathrm{Leb}(E_m)\,\mathrm{Leb}(E_n)$$

whereas the second summand is

$$\sum_{n=M}^{N} 4\frac{\psi(2^n)}{2^n} \sum_{q=2^{M-1}}^{2^{n-1}-1} \phi(q) \leq \sum_{n=M}^{N} C\,\mathrm{Leb}(E_n).$$

[1] One only needs the qualitative version $\frac{\sum \phi(n)}{N} \to c > 0$ to achieve this.

Consequently,

$$\sum_{m,n=M,\dots,N} \operatorname{Leb}(E_m \cap E_n) \le 2C \sum_{m,n=M,\dots,N} \operatorname{Leb}(E_m)\operatorname{Leb}(E_n) + 3C\sum_{n=M}^{N}\operatorname{Leb}(E_n)$$
$$= 2C\left(\sum_{n=M}^{N}\operatorname{Leb}(E_n)\right)^2 + 3C\sum_{n=M}^{N}\operatorname{Leb}(E_n).$$

Since $\sum \operatorname{Leb}(E_n)$ diverges, there exists $C' > 0$ (independent of M) such that

$$\sum_{n=M}^{N}\operatorname{Leb}(E_n) < C'\left(\sum_{n=M}^{N}\operatorname{Leb}(E_n)\right)^2$$

for all N large enough. Thus

$$\sum_{m,n=M,\dots,N} \operatorname{Leb}(E_m \cap E_n) \le C''\left(\sum_{n=M}^{N}\operatorname{Leb}(E_n)\right)^2$$

for some $C'' > 0$ and N large enough, completing the proof.

1.12. **Average of Euler's totient function.**

Lemma 1.28. *For any integer $N \ge e$, one has*

$$\left|\sum_{n=1}^{N}\phi(n) - \frac{1}{2\zeta(2)}N^2\right| \le 5N\ln N$$

where $\zeta(s) := \sum_{n\in\mathbb{Z}^+}\frac{1}{n^s}$ is the usual Riemann zeta function.

One may note that $\sum\phi(n)$ is counting primitive integral vectors in a cone.

The proof is based on "Fubini", "change of variable" and the Mobius function:

Definition 1.29. *Decompose a positive integer $n \ne 1$ into products of distinct prime numbers $n = \prod_{i=1}^{k} p_i^{d_i}$ with $d_i \in \mathbb{Z}^+$ and $k \in \mathbb{Z}^+$. Define the Mobius function $\mu : \mathbb{Z}^+ \to \{-1,0,1\}$ by*

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \ne 1,\ d_i = 1 \text{ for every } i; \\ 1 & \text{if } n = 1; \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 1.30. *For $n \in \mathbb{Z}^+$, one has*

$$\sum_{d|n}\mu(n) = \begin{cases} 0 & n \ne 1 \\ 1 & n = 1 \end{cases}$$

Proof. $0 = (1-1)^n = \sum\binom{n}{j}(-1)^j = \sum_{d|n}\mu(n)$. □

Lemma 1.31.

$$\sum_{d\in\mathbb{Z}^+}\frac{\mu(d)}{d^2}\cdot\sum_{n\in\mathbb{Z}^+}\frac{1}{n^2} = 1, \text{ or equivalently, } \sum_{d\in\mathbb{Z}^+}\frac{\mu(d)}{d^2} = \zeta(2)^{-1}.$$

Proof. Expand the product and apply the lemma above. □

Proof of Lemma 1.28.

$$\sum_{n=1}^{N}\phi(n) = \sum_{n=1}^{N}\sum_{m=1,\dots,n;\,(m,n)=1} 1 = \sum_{n=1}^{N}\sum_{m=1}^{n}\sum_{d|(m,n)}\mu(d)$$
$$= \sum_{n=1}^{N}\sum_{d|n}\sum_{m=1,\dots,n;\,d|m}\mu(d) = \sum_{n=1}^{N}\sum_{d|n}\frac{n}{d}\mu(d)$$
$$= \sum_{\{(m,d),\ md\le N\}} m\mu(d) = \sum_{d=1}^{N}\mu(d)\sum_{m=1}^{\lfloor\frac{N}{d}\rfloor} m$$
$$= \sum_{d=1}^{N}\mu(d)\left(\frac{1}{2}\frac{N^2}{d^2} + \operatorname{error}_1(d)\right)$$

where

$$|\mathrm{error}_1(d)| = \left| \int_0^{\lfloor \frac{N}{d} \rfloor} x\mathrm{dx} + \int_{\lfloor \frac{N}{d} \rfloor}^{\frac{N}{d}} x\mathrm{dx} - \sum_{m=1}^{\lfloor \frac{N}{d} \rfloor} m \right| \leq \frac{1}{2}\frac{N}{d} + \frac{N}{d} \leq 2\frac{N}{d}.$$

So if $N \geq e$,

$$\left| \sum_{d=1}^{N} \mathrm{error}_1(d) \right| \leq \sum_{d=1}^{N} 2\frac{N}{d} \leq 2N \left(1 + \sum_{d=2}^{N} \frac{1}{d} - \int_1^N \frac{1}{x}\mathrm{dx} \right) \leq 2N(\ln(N) + 1) \leq 4N\ln(N).$$

Therefore,

$$\sum_{n=1}^{N} \phi(n) = \frac{N^2}{2} \sum_{d=1}^{+\infty} \frac{\mu(d)}{d^2} + \mathrm{error}_2(d)$$

with

$$|\mathrm{error}_2(d)| \leq 4N\ln(N) + \frac{N^2}{2} \sum_{d=N+1}^{\infty} \frac{1}{d^2} \leq 4N\ln(N) + \frac{N^2}{2} \int_N^{\infty} \frac{1}{x^2}\mathrm{dx} \leq 5N\ln(N)$$

if $N \geq e$.

□

1.13. **Exercises.**

Exercise A. *Prove that there exists $c_0 > 0$ such that for every $q \in \mathbb{Z}^+$, $q^2\langle q\sqrt[3]{2}\rangle > c_0$.*

Exercise B. *Prove Cassels' zero-one law Theorem 1.18 without assuming ψ to be non-increasing.*

Exercise C. *Find two non-increasing sequences of positive numbers $(a_n)_{n\in\mathbb{Z}^+}$ and $(b_n)_{n\in\mathbb{Z}^+}$ such that $\sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} b_n = +\infty$ but $\sum_{n=1}^{\infty} \min\{a_n, b_n\} < +\infty$.*

Define (N is a positive integer)

$$\mathcal{L} := \left\{(x, y) \in \mathbb{Z}^2 \;\middle|\; \gcd(x, y) = 1,\ 0 < x < y\right\},$$
$$\mathcal{L}_N := \{(x, y) \in \mathcal{L} \mid y < N\}.$$

For every $(x, y) \in \mathcal{L}$, define $\pi(x, y) := (1, \frac{y}{x}) \in \{1\} \times (0, 1)$. For every $N \in \mathbb{Z}^+$, define a measure μ_N on $\{1\} \times (0, 1)$ by

$$\mu_N := \frac{1}{\#\mathcal{L}_N} \sum_{(x,y)\in\mathcal{L}_N} \delta_{\pi(x,y)}$$

where $\delta_{(x,y)}$ is the Dirac measure supported on (x, y) defined by

$$\delta_{(x,y)}(E) = \begin{cases} 1 & \text{if } (x, y) \in E \\ 0 & \text{if } (x, y) \notin E. \end{cases}$$

Exercise D. *Prove that (μ_N) converges, under the weak$*$ topology, to the standard Lebesgue measure on $\{1\} \times (0, 1)$. For simplicity, you are only required to show the following: for every interval $(a, b) \subset (0, 1)$, one has*

$$\lim_{N\to\infty} \mu_N(\{1\} \times (a, b)) = b - a.$$

(Hint: the proof of Lemma 1.28 might be helpful)

Notation

Vectors in $\mathbb{R}^n$, by default, are written as column vectors. For a few $\mathbf{x}_1, ..., \mathbf{x}_k$, write $(\mathbf{x}_1, ..., \mathbf{x}_k)$ for the n-by-k matrix whose i-th column is given by $\mathbf{x}_i$. We use I_2 to denote the two-by-two identity matrix.

2. Lecture 2, Space of Lattices of $\mathbb{R}^2$, Dani's Correspondence and Ergodic Theory

One may consult Cassels' book [Cas59] for facts about lattices in $\mathbb{R}^n$. For an introduction to ergodic theory, we recommend Einsiedler–Ward's book [EW11]. The proof of mixing of the geodesic flow is taken from Witte Morris' excellent book [Mor15]. For relation between Khintchine's theorem and exponential mixing, which is not discussed here, see the work of Kleinbock–Margulis [KM99]. The interaction between homogeneous dynamics and Diophantine approximation (especially the metric aspects) is very fruitful. See [Kle23] for a survey.

2.1. Prelude.

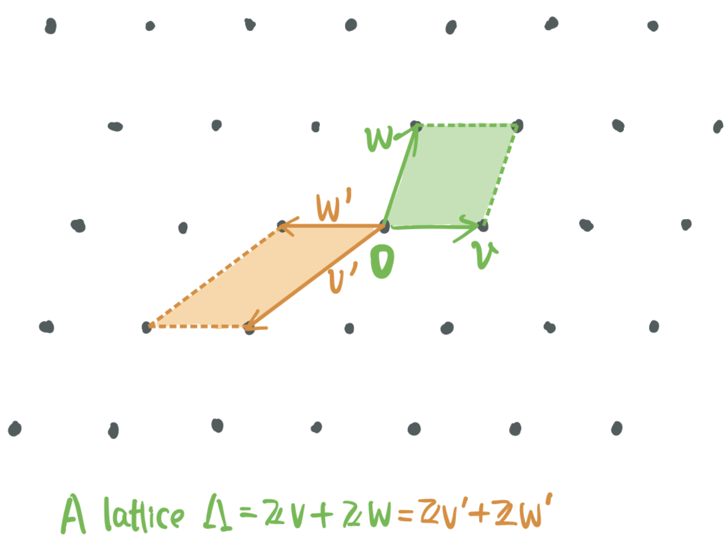
Certain problems in Diophantine approximations can be restated in terms of lattices in $\mathbb{R}^n$ (the study of such objects is called "geometry of numbers"). Rather than studying individual lattices one-by-one, it is fruitful to study all lattices at the same time. It turns out that this space allows the transitive action of a linear group. Hence tools from linear algebra can be applied. Moreover, this (non-compact) space has a finite invariant measure. Therefore, tools from ergodic theory kick in.

Towards the end of this lecture, we will provide an alternative proof of **BAD** having zero Lebesgue measure from this point of view.

2.2. Unimodular lattices in $\mathbb{R}^2$.

Definition 2.1. *A discrete subgroup $\Lambda \leq \mathbb{R}^2$ is said to be a **lattice** iff there exist two linearly independent vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^2$ such that $\Lambda = \mathbb{Z}\mathbf{v} + \mathbb{Z}\mathbf{w}$. The co-volume of a lattice, denoted as $\|\Lambda\|$, is defined to be $|\det(\mathbf{v}, \mathbf{w})| = \|\mathbf{v} \wedge \mathbf{w}\|$. A lattice is said to be **unimodular** iff its co-volume is equal to one. A vector $\mathbf{v}$ in a lattice Λ is said to be **primitive** iff $\mathbb{Q}\mathbf{v} \cap \Lambda = \mathbb{Z}\mathbf{v}$.*

Definition 2.2. *Let X_2 denote the set of all unimodular lattices in $\mathbb{R}^2$.*



Lemma 2.3. *Let Λ be a lattice of $\mathbb{R}^2$ and $F \subset \mathbb{R}^2$ be a Borel subset. If $F \cap (F + \mathbf{v}) = \emptyset$ for every nonzero $\mathbf{v} \in \Lambda$, then $\mathrm{Leb}(F) \leq \|\Lambda\|$. On the other hand, if $\mathbb{R}^2 = \bigcup_{\mathbf{v}\in\Lambda} F + \mathbf{v}$, then $\mathrm{Leb}(F) \geq \|\Lambda\|$.*

If both conditions are met, we call F a **strict fundamental domain** of Λ.

Proof. Note that there exists a strict fundamental domain F_0 for Λ with $\mathrm{Leb}(F_0) = \|\Lambda\|$. For instance, if $\Lambda = \mathbb{Z}\mathbf{v} + \mathbb{Z}\mathbf{w}$, then F_0 can be taken to be $[0,1)\mathbf{v} + [0,1)\mathbf{w}$. For another Borel measurable subset $F \subset \mathbb{R}^2$, let $F_{\mathbf{v}} := F \cap (F_0 - \mathbf{v})$ for $\mathbf{v} \in \Lambda$. Then $F = \bigsqcup_{\mathbf{v}\in\Lambda} F_{\mathbf{v}}$ and hence $\mathrm{Leb}(F) = \sum \mathrm{Leb}(F_{\mathbf{v}}) = \sum \mathrm{Leb}(F_{\mathbf{v}} + \mathbf{v})$.

First assume $F \cap (F+\mathbf{v}) = \emptyset$ for every nonzero $\mathbf{v} \in \Lambda$. Then $(F_{\mathbf{v}} + \mathbf{v})_{\mathbf{v}\in\Lambda}$ are disjoint from each other since $(F+\mathbf{v})$'s are. So

$$\mathrm{Leb}(F) = \sum_{\mathbf{v}\in\Lambda} \mathrm{Leb}(F_{\mathbf{v}} + \mathbf{v}) = \mathrm{Leb}(\bigcup_{\mathbf{v}\in\Lambda} F_{\mathbf{v}} + \mathbf{v}) \leq \mathrm{Leb}(F_0) = 1.$$

Next assume $\mathbb{R}^2 = \bigcup_{\mathbf{v}\in\Lambda} F + \mathbf{v}$. Then $\bigcup F_{\mathbf{v}} + \mathbf{v} = F_0$. Thus

$$\mathrm{Leb}(F) = \sum_{\mathbf{v}\in\Lambda} \mathrm{Leb}(F_{\mathbf{v}} + \mathbf{v}) \geq \mathrm{Leb}(F_0) = 1.$$

□

We equip X_2 with the following topology: A subset $U \subset \mathrm{X}_2$ is open iff for every $\Lambda \in U$, say $\Lambda = \mathbb{Z}\mathbf{v} + \mathbb{Z}\mathbf{w}$, there exists $\varepsilon > 0$ (dependent on the choice of $\mathbf{v}, \mathbf{w}$) such that every unimodular lattice $\Lambda' = \mathbb{Z}\mathbf{v}' + \mathbb{Z}\mathbf{w}'$ with $\|\mathbf{v} - \mathbf{v}'\| < \varepsilon$, $\|\mathbf{w} - \mathbf{w}'\| < \varepsilon$ belongs to U. Equivalently, we equip X_2 with the Chabauty topology.

Lemma 2.4. X_2 *is a separable metrizable space.*

There are different ways of showing X_2 is metrizable. For instance, one can show that for every $\Lambda = \mathbb{Z}\mathbf{v} + \mathbb{Z}\mathbf{w} \in \mathrm{X}_2$ and every $\varepsilon > 0$, there exists $\Lambda' = \mathbb{Z}\mathbf{v}' + \mathbb{Z}\mathbf{w}' \in \mathrm{X}_2$ with $\|\mathbf{v}' - \mathbf{v}\|, \|\mathbf{w}' - \mathbf{w}\| < \varepsilon$ and $\mathbf{v}', \mathbf{w}' \in \mathbb{Q}^2$. This would imply that X_2 is regular[2] and has a countable basis[3]. Then invoke Urysohn's metrization theorem.

Note that there exist distinct $\Lambda, \Lambda' \in \mathrm{X}_2$ such that for every $\varepsilon > 0$, there exist $\mathbf{v}, \mathbf{w}, \mathbf{w}, \mathbf{w}'$ with $\Lambda = \mathbb{Z}\mathbf{v} + \mathbb{Z}\mathbf{w}$ and $\mathbf{v}', \mathbf{w}' \in \Lambda'$ such that $\|\mathbf{v} - \mathbf{v}'\| < \varepsilon$, $\|\mathbf{w} - \mathbf{w}'\| < \varepsilon$. However, it is not clear to me whether one can further require $\mathbf{v}', \mathbf{w}'$ to form a $\mathbb{Z}$-basis of Λ'.

2.3. **Systole function and Mahler's criterion.**

Definition 2.5. *For a lattice* Λ, *let* $\mathrm{sys}(\Lambda) := \inf_{\mathbf{v}\neq\mathbf{0}\in\Lambda} \|\mathbf{v}\|$.

Theorem 2.6. $\mathrm{sys} : \mathrm{X}_2 \to \mathbb{R}_{>0}$ *is a bounded proper continuous function.*

We show the properness and boundedness of sys below. Continuity is left as an exercise.

Proof. It suffices to show that, given $c_0 > 0$, for every sequence $(\Lambda_n) \subset \mathrm{X}_2$ with $\mathrm{sys}(\Lambda_n) > c_0$ for all n, there exists a convergent subsequence. To prove this claim, it suffices, for every $\Lambda \in \mathrm{X}_2$ with $\mathrm{sys}(\Lambda) > c_0$, to find a constant $C > 1$ (depending on c_0 but not Λ) such that $\Lambda = \mathbb{Z}\mathbf{v} + \mathbb{Z}\mathbf{w}$ for some $\|\mathbf{v}\|, \|\mathbf{w}\| < C$.

Now fix such a c_0 and Λ. Let $\mathbf{v}_0 \in \Lambda$ be such that

$$\|\mathbf{v}_0\| = \inf \{\|\mathbf{v}\| \mid \mathbf{v} \in \Lambda \setminus \{\mathbf{0}\}\}.$$

Let $\mathbf{w}_0 \in \Lambda$[4] be such that

$$\mathrm{dist}(\mathbf{w}_0, \mathbb{R}\mathbf{v}_0) = \inf \{\mathrm{dist}(\mathbf{w}, \mathbb{R}\mathbf{v}_0) \mid \mathbf{w} \in \Lambda \setminus \mathbb{R}\mathbf{v}_0\}.$$

We claim that $\Lambda = \mathbb{Z}\mathbf{v}_0 + \mathbb{Z}\mathbf{w}_0$. Indeed, if $\Lambda = \mathbb{Z}\mathbf{v}_1 + \mathbb{Z}\mathbf{w}_1$ then $\mathbf{v}_0 = a\mathbf{v}_1 + b\mathbf{w}_1$ with $\gcd(a, b) = 1$. If $ad + bc = 1$ for some $c, d \in \mathbb{Z}$, then $\mathbf{w}_0' := (-c, d)^{\mathrm{tr}} \in \Lambda$ satisfies $\Lambda = \mathbb{Z}\mathbf{v}_0 + \mathbb{Z}\mathbf{w}_0'$. Write $\mathbf{w}_0 = e\mathbf{v}_0 + f\mathbf{w}_0'$, then $f = \pm 1$ as $\mathbf{w}_0$ minimizes the distance to $\mathbb{R}\mathbf{v}_0$. So $\mathbf{w}_0'$ can be written as integral combinations of $\mathbf{v}_0$ and $\mathbf{w}_0$. Thus $\Lambda = \mathbb{Z}\mathbf{v}_0 + \mathbb{Z}\mathbf{w}_0$.

It remains to give an upper bound on $\|\mathbf{v}_0\|$ and $\mathrm{dist}(\mathbf{w}_0, \mathbb{R}\mathbf{v}_0)$ in terms of c_0 (replacing $\mathbf{w}_0$ by $\mathbf{w}_0 - n\mathbf{v}_0$ for suitable n_0 would give an upper bound for $\|\mathbf{w}_0\|$).

Let $F := [0, 2) \times [0, 2)$. Then $\mathrm{Leb}(F) > 1$. By Lemma 2.3, for some non-zero $\mathbf{v} \in \Lambda$, $\mathbf{v} + F \cap F \neq \emptyset$. In other words, $\mathbf{v} = \mathbf{x}_1 - \mathbf{x}_2$ for some $\mathbf{x}_i \in F$. Thus

$$\|\mathbf{v}_0\| \leq \|\mathbf{v}\| \leq 2\sqrt{2}.$$

Also note that $\Lambda \cap \mathbb{R}\mathbf{v}_0 = \mathbb{Z}\mathbf{v}_0$.

[2]For every x and every neighborhood $\mathcal{N}$ of x, there exists a smaller one whose closure is contained in $\mathcal{N}$.

[3]Countably many open subsets that 1. cover X_2, and 2.any intersection of two containing some x contains a third one containing the same x.

[4]As remarked by H.Li, one can simply take $\mathbf{w}_0$ to be any vector such that $\Lambda = \mathbb{Z}\mathbf{v}_0 + \mathbb{Z}\mathbf{w}_0$. As the covolume of Λ is one, the distance from $\mathbf{w}_0$ to $\mathbb{R}\mathbf{v}_0$ must be bounded from above. This gives a shorter proof. But the proof here generalizes.

Pick some unit vector $\mathbf{y}_0$ orthogonal to $\mathbf{v}_0$. Let $F' := (0,1)\mathbf{v}_0 + (0,C)\mathbf{y}_0$ with $C := \frac{2}{c_0}$. Then

$$\mathrm{Leb}(F') = \frac{2\,\|\mathbf{v}_0\|}{c_0} > 1.$$

Thus there exists some non-zero $\mathbf{w} \in (F' - F') \cap \Lambda$. We can write $\mathbf{w} = w_1\mathbf{v}_0 + w_2\mathbf{y}_0$ for some $w_1 \in (-1,1)$ and $w_2 \in (-C,C)$. If $w_2 = 0$, then w_1 has to be integral, so is also 0. This contradicts against the fact that $\mathbf{w}$ is nonzero. So $\mathbf{w} \notin \mathbb{R}\mathbf{v}_0$ and

$$\mathrm{dist}(\mathbf{w}_0, \mathbb{R}\mathbf{v}_0) \leq \mathrm{dist}(\mathbf{w}, \mathbb{R}\mathbf{v}_0) = |w_2| \leq C = \frac{2}{c_0}.$$

The proof is complete. $\square$

Corollary 2.7. X_2 *is non-compact.*

A subset B of X_2 is said to be **bounded** iff there exists $c > 0$ such that $\mathrm{sys}(\Lambda) > c$ for every $\Lambda \in B$. Otherwise B is **unbounded**. A subset of X_2 is bounded iff it is precompact by Mahler's criterion.

A sequence $(x_n)_{n\in\mathbb{Z}^+}$ (or a subset indexed by positive real numbers $(x_t)_{t\in\mathbb{R}^+}$) in a topological space X is said to be **divergent** iff $\lim_{n\to+\infty} \mathrm{sys}(x_n) = 0$ (resp. $\lim_{t\to+\infty} = 0$). By Mahler's criterion, $(x_n)_{n\in\mathbb{Z}^+}$ (resp. $(x_t)_{t\in\mathbb{R}^+}$) is divergent iff for any compact subset $C \subset X$, there exists N such that for every $n > N$ (resp. every $t > N$), $x_n \notin C$ (resp. $x_t \notin C$).

2.4. **Group action.** The set $\mathbf{SL}_2(\mathbb{R}) := \{\text{2-by-2 real matrices with determinant 1}\}$ is naturally a topological space (subspace topology from $\mathbb{R}^4$) as well as a group (matrix multiplication). It is a **topological group** as

$$\begin{aligned} \mathbf{SL}_2(\mathbb{R}) \times \mathbf{SL}_2(\mathbb{R}) &\to \mathbf{SL}_2(\mathbb{R}) \\ (g,h) &\mapsto gh \end{aligned}$$

and $g \mapsto g^{-1}$ from $\mathbf{SL}_2(\mathbb{R})$ to itself are continuous.

The group $\mathbf{SL}_2(\mathbb{R})$ acts on X_2 by $(g,\Lambda) \mapsto g\Lambda := \{g\mathbf{v},\ \mathbf{v} \in \Lambda\}$. The action is continuous in the sense that

$$\begin{aligned} \mathbf{SL}_2(\mathbb{R}) \times \mathrm{X}_2 &\to \mathrm{X}_2 \\ (g,\Lambda) &\mapsto g\Lambda \end{aligned}$$

is continuous.

Lemma 2.8. *The map $g \mapsto g.\mathbb{Z}^2$ from $\mathbf{SL}_2(\mathbb{R})$ to X_2 is continuous and open. Moreover, it factors through a homeomorphism $\mathbf{SL}_2(\mathbb{R})/\mathbf{SL}_2(\mathbb{Z}) \to \mathrm{X}_2$.*

There are a few subgroups of $\mathbf{SL}_2(\mathbb{R})$ that we are particularly interested in. First,

$$\mathrm{U}^+ := \left\{ \mathbf{u}_t^+ := \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \,\middle|\, t \in \mathbb{R} \right\}$$

is a one-parameter subgroup (that is, $t \mapsto \mathbf{u}_t^+$ from $(\mathbb{R},+)$ to U^+ gives an isomorphism of topological groups) consisting of unipotent matrices. Its action on X_2 is sometimes referred as a horocycle/unipotent flow. Also,

$$\mathrm{A} := \left\{ \mathbf{a}_t := \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix} \,\middle|\, t \in \mathbb{R} \right\}$$

is a one-parameter subgroup consisting of diagonal matrices. Its action on X_2 is sometimes called a geodesic/diagonal flow.

2.4.1. *An explicit metric on* X_2. Once we realize X_2 as a homogeneous space, we can equip it with a metric as follows.

For $A \in \mathbf{SL}_2(\mathbb{R})$, let $\|A\|_{\mathrm{op}}$ denote the operator norm w.r.t. Euclidean norm:

$$\|A\|_{\mathrm{op}} := \sup_{\mathbf{v}\neq 0\in\mathbb{R}^2} \frac{\|A.\mathbf{v}\|}{\|\mathbf{v}\|}.$$

Define a metric on $\mathbf{SL}_2(\mathbb{R})$ by

$$\mathrm{dist}(g,h) := \log\left\{1 + \left\|gh^{-1} - \mathrm{I}_2\right\|_{\mathrm{op}} + \left\|hg^{-1} - \mathrm{I}_2\right\|_{\mathrm{op}}\right\}$$

Once can verify that $\operatorname{dist}(g\gamma, h\gamma)$ for every $\gamma \in \mathbf{SL}_2(\mathbb{R})$. Then,

$$\operatorname{dist}(g\mathbb{Z}^2, h\mathbb{Z}^2) := \inf \{\operatorname{dist}(g, h\gamma) \mid \gamma \in \mathbf{SL}_2(\mathbb{Z})\}$$

define a metric on $\mathbf{SL}_2(\mathbb{R})/\,\mathbf{SL}_2(\mathbb{Z})$.

2.5. **Dani correspondence.** For a real number α, let

$$\Lambda_\alpha := \mathbb{Z}\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \mathbb{Z}\begin{bmatrix} \alpha \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix}\mathbb{Z}^2 = \mathbf{u}_\alpha^+\mathbb{Z}^2,$$

a unimodular lattice (in X_2).

Lemma 2.9 (Dani correspondence)**.** *A real number α is badly approximable iff $(\mathbf{a}_t\Lambda_\alpha)_{t>0}$ is bounded in X_2.*

Note that for every α, the full orbit $(\mathbf{a}_t\Lambda_\alpha)_{t\in\mathbb{R}}$ is unbounded. Actually, $(\mathbf{a}_t\Lambda_\alpha)$ as $t \to -\infty$ diverges.

Proof. For every $(x, y)^{\mathrm{tr}} \in \mathbf{a}_t\Lambda_\alpha$, there exists $(m, n) \in \mathbb{Z}^2$ such that

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e^t(m + n\alpha) \\ e^{-t}n \end{bmatrix}.$$

So $\operatorname{sys}(\mathbf{a}_t\Lambda_\alpha) \geq \varepsilon$ for some $\varepsilon > 0$ iff for some $\varepsilon' > 0$, for every $(m, n) \in \mathbb{Z}^2 \setminus \{\mathbf{0}\}$ and $t > 0$,

$$e^t\,|m + n\alpha| \geq \varepsilon',\ \ e^{-t}\,|n| \geq \varepsilon'. \tag{7}$$

Assume α is bad, namely, there exists $c_0 \in (0, 1)$ such that for every $(p, q) \in \mathbb{Z}^2$ with $q \neq 0$, $|q|\,|p + q\alpha| > c_0$. So for $\mathbf{0} \neq (x, y)^{\mathrm{tr}} = (e^t(m + n\alpha), e^{-t}n)^{\mathrm{tr}} \in \mathbf{a}_t\Lambda_\alpha$, if $y \neq 0$, then

$$|xy| = |n||m + n\alpha| > c_0,\ \text{implying}\ (x^2 + y^2)^{\frac{1}{2}} \geq \sqrt{2|xy|} > \sqrt{2c_0}.$$

If $y = 0$, then $(x, y)^{\mathrm{tr}} = (e^t m, 0)^{\mathrm{tr}}$. Hence $\|(x, y)^{\mathrm{tr}}\| \geq 1$. Anyway, we have shown that every non-zero vector of $\mathbf{a}_t\Lambda_\alpha$ has norm at least $\sqrt{c_0}$.

Conversely, suppose $\operatorname{sys}(\mathbf{a}_t\Lambda_\alpha) > c_1 > 0$ for all $t > 0$. For every $(p, q) \in \mathbb{Z}^2$ with $q > 0$, take $t_q > 0$ such that $e^{t_q} = \frac{2q}{c_1}$. Then $\|(e^{t_q}(p + q\alpha), e^{-t_q}q)\| \geq c_1$. But $|e^{-t_q}q|^2 = \frac{c_1^2}{4}$, so

$$|e^{t_q}(p + q\alpha)| \geq \frac{\sqrt{3}}{2}c_1 \implies q\,|p + q\alpha| \geq \frac{\sqrt{3}}{4}c_1^2.$$

The proof is now complete. $\square$

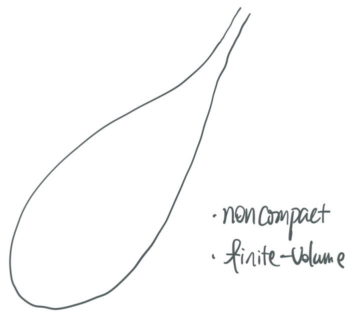
2.6. **Invariant measures on X_2.**

Definition 2.10. *Let $f : X \to Y$ be a Borel measurable map between two topological spaces X and Y. Given a measure μ on $(X, \mathscr{B}_X)$ ($\mathscr{B}_X$ is the Borel σ-algebra, the smallest σ-algebra containing all open subsets), we define $f_*\mu$ to be the measure on Y satisfying $f_*\mu(E) := \mu(f^{-1}(E))$ for every $E \in \mathscr{B}_Y$. If $X = Y$ and $f_*\mu = \mu$, we say that f preserves the measure μ. If G is a group acting on X by Borel measurable maps such that $g_*\mu = \mu$ for every $g \in G$, then we say that μ is G-invariant.*

Lemma 2.11. *There exists a locally finite $\mathbf{SL}_2(\mathbb{R})$-invariant measure $\mathrm{m}_{\mathrm{X}_2}$ on X_2.*

There are different ways to see the existence of $\mathrm{m}_{\mathrm{X}_2}$. For instance, one may equip $\mathbf{SL}_2(\mathbb{R})$ with a right invariant Riemannian metric and then $\mathbf{SL}_2(\mathbb{R})/\,\mathbf{SL}_2(\mathbb{Z})$ will inherit a Riemannian metric. One can check that the volume form induced from such a metric is $\mathbf{SL}_2(\mathbb{R})$-invariant. We will give an explicit construction of an invariant measure on $\mathbf{SL}_2(\mathbb{R})$ and then induce one on the quotient space in the next subsection. What is less trivial is that:

Theorem 2.12. $\mathrm{m}_{\mathrm{X}_2}$ *is a finite measure.*



Henceforth, we normalize $\mathrm{m}_{\mathrm{X}_2}$ to be a **probability measure**, namely, $\mathrm{m}_{\mathrm{X}_2}(\mathrm{X}_2) = 1$.

2.7. **A construction of the invariant measure.**

2.7.1. *Explicit construction of invariant measures on* $\mathbf{SL}_2(\mathbb{R})$*.* Let

$$\mathcal{O}_1 := \left\{ \begin{bmatrix} x & y \\ z & w \end{bmatrix} \in \mathbf{SL}_2(\mathbb{R}) \,\middle|\, x \neq 0 \right\}, \quad \mathcal{O}_2 := \left\{ \begin{bmatrix} x & y \\ z & w \end{bmatrix} \in \mathbf{SL}_2(\mathbb{R}) \,\middle|\, z \neq 0 \right\}$$

be two open subsets covering $\mathbf{SL}_2(\mathbb{R})$, each of which can be parametrized by domains in Euclidean spaces:

$$\mathcal{O}'_1 := \left\{ (x,y,z) \in \mathbb{R}^3 \,\middle|\, x \neq 0 \right\} \xrightarrow[\simeq]{\varphi_1} \mathcal{O}_1$$

$$\mathcal{O}'_2 := \left\{ (x,z,w) \in \mathbb{R}^3 \,\middle|\, z \neq 0 \right\} \xrightarrow[\simeq]{\varphi_2} \mathcal{O}_2$$

where

$$\varphi_1(x,y,z) := \begin{bmatrix} x & y \\ z & \frac{1+yz}{x} \end{bmatrix}, \ \varphi_2(x,z,w) := \begin{bmatrix} x & \frac{xw-1}{z} \\ z & w \end{bmatrix}$$

are homeomorphisms.

Lemma 2.13. *The map* $\varphi_{12}(x,y,z) := (x, z, \frac{1+yz}{x})$ *from* $\{(x,y,z) \in \mathcal{O}_1,\ z \neq 0\}$ *to* $\{(x,z,w) \in \mathcal{O}_2,\ x \neq 0\}$ *sends* $(\varphi_{12})_* \left|\frac{\mathrm{dxdydz}}{x}\right| = \left|\frac{\mathrm{dxdzdw}}{z}\right|$*. Therefore*

$$(\varphi_1)_* \left|\frac{\mathrm{dxdydz}}{x}\right| \quad \textit{and} \quad (\varphi_2)_* \left|\frac{\mathrm{dxdzdw}}{z}\right|$$

glue to a locally finite measure on $\mathbf{SL}_2(\mathbb{R})$*. Also,* $\{(x,y,z) \in \mathcal{O}_1,\ z = 0\}$ *has measure zero under* $\left|\frac{\mathrm{dxdydz}}{x}\right|$*. Similarly* $\{(x,z,w) \in \mathcal{O}_2,\ x = 0\}$ *has measure zero under* $\left|\frac{\mathrm{dxdzdw}}{z}\right|$*.*

Proof. Direct calculation. Note that by differentiating $xw - yz = 1$, one obtains $w\mathrm{dx} + x\mathrm{dw} = y\mathrm{dz} + z\mathrm{dy}$. □

Let $\mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}$ denote this measure.

2.7.2. *Invariance property.* Define

$$\mathrm{U}^+ := \left\{ \mathbf{u}^+_t := \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \,\middle|\, t \in \mathbb{R} \right\}.$$

and

$$\mathrm{U}^- := \left\{ \mathbf{u}^-_t := \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix} \,\middle|\, t \in \mathbb{R} \right\}.$$

Lemma 2.14. $\mathbf{SL}_2(\mathbb{R})$ *is generated by the two subgroups* U^+ *and* U^-*.*

Proof. Left as exercise. □

By restricting to $\mathcal{O}_1$ or $\mathcal{O}_2$ respectively, it is easy to verify that

Lemma 2.15. $\mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}$ *is invariant under the left multiplication by* $\mathbf{SL}_2(\mathbb{R})$*.*

By similar reasoning[5], using additionally

$$\mathcal{O}_3 := \left\{ \begin{bmatrix} x & y \\ z & w \end{bmatrix} \in \mathbf{SL}_2(\mathbb{R}) \;\middle|\; y \neq 0 \right\},$$

one can show that

Lemma 2.16. $\mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}$ *is invariant under the right multiplication by* $\mathbf{SL}_2(\mathbb{R})$*.*

2.7.3. *Strict fundamental domain.* A Borel subset $\mathcal{F} \subset \mathbf{SL}_2(\mathbb{R})$ is said to be a **strict fundamental domain** for $\mathbf{SL}_2(\mathbb{Z})$ iff

$$\mathbf{SL}_2(\mathbb{R}) = \bigsqcup_{\gamma \in \mathbf{SL}_2(\mathbb{Z})} \mathcal{F} \cdot \gamma.$$

Lemma 2.17. *Strict fundamental domains exist.*

Proof. First we choose a small open neighborhood $\mathcal{N}$ of identity in $\mathbf{SL}_2(\mathbb{R})$ such that $\mathcal{N}\gamma \cap \mathcal{N} = \emptyset$ for every non-identity element γ in $\mathbf{SL}_2(\mathbb{Z})$. Choose a sequence $(g_n) \subset \mathbf{SL}_2(\mathbb{R})$ such that

$$\mathbf{SL}_2(\mathbb{R}) = \bigcup g_n \cdot \mathcal{N}.$$

Then we define

$$\begin{aligned} &V_1 := g_1\mathcal{N} \\ &V_2 := g_2\mathcal{N} \setminus g_1\mathcal{N}\Gamma \\ &V_3 := g_3\mathcal{N} \setminus (g_1\mathcal{N}\Gamma \cup g_2\mathcal{N}\Gamma) \\ &\quad .... \end{aligned}$$

From the definition, V_2 lives in the complement of $V_1\Gamma$, V_3 lives in the complement of $(V_1 \cup V_2)\Gamma$.... Therefore, $V_i \cap V_j\gamma = \emptyset$ for every $i \neq j$ and $\gamma \in \Gamma$. Moreover, by the choice of $\mathcal{N}$, $V_i \cap V_i\gamma = \emptyset$ for non-identity $\gamma \in \mathbf{SL}_2(\mathbb{Z})$. Thus if we let

$$\mathcal{F} := \bigcup_{i=1}^{\infty} V_i,$$

then $\mathcal{F} \cap \mathcal{F}\gamma = \emptyset$ for every $\gamma_{\neq \mathrm{id}} \in \mathbf{SL}_2(\mathbb{Z})$. On the other hand, for $g \in \mathbf{SL}_2(\mathbb{R})$, if n_g is the smallest positive integer n such that $g \in g_n\mathcal{N}\Gamma$, then $g \in V_{n_g}\Gamma \subset \mathcal{F}\Gamma$ by the definition of V_n's. The proof is thus complete. $\square$

2.7.4. *The invariant measure on the quotient.* Fix some strict fundamental domain $\mathcal{F}$ and let $\mathrm{m}_{\mathcal{F}}$ be the restriction of $\mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}$ to $\mathcal{F}$. Let $\pi : \mathbf{SL}_2(\mathbb{R}) \to \mathbf{SL}_2(\mathbb{R})/\mathbf{SL}_2(\mathbb{Z})$ be the natural quotient and $\pi_{\mathcal{F}}$ denote the induced bijection $\mathcal{F} \to \mathbf{SL}_2(\mathbb{R})/\mathbf{SL}_2(\mathbb{Z})$. Let $\mathrm{m}_{[\mathcal{F}]} := (\pi_{\mathcal{F}})_*\mathrm{m}_{\mathcal{F}}$.

Lemma 2.18. *If* $\mathcal{O} \subset \mathbf{SL}_2(\mathbb{R})$ *is such that* π *restricted to* $\mathcal{O}$ *is injective, then*

$$(\pi_{\mathcal{O}})_* \left(\mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}|_{\mathcal{O}}\right) = \mathrm{m}_{[\mathcal{F}]}|_{\pi(\mathcal{O})}.$$

Consequently, $\mathrm{m}_{[\mathcal{F}]}$ *is independent of the choice of strict fundamental domains and* $\mathrm{m}_{[\mathcal{F}]}$ *is invariant under the left action of* $\mathbf{SL}_2(\mathbb{R})$*.*

Proof. It suffices to show $\mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}(\mathcal{O}) = \mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}(\pi_{\mathcal{F}}^{-1}(\pi(\mathcal{O})))$ for every such $\mathcal{O}$ as in the statement.

For every $\gamma \in \mathbf{SL}_2(\mathbb{Z})$, let

$$\mathcal{O}_\gamma := \{x \in \mathcal{O} \mid x\gamma \in \mathcal{F}\}.$$

By assumption, elements from $(\mathcal{O}_\gamma)_{\gamma \in \mathbf{SL}_2(\mathbb{Z})}$ (resp. $(\mathcal{O}_\gamma \cdot \gamma)_{\gamma \in \mathbf{SL}_2(\mathbb{Z})}$) are disjoint from each other. Hence

$$\mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}(\mathcal{O}) = \mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}(\cup \mathcal{O}_\gamma) = \mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}(\cup \mathcal{O}_\gamma \cdot \gamma) = \mathrm{m}_{\mathcal{F}}(\mathcal{O}).$$

$\square$

This finishes the proof of Lemma 2.11. The local finiteness follows from the lemma above and the fact that $\mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}$ is locally finite.

[5] Alternatively, as remarked by H.Li, one can verify the invariance of measure under the transpose map on $\mathcal{O}_1$. Right invariance then follows from the left invariance.

2.8. **Ergodicity and mixing.**

Definition 2.19. *The action of* $\mathrm{A} \curvearrowright (\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$ *is said to be*

- ***ergodic*** *iff for every Borel subset* $B \subset \mathrm{X}_2$ *that is* A*-invariant (i.e.,* $a.B = B$ *for every* $a \in \mathrm{A}$*), one has* $\mathrm{m}_{\mathrm{X}_2}(B) = 0$ *or* $\mathrm{m}_{\mathrm{X}_2}(\mathrm{X}_2 \setminus B) = 0$*;*
- ***mixing*** *iff for every divergent sequence* $(a_n) \in \mathrm{A}$ *and Borel subsets* B, C*, one has*

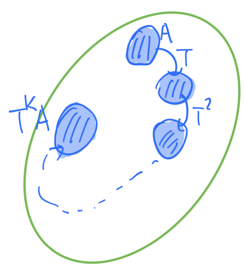
$$\lim_{n\to\infty} \mathrm{m}_{\mathrm{X}_2}(B \cap a_n^{-1}.C) = \mathrm{m}_{\mathrm{X}_2}(B)\mathrm{m}_{\mathrm{X}_2}(C).$$

Lemma 2.20. *Mixing implies ergodicity.*

Proof. Indeed, let B be an A-invariant subset and let (a_n) be a divergent sequence in A. Then by mixing,

$$\lim_{n\to\infty} \mathrm{m}_{\mathrm{X}_2}(B \cap a_n^{-1}.B) = \mathrm{m}_{\mathrm{X}_2}(B)^2.$$

By A-invariance, the left hand side is $\mathrm{m}_{\mathrm{X}_2}(B)$. Then $\mathrm{m}_{\mathrm{X}_2}(B)^2 = \mathrm{m}_{\mathrm{X}_2}(B)$ implies $\mathrm{m}_{\mathrm{X}_2}(B) = 0$ or 1. So we are done. □



Ergodicity means that every set will eventually visit "everywhere" in the space

Mixing is the stronger claim that TᵏA "equidistributes" in the space.

We are going to prove that the A-action on X_2 is mixing via a little functional analysis.

Theorem 2.21. *The action of* $\mathrm{A} \curvearrowright (\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$ *is mixing.*

Remark 2.22. *Since we have not proved that* $\mathrm{m}_{\mathrm{X}_2}$ *is finite, in the possible case when* $|\mathrm{m}_{\mathrm{X}_2}|$ *is infinite, the proof below shows that* $\lim_{t\to\infty}\langle U_{\mathbf{a}_t}\phi, \psi\rangle = 0$ *for any two* L^2 *functions* ϕ, ψ*.*

2.9. **The associated unitary representation.** Let

$$\mathrm{L}^2(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2}) := \left\{ f : \mathrm{X}_2 \to \mathbb{C} \text{ measurable} \;\middle|\; \int |f|^2 \mathrm{m}_{\mathrm{X}_2} < +\infty \right\};$$
$$\mathrm{L}^2_0(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2}) := \left\{ f \in \mathrm{L}^2(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2}) \;\middle|\; \int f \mathrm{m}_{\mathrm{X}_2} = 0 \right\}$$

(note that L^2 functions are in L^1 since $\mathrm{m}_{\mathrm{X}_2}$ is finite) with inner product denoted by

$$\langle f, g\rangle := \int_{\mathrm{X}_2} f(x)g(x)\mathrm{m}_{\mathrm{X}_2}(x).$$

Also, $\|f\|_2 := \sqrt{\langle f, f\rangle}$.

As usual, we identify two functions $f, g \in \mathrm{L}^2_0(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$ if they are equal almost surely. Then $\mathrm{L}^2_0(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$ with this inner product is a separable (i.e., has a countable dense subset) real Hilbert space.

Note that the $\mathbf{SL}_2(\mathbb{R})$ action on $(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$ induces an action of $\mathbf{SL}_2(\mathbb{R})$ on $\mathrm{L}^2_0(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$ defined by

$$U_g f(x) := f(g^{-1}x).$$

Lemma 2.23. *The action has the following properties:*

1. *For each* $g \in \mathbf{SL}_2(\mathbb{R})$*,* $U_g : \mathrm{L}^2_0(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2}) \to \mathrm{L}^2_0(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$ *is a unitary operator;*

2. *For every $\varepsilon > 0$ and $f \in \mathrm{L}^2_0(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$, there exists a neighborhood $\mathcal{O}_\varepsilon$ of the identity matrix in $\mathbf{SL}_2(\mathbb{R})$ such that for every $g \in \mathcal{O}_\varepsilon$,*

$$\|U_g f - f\|_2 \leq \varepsilon.$$

Proof. Take $g \in \mathbf{SL}_2(\mathbb{R})$. Since the action of g preserves $\mathrm{m}_{\mathrm{X}_2}$, we have $\int f(gx)\mathrm{m}_{\mathrm{X}_2}(x) = \int f(x)\mathrm{m}_{\mathrm{X}_2}(x)$ for every integrable function f. For $\phi \in \mathrm{L}^2(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$, by applying this equality to $f = |\phi|^2$, we see that $\|U_g\phi\|_2 = \|\phi\|_2$.

For the second part, note that the set $C_c(\mathrm{X}_2)$ of compactly supported functions are dense in $\mathrm{L}^2(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$ (for instance, see [Rud87, Theorem 3.14]). For every $\varepsilon > 0$ and $f \in \mathrm{L}^2(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$, find $\phi \in C_c(\mathrm{X}_2)$ such that $\|\phi - f\|_2 \leq 0.1\varepsilon$. Since ϕ is uniformly continuous, find $\delta > 0$ such that $d(x, y) < \delta \implies |\phi(x) - \phi(y)| < 0.1\varepsilon$. Fixing a relatively compact neighborhood of identity $\mathcal{O}_0$, then $\overline{\mathcal{O}_0^{-1}.\operatorname{supp}(\phi)}$ is still compact. Thus, we can find $\mathcal{O}_\varepsilon \subset \mathcal{O}_0$, a neighborhood of the identity, such that for every $g \in \mathcal{O}_\varepsilon$ and $x \in C := \overline{\mathcal{O}_0^{-1}.\operatorname{supp}(\phi)} \cup \operatorname{supp}(\phi)$,

$$|\phi(gx) - \phi(x)| < 0.1\varepsilon.$$

Consequently,

$$\int_{\mathrm{X}_2} |\phi(gx) - \phi(x)|^2 \, \mathrm{m}_{\mathrm{X}_2}(x) = \int_C |\phi(gx) - \phi(x)|^2 \, \mathrm{m}_{\mathrm{X}_2}(x) \leq (0.1\varepsilon)^2,$$

which implies $\|U_g\phi - \phi\|_2 < 0.1\varepsilon$. Therefore, for $g \in \mathcal{O}_\varepsilon$

$$\|U_g f - f\|_2 \leq \|U_g\phi - \phi\|_2 + \|U_g\phi - U_g f\| + \|\phi - f\| \leq 0.1\varepsilon + 0.1\varepsilon + 0.1\varepsilon < \varepsilon.$$

So the proof completes. $\square$

2.10. **Mixing of the geodesic flow.** In this subsection we prove Theorem 2.21. We need to show that for $\phi, \psi \in \mathrm{L}^2_0(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$ and a divergence sequence $(a_n) \in \mathrm{A}^+$ (namely, assume the $(1,1)$ entries of matrices a_n diverge to $+\infty$. The other case when they diverge to $-\infty$ is similar), one has

$$\lim_{n\to\infty} \langle U_{a_n}\phi, \psi\rangle = 0.$$

For simplicity write $\mathrm{L}^2_0 := \mathrm{L}^2_0(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$.

2.10.1. *The basics.* As L^2_0 is separable, by applying a diagonal argument, we assume that

$$\lim_{n\to\infty} \langle U_{a_{n_k}}\phi, \psi\rangle \text{ exists}, \ \forall\, \phi, \psi \in \mathrm{L}^2_0$$

for some subsequence n_k. It suffices to show that for this subsequence, the limit above is zero for every $\phi, \psi \in \mathrm{L}^2_0$.

For each $\phi \in \mathrm{L}^2_0$, the map $\psi \mapsto \lim_{k\to\infty}\langle U_{a_{n_k}}\phi, \psi\rangle$ is linear and bounded since

$$\left|\lim_{k\to\infty} \langle U_{a_{n_k}}\phi, \psi\rangle\right| = \lim_{k\to\infty} \left|\langle U_{a_{n_k}}\phi, \psi\rangle\right| \leq \|\phi\|_2 \|\psi\|_2 .$$

By Riesz's lemma, there exists some element in L^2_0, denoted as $E(\phi)$, such that $\langle E(\phi), \psi\rangle = \lim_{k\to\infty}\langle U_{a_{n_k}}\phi, \psi\rangle$ for every $\psi \in \mathrm{L}^2_0$.

Next we note that $\phi \mapsto E(\phi)$ is a bounded operator. Linearity is clear. To show that it is bounded, apply the computation above to $\psi := E(\phi)$,

$$\|E(\phi)\|_2^2 = \langle E(\phi), E(\phi)\rangle \leq \|\phi\|_2 \|E(\phi)\|_2 \implies \|E(\phi)\|_2 \leq \|\phi\|_2 .$$

Let E^* be the adjoint operator of E, then

$$\langle E^*\phi, \psi\rangle = \lim_{k\to\infty} \langle U_{a_{n_k}^{-1}}\phi, \psi\rangle.$$

2.10.2. *Almost invariant functions are constants.* Next we are going to show that the image of E is pointwisely fixed by $\mathbf{SL}_2(\mathbb{R})$ and is hence zero by the following lemma

Lemma 2.24. *Let $f \in \mathrm{L}^2_0$. If for every $g \in \mathbf{SL}_2(\mathbb{R})$, $f(g.x) = f(x)$ for almost every $x \in \mathrm{X}_2$, then f is a constant function a.e.*

Proof. Consider the set

$$F := \{(g,x) \in \mathbf{SL}_2(\mathbb{R}) \mid f(g.x) \neq f(x)\}\,.$$

By Fubini theorem, $\mathrm{m}_{\mathrm{X}_2}(F) = 0$. Let $F_x := \{g \in \mathbf{SL}_2(\mathbb{R}) \mid f(g.x) \neq f(x)\}$. Apply Fubini again

$$\mathrm{m}_{\mathrm{X}_2}(F) = \int_{x \in \mathrm{X}_2} \mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}(F_x)\, \mathrm{m}_{\mathrm{X}_2}(x)$$

So there exists $x_0 \in \mathrm{X}_2$ such that $f(g.x_0) = f(x_0)$ for almost all $g \in \mathbf{SL}_2(\mathbb{R})$. Thus f is equal to $f(x_0)$ a.e. □

Recall from Lemma 2.14 that $\mathbf{SL}_2(\mathbb{R})$ is generated by U^+ and U^-. Therefore one only needs to show the invariance by

$$\mathrm{U}^+ := \left\{ \mathbf{u}_t^+ := \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \,\middle|\, t \in \mathbb{R} \right\}.$$

and

$$\mathrm{U}^- := \left\{ \mathbf{u}_t^- := \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix} \,\middle|\, t \in \mathbb{R} \right\}.$$

2.10.3. *The easy part.* For simplicity we assume $n_k = k$. Elements from U^+ and U^- enjoy the following properties

$$\lim_{n\to\infty} a_n^{-1} u a_n = \mathrm{I}_2,\ \forall\, u \in \mathrm{U}^+; \quad \lim_{n\to\infty} a_n v a_n^{-1} = \mathrm{I}_2,\ \forall\, v \in \mathrm{U}^-.$$

Combined with Lemma 2.23, for an element $u \in \mathrm{U}^+$, one gets

$$\begin{aligned} \langle U_u E(\phi), \psi \rangle &= \lim_{n\to\infty} \langle U_u U_{a_n} \phi, \psi \rangle = \lim_{n\to\infty} \langle U_{a_n^{-1} u a_n} \phi, U_{a_n^{-1}} \psi \rangle \\ &= \lim_{n\to\infty} \langle \phi, U_{a_n^{-1}} \psi \rangle = \langle E(\phi), \psi \rangle. \end{aligned}$$

for all $\phi, \psi \in \mathrm{L}^2_0$. Thus $U_u \circ E = E$.

Similarly, for $v \in \mathrm{U}^-$, one has $U_v \circ E^* = E^*$.

2.10.4. *The trick.*

Lemma 2.25. *Let E, E^* be as above. Then* $\ker(E) = \ker(E^*)$.

Note that in general the kernel of a linear operator is not the same as its adjoint.

Proof. It suffices to note that

$$\begin{aligned} \langle E(\phi), E(\phi) \rangle &= \lim_{n\to\infty} \langle U_{a_n} \phi, E(\phi) \rangle \\ &= \lim_{n\to\infty} \lim_{m\to\infty} \langle U_{a_n} \phi, U_{a_m} \phi \rangle \\ &= \lim_{n\to\infty} \lim_{m\to\infty} \langle U_{a_m^{-1}} \phi, U_{a_n^{-1}} \phi \rangle = \langle E^*(\phi), E^*(\phi) \rangle. \end{aligned}$$

□

By results from the last subsubsection, $E \circ (U_v - \mathrm{I}_2) = 0$ for every $v \in \mathrm{U}^-$. The lemma then implies that $E^* \circ (U_v - \mathrm{I}_2) = 0$. Taking the adjoint, we get $U_v \circ E = E$. The proof of Theorem 2.21 is now complete.

2.11. **Another proof of Leb(BAD) being zero.** Here we give an alternative proof of the fact that the set of badly approximable numbers has Lebesgue measure zero. We assume $\mathrm{Leb}(\mathbf{BAD}) > 0$ and derive a contradiction.

We fix some $\varepsilon > 0$ and let

$$\begin{aligned} \mathcal{O}_\varepsilon &:= \left\{ \begin{bmatrix} e^s & 0 \\ 0 & e^{-s} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix} .\mathbb{Z}^2 \,\middle|\, s, t \in (-\varepsilon, \varepsilon),\ r \in (0,1] \right\} \\ &= \mathbf{a}_{(-\varepsilon,\varepsilon)} \mathbf{u}^-_{(-\varepsilon,\varepsilon)} \mathbf{u}^+_{(0,1]} .\mathbb{Z}^2. \end{aligned}$$

Let $\mathrm{Obt} : \mathbf{SL}_2(\mathbb{R}) \to \mathrm{X}_2$ defined by $g \mapsto g.\mathbb{Z}^2$. For $\varepsilon > 0$ small enough, we assume that

$$(s,t,r) \mapsto \begin{bmatrix} e^s & 0 \\ 0 & e^{-s} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 \\ t & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix}$$

on $(-\varepsilon, \varepsilon)^2 \times [0,1]$ is a homeomorphism onto its image.

Lemma 2.26. *There exists a positive continuous function $\varphi : (-\varepsilon, \varepsilon) \times (-\varepsilon, \varepsilon) \times (0, 1) \to \mathbb{R}$ such that*

$$\mathrm{m}_{\mathrm{X}_2}|_{\mathcal{O}_\varepsilon} = \mathrm{Obt}_* \left(\varphi |\mathrm{dsdtdr}|\right).$$

For $n \in \mathbb{Z}^+$, define

$$\mathbf{BAD}_n := \left\{ r \in \mathbf{BAD} \;\middle|\; \mathrm{sys}(\mathbf{a}_t.\Lambda_r) \geq \frac{1}{n},\ \forall t > 0 \right\}$$

By Dani correspondence, $\mathbf{BAD} = \bigcup_{n \in \mathbb{Z}^+} \mathbf{BAD}_n$. Thus $\mathrm{Leb}(\mathbf{BAD}_{n_0}) > 0$ for some $n_0 \in \mathbb{Z}^+$. Let $\mathcal{O}_\varepsilon(\mathbf{BAD}_{n_0})$ be the subset of $\mathcal{O}_\varepsilon$ where $r \in \mathbf{BAD}_{n_0}$. By Lemma 2.26,

$$\mathrm{m}_{\mathrm{X}_2}(\mathcal{O}_\varepsilon(\mathbf{BAD}_{n_0})) > 0.$$

Let

$$B_n := \overline{\bigcup_{s \geq n} \mathbf{a}_s.\mathcal{O}_\varepsilon(\mathbf{BAD}_{n_0})}$$

$$B := \bigcap_{n \in \mathbb{Z}^+} B_n = \left\{ x = \lim_{n \to \infty} \mathbf{a}_{s_n}.x_n \text{ for some } (s_n) \to +\infty,\ (x_n) \subset \mathcal{O}_\varepsilon(\mathbf{BAD}_{n_0}) \right\}.$$

Since each B_n contains $\mathbf{a}_s.\mathcal{O}_\varepsilon(\mathbf{BAD}_{n_0})$ for some s, we have

$$\mathrm{m}_{\mathrm{X}_2}(B_n) \geq \mathrm{m}_{\mathrm{X}_2}\left(\mathbf{a}_s.\mathcal{O}_\varepsilon(\mathbf{BAD}_{n_0})\right) = \mathrm{m}_{\mathrm{X}_2}(\mathcal{O}_\varepsilon(\mathbf{BAD}_{n_0}))$$

for every n. Hence $\mathrm{m}_{\mathrm{X}_2}(B) \geq \mathrm{m}_{\mathrm{X}_2}(\mathcal{O}_\varepsilon(\mathbf{BAD}_{n_0})) > 0$.

On the other hand, B is A-invariant as well as contained in

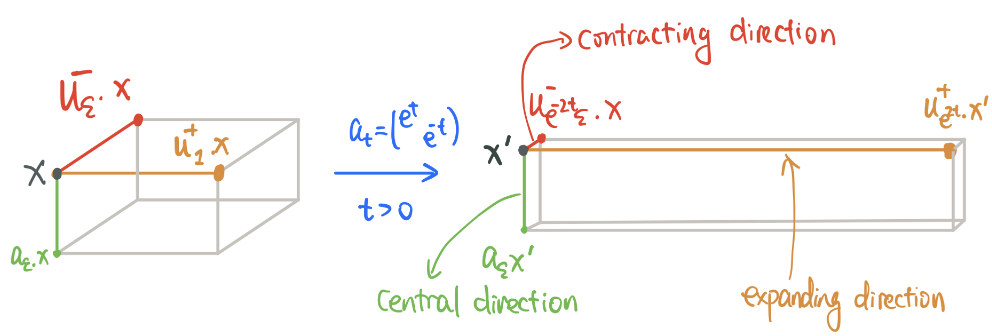
$$\mathscr{C} := \mathbf{a}_{(-\varepsilon,\varepsilon)}.\left\{ \Lambda \in \mathrm{X}_2 \;\middle|\; \mathrm{sys}(\Lambda) \geq \frac{1}{n_0} \right\}.$$

(Exercise: check this is indeed true) So $\mathrm{m}_{\mathrm{X}_2}(\mathbf{a}_t B \cap \mathscr{C}) = \mathrm{m}_{\mathrm{X}_2}(\mathscr{C}) > 0$ for all $t \in \mathbb{R}$. On the other hand, mixing implies that

$$\lim_{t \to \infty} \mathrm{m}_{\mathrm{X}_2}(\mathbf{a}_t B \cap \mathscr{C}) = \begin{cases} \mathrm{m}_{\mathrm{X}_2}(B) \cdot \mathrm{m}_{\mathrm{X}_2}(\mathscr{C}) & \text{if } |\mathrm{m}_{\mathrm{X}_2}| = 1 \\ 0 & \text{if } |\mathrm{m}_{\mathrm{X}_2}| = \infty. \end{cases}$$

In any case, we have arrived at a contradiction.

Here is an explanation of this thickening trick by picture.



2.12. **Exercises.** Two lattices $\Lambda_1, \Lambda_2 \subset \mathbb{R}^2$ are said to be **commensurable** iff $\Lambda_1 \cap \Lambda_2$ is a finite-index subgroup in both Λ_1 and Λ_2.

Exercise A. *Let $\Lambda_0 \in \mathrm{X}_2$ be a unimodular lattice, then the set*

$$\{\Lambda \in \mathrm{X}_2 \mid \Lambda \text{ is commensurable with } \Lambda_0\}$$

is dense in X_2.

Recall $\mathrm{A} = \left\{ \mathbf{a}_t := \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix} \;\middle|\; t \in \mathbb{R} \right\}$.

Exercise B. *Assume $\Lambda_1, \Lambda_2 \in \mathrm{X}_2$ are commensurable. Show that*

1. $(\mathbf{a}_t \Lambda_1)_{t>0}$ *diverges iff* $(\mathbf{a}_t \Lambda_2)_{t>0}$ *diverges;*
2. $\mathrm{A}.\Lambda_1$ *is bounded (i.e., closure is compact) iff* $\mathrm{A}.\Lambda_2$ *is bounded.*

Exercise C. *For $\varepsilon > 0$, let $B_\varepsilon := \{x \in \mathbb{R}^2,\ \|x\| < \varepsilon\}$. Show that for any $\Lambda \in \mathrm{X}_2$, one has $\Lambda \cap B_1 \subset \mathbb{Z}.\mathbf{v}$ for some $\mathbf{v} \in \Lambda$.*

Exercise D. *For $\alpha \in [0,1)$, let $\Lambda_\alpha \in \mathrm{X}_2$ be as in the lecture notes. Show that $(\mathbf{a}_t.\Lambda_\alpha)_{t>0}$ diverges iff $\alpha \in \mathbb{Q}$.*

(Hint: you might want to use Exercise C.)

Exercise E. *Though it is claimed in Theorem 2.6, we did not explain why* sys *is a continuous function. Prove it. Also, can you find $\sup_{\Lambda \in \mathrm{X}_2} \mathrm{sys}(\Lambda)$?*

Exercise F. *We know that for some constant $C > 0$, for every irrational number α, there are infinitely many $q \in \mathbb{Z}^+$ such that $q\langle q\alpha \rangle < C$. Use the boundedness of systole function to give another proof of this fact.*

Definition 2.27. *We define*

$$\mathrm{Y}_2 := \left\{ (\Lambda, \mathbf{v} + \Lambda) \;\middle|\; \Lambda \in \mathrm{X}_2,\ \mathbf{v} + \Lambda \in \mathbb{R}^2/\Lambda \right\}$$

An element $(\Lambda, \mathbf{v}+\Lambda) \in \mathrm{Y}_2$[6] *is referred to as a **unimodular grid**. A sequence $(\Lambda_n, \mathbf{v}_n + \Lambda_n)$ converges to $(\Lambda, \mathbf{v}+\Lambda)$ iff there are $\mathbf{x}_n, \mathbf{y}_n, \mathbf{v}'_n \in \mathbb{R}^2$ and $\mathbf{x}, \mathbf{y}, \mathbf{v}' \in \mathbb{R}^2$, such that*

$$\Lambda_n = \mathbb{Z}\mathbf{x}_n + \mathbb{Z}\mathbf{y}_n,\ \mathbf{v}_n + \Lambda_n = \mathbf{v}'_n + \Lambda_n,\ \Lambda = \mathbb{Z}\mathbf{x} + \mathbb{Z}\mathbf{y},\ \mathbf{v} + \Lambda = \mathbf{v}' + \Lambda;$$

$$(\mathbf{x}_n) \text{ converges to } \mathbf{x},\ (\mathbf{y}_n) \text{ converges to } \mathbf{y},\ \text{and } (\mathbf{v}'_n) \text{ converges to } \mathbf{v}.$$

Also note that $\mathbf{SL}_2(\mathbb{R})$ acts on Y_2 by $(g, \mathbf{v} + \Lambda) \mapsto g\mathbf{v} + g\Lambda$.

Exercise G. *Let B_ε be as above. Show that for $\varepsilon > 0$ sufficiently small (say, $\varepsilon = 0.01$ should suffice), for any unimodular grid $(\Lambda, \mathbf{v}+\Lambda)$, one has that $B_\varepsilon \cap (\mathbf{v}+\Lambda)$ is contained in a line (not necessarily passing through the origin).*

For $\alpha, \beta \in [0,1)$, define a unimodular grid by $y_{\alpha,\beta} = (\Lambda, (\beta, 0)^{\mathrm{tr}} + \Lambda) \in \mathrm{Y}_2$[7].

Exercise H. *Take $\alpha, \beta \in [0,1)$. The following two are equivalent:*

1. *for any $\varepsilon > 0$, there exists $t_0 > 0$ such that for all $t > t_0$, $\mathbf{a}_t.y_{\alpha,\beta} \cap B_\varepsilon \neq \emptyset$;*
2. *$\beta \in \mathbb{Z} + \mathbb{Z}\alpha$.*

Exercise I. *Show that for some constant $C > 0$, for every $\alpha, \beta \in [0,1)$ with $\beta \notin \mathbb{Z} + \mathbb{Z}\alpha$, there are infinitely many $q \in \mathbb{Z}^+$, such that $q\langle q\alpha + \beta\rangle < C$.*

In the first lecture, the homogeneous version was deduced from a theorem of Dirichlet, which is no longer true in the inhomogeneous setting.

Exercise J. *Prove that for any $c > 0$, there exist $\alpha, \beta \in \mathbb{R}$ such that there exists infinitely many $N \in \mathbb{Z}^+$ such that for every $q \in \{0, 1, ..., N-1\}$,*

$$\langle q\alpha + \beta \rangle > \frac{c}{N}.$$

The purpose of the following two exercises is to show you a curious calculation.

Exercise K. *Define*

$$\mathcal{R} := \left\{ (x,y) \in \mathbb{R}^2 \;\middle|\; -\frac{1}{2} < x < \frac{1}{2},\ x^2 + y^2 > 1,\ y > 0 \right\}$$

Calculate the following double integral

$$\int_{\mathcal{R}} \frac{\mathrm{dxdy}}{y^2} = \frac{\pi}{3}.$$

Exercise L. *Let p be a prime number, show that $\#\,\mathbf{SL}_2(\mathbb{Z}/p\mathbb{Z}) = (p^2-1)p$.*

(Hint: find out $\#\,\mathbf{GL}_2(\mathbb{Z}/p\mathbb{Z})$ first.)

Remark. It is not necessary to read this, but here are some contexts about the exercises K and L above. Exercise K shows that with respect to the volume form as defined by "$\frac{\mathrm{dxdydz}}{|x|}$", any (strict) fundamental domain for $\mathbf{SL}_2(\mathbb{Z})$ has volume $\frac{\pi}{3} \times \frac{\pi}{2} = \frac{\pi^2}{6}$. With respect to the same volume form, one can show that the volume of $\mathbf{SL}_2(\mathbb{Z}_p)$ (p-adic integers) is equal to $p^{-3}\,|\mathbf{SL}_2(\mathbb{Z}/p\mathbb{Z})|$ (which is equal to $1 - p^{-2}$ by Exercise L). The fact that $\zeta(2) = \frac{\pi^2}{6}$ shows that

$$\left(\frac{\pi}{3} \times \frac{\pi}{2}\right) \cdot \prod \frac{(p^2-1)p}{p^3} = \frac{\pi^2}{6} \cdot \prod (1 - p^{-2}) = 1.$$

[6] this is sometimes abbreviated as $\mathbf{v} + \Lambda$

[7] we take transpose of a row vector as by convention, we write vectors as column vectors

By putting things together we have

$$\mathrm{Vol}(\mathbf{SL}_2(\mathbb{R})/\mathbf{SL}_2(\mathbb{Z})) \times \prod_p \mathrm{Vol}(\mathbf{SL}_2(\mathbb{Z}_p)) = 1.$$

In adelic language, this can be restated as $\mathrm{Vol}(\mathbf{SL}_2(\mathbb{A}_{\mathbb{Q}})/\mathbf{SL}_2(\mathbb{Q})) = 1$. □

Appendix. Let

$$T^1(H^2) := \{(\lambda, v) \in \mathbb{C} \times \mathbb{C} \mid Im(\lambda) > 0,\ |v| = Im(\lambda)\}.$$

Then $\mathbf{SL}_2(\mathbb{R})$ acts (a right action) on $T^1(H^2)$ by

$$\left(\begin{bmatrix} x & y \\ z & \frac{1+yz}{x} \end{bmatrix}, (\lambda, v) \right) \mapsto \left(\frac{w\lambda - y}{-z\lambda + x}, \frac{v}{(-z\lambda + x)^2} \right).$$

Note that the following map

$$\begin{aligned} [0, 2\pi) \times \mathbb{R} \times \mathbb{R} &\to \mathbf{SL}_2(\mathbb{R}) \\ (a, b, c) &\mapsto \begin{bmatrix} \cos(a) & -\sin(a) \\ \sin(a) & \cos(a) \end{bmatrix} \cdot \begin{bmatrix} 1 & b \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} e^c & 0 \\ 0 & e^{-c} \end{bmatrix}. \end{aligned}$$

is a Borel isomorphism (actually a diffeomorphism if we identify 2π with 0) sending the measure

$$|\mathrm{da} \wedge \mathrm{db} \wedge \mathrm{dc}| \to \mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}.$$

Under this coordinates, the orbit map of $\mathbf{SL}_2(\mathbb{R})$ based at $(i, i) \in T^1(H^2)$ is given by

$$(a, b, c) \mapsto \left(e^{-2c}(i - b), e^{-2c}(\cos(2a)i - \sin(2a))\right)$$

This map induces a Borel isomorphism from $[0, \pi) \times \mathbb{R} \times \mathbb{R}$ to $T^1(H^2)$ and pushes the measure

$$|\mathrm{da} \wedge \mathrm{db} \wedge \mathrm{dc}| \to \frac{|\mathrm{dx} \wedge \mathrm{dy} \wedge \mathrm{d}\tau|}{4y^2}$$

where we identify (a Borel isomorphism) $\mathbb{R} \times \mathbb{R}^+ \times [0, 2\pi)$ with $T^1(H^2)$ via

$$(x, y, \tau) \mapsto (x + iy, ye^{i\tau}).$$

Notation

Let I_2 be the two-by-two identity matrix.

Fix a right invariant metric $d^{\mathbf{SL}_2(\mathbb{R})}$ on $\mathbf{SL}_2(\mathbb{R})$ compatible with its topology. Let d^{X_2} be the induced metric on $\mathbf{SL}_2(\mathbb{R})/\mathbf{SL}_2(\mathbb{Z}) \cong \mathrm{X}_2$ defined by

$$d^{\mathrm{X}_2}(g_1 \mathbf{SL}_2(\mathbb{Z}), g_2 \mathbf{SL}_2(\mathbb{Z})) := \inf_{\gamma_1, \gamma_2 \in \mathbf{SL}_2(\mathbb{Z})} d^{\mathbf{SL}_2(\mathbb{R})}(g_1\gamma_1, g_2\gamma_2) = \inf_{\gamma \in \mathbf{SL}_2(\mathbb{Z})} d^{\mathbf{SL}_2(\mathbb{R})}(g_1, g_2\gamma).$$

Note that this inf can actually be obtained by some $\gamma \in \mathbf{SL}_2(\mathbb{Z})$.

For $\delta > 0$ and $x \in \mathrm{X}_2$, let

$$B(\delta) := \left\{ g \in \mathbf{SL}_2(\mathbb{R}) \;\middle|\; d^{\mathbf{SL}_2(\mathbb{R})}(g, \mathrm{I}_2) < \delta \right\},\ \ B_x^{\mathrm{X}_2}(\delta) := \left\{ y \in \mathrm{X}_2 \mid d^{\mathrm{X}_2}(x, y) < \delta \right\}.$$

For simplicity, we will write $d := d^{\mathbf{SL}_2(\mathbb{R})}$ and $B_x(\delta) = B_x^{\mathrm{X}_2}(\delta)$. Hopefully no confusion shall arise.

3. Appendix to Lecture 2, Injectivity Radius

Let π denote the natural quotient map $\mathbf{SL}_2(\mathbb{R}) \to \mathbf{SL}_2(\mathbb{R})/\mathbf{SL}_2(\mathbb{Z})$.

3.1. Injectivity radius.

Definition 3.1. *For $x \in \mathrm{X}_2 \cong \mathbf{SL}_2(\mathbb{R})/\mathbf{SL}_2(\mathbb{Z})$, choose $g_x \in \mathbf{SL}_2(\mathbb{R})$ such that $x = g_x \mathbf{SL}_2(\mathbb{Z})$, define*[8]

$$\mathrm{InjRad}(x) := \frac{1}{10} \inf \left\{ d(g_x\gamma, g_x) \mid \gamma_{\neq \mathrm{I}_2} \in \mathbf{SL}_2(\mathbb{Z}) \right\},$$

which is independent of the choice of g_x.

Lemma 3.2. *Let $\mathscr{C} \subset \mathrm{X}_2$ be a compact subset, then there exists $c > 0$ such that $\mathrm{InjRad}(x) > c$ for every $x \in \mathscr{C}$.*

[8] We are content with this rather coarse definition of injectivity radius, which might be different from the one you are used to.

Proof. Since every compact subset of X_2 is contained in the image under π of some compact subset of X_2, it suffices to show that

$$\inf\left\{d(\mathrm{I}_2, g\gamma g^{-1}) \;\middle|\; g \in \mathscr{C}',\ \gamma_{\neq \mathrm{I}_2} \in \mathbf{SL}_2(\mathbb{Z})\right\} > 0$$

for every compact subset $\mathscr{C}'$ of $\mathbf{SL}_2(\mathbb{R})$.

Let $\Gamma(1)$ be the collection of $\gamma \in \Gamma \setminus \{\mathrm{I}_2\}$ such that $g\gamma g^{-1} \in \overline{B(1)}$ for some $g \in \mathscr{C}'$. Here $B(1) := \{d(g, \mathrm{I}_2) < 1\}$. Since the map $(g, h) \mapsto g^{-1}hg$ is continuous, we know that the union of $g^{-1}\overline{B(1)}g$ as g varies in $\mathscr{C}'$ is compact. Hence $\Gamma(1)$ is a compact subset of a discrete subset $\Gamma \setminus \{\mathrm{I}_2\}$, which must be finite. Say, $\Gamma(1) = \{\gamma_1, ..., \gamma_l\}$.

Then

$$\inf\left\{d(\mathrm{I}_2, g\gamma g^{-1}) \;\middle|\; g \in \mathscr{C}',\ \gamma_{\neq \mathrm{I}_2} \in \mathbf{SL}_2(\mathbb{Z})\right\}$$

is either at least one or is equal to

$$\inf\left\{d(\mathrm{I}_2, g) \;\middle|\; g \in \bigcup_{c\in\mathscr{C}'}\bigcup_{i=1}^{l} c\gamma_i c^{-1}\right\} > 0.$$

But

$$\bigcup_{c\in\mathscr{C}'}\bigcup_{i=1}^{l} c\gamma_i c^{-1} = \bigcup_{i=1}^{l}\bigcup_{c\in\mathscr{C}'} c\gamma_i c^{-1}$$

is compact. Also, it does not contain I_2. Therefore it must have positive distance away from I_2. So we are done. $\square$

Lemma 3.3. *For $x \in \mathrm{X}_2$ and $\delta < \mathrm{InjRad}(x)$, the natural map*

$$\mathrm{Obt}_x : B(\delta) \to B(\delta).x$$
$$g \mapsto g.x$$

is an isometry between $(B(\delta), d) \cong (B_x^{\mathrm{X}_2}(\delta), d^{\mathrm{X}_2})$. In particular, $B_x^{\mathrm{X}_2}(\delta) = B(\delta).x$.

Proof. For $g_1, g_2 \in B(\delta)$, we need to show that

$$\inf_{\gamma\in\mathbf{SL}_2(\mathbb{Z})} d(g_1 g_x, g_2 g_x \gamma) = d(g_1, g_2).$$

In different words,

$$d(g_1 g_x, g_2 g_x \gamma) > d(g_1, g_2), \quad \forall\, \gamma_{\neq \mathrm{I}_2} \in \mathbf{SL}_2(\mathbb{Z}).$$

This can be seen from the following inequalities:

$$\begin{aligned} d(g_2 g_x, g_2 g_x \gamma) &> d(g_x, g_x\gamma) - d(g_x, g_2 g_x) - d(g_2 g_x \gamma, g_x \gamma) \\ &= d(g_x, g_x\gamma) - d(\mathrm{I}_2, g_2) - d(g_2, \mathrm{I}_2) \\ &> 10\delta - \delta - \delta = 8\delta. \end{aligned}$$

Then

$$\begin{aligned} d(g_1 g_x, g_2 g_x \gamma) &\geq d(g_2 g_x, g_2 g_x\gamma) - d(g_1 g_x, g_2 g_x) \\ &> d(g_2 g_x, g_2 g_x \gamma) - 2\delta > 8\delta - 2\delta = 6\delta. \end{aligned}$$

But $d(g_1, g_2) < 2\delta$. So we are done. The last claim follows from the definition of the distance function on the quotient. $\square$

3.2. **Integration in local coordinates.** For $\eta > 0$, define

$$\mathcal{O}_\eta := \left\{\mathbf{a}_r \mathbf{u}_s^- \mathbf{u}_t^+ \;\middle|\; r, s, t \in (-\eta, \eta)\right\}.$$

By explicit calculation, one can show that $\mathcal{O}_\eta$ is an open neighborhood of the identity element in $\mathbf{SL}_2(\mathbb{R})$ for every $\eta > 0$.

We fix $\eta_0 > 0$ small enough such that

$$(-\eta_0, \eta_0)^3 \mapsto \mathcal{O}_{\eta_0}$$
$$(r, s, t) \mapsto \mathbf{a}_r \mathbf{u}_s^- \mathbf{u}_t^+$$

is a homeomorphism. We find ϕ_{η_0}, a positive continuous function on $[-\eta_0, \eta_0]^3$, such that for every $f \in L^1(\mathbf{SL}_2(\mathbb{R}), \mathrm{m}_{\mathbf{SL}_2(\mathbb{R})})$,

$$\int_{z\in\mathcal{O}_{\eta_0}} f(z)\mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}(z) = \int_{-\eta_0}^{\eta_0}\int_{-\eta_0}^{\eta_0}\int_{-\eta_0}^{\eta_0} f(\mathbf{a}_r \mathbf{u}_s^- \mathbf{u}_t^+)\,\phi_{\eta_0}(r, s, t)\,\mathrm{drdsdt}.$$

Fix a constant $C_1 > 1$ such that $\|\phi_{\eta_0}\|_{\sup} \leq C_1$.

By the relation between $\mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}$ and $\mathrm{m}_{\mathrm{X}_2}$, one can show that

Lemma 3.4. *Let $x \in \mathrm{X}_2$ and $\delta < \mathrm{InjRad}(x)$. Let $0 < \eta < \eta_0$ be such that $\mathcal{O}_\eta \subset B(\delta)$. Then for every $f \in L^1(\mathrm{X}_2, \mathrm{m}_{\mathrm{X}_2})$,*

$$\int_{z \in \mathcal{O}_\eta . x} f(z) \mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}(z) = \int_{-\eta}^{\eta} \int_{-\eta}^{\eta} \int_{-\eta}^{\eta} f(\mathbf{a}_r \mathbf{u}_s^- \mathbf{u}_t^+ . x)\, \phi_{\eta_0}(r,s,t)\, \mathrm{drdsdt}.$$

Proof. This follows from Lemma 3.3. □

3.3. **Uniform mixing in a weak sense.** The main result of this appendix is the following very weak form of equidistribution of expanding unipotent trajectories. The point is the uniformity as the base points vary in a compact subset.

Theorem 3.5. *Fix $y_0 \in \mathrm{X}_2$, $\varepsilon_0 \in (0,1)$ and a compact subset $\mathscr{C}$ of X_2. There exist $\delta, T > 0$ and $M \in 2\mathbb{Z}^+$ such that for every $x \in \mathscr{C}$ and $T' > T$,*

$$\int_{-0.5}^{0.5} \mathbf{1}_{B_{y_0}(\varepsilon_0)}(\mathbf{a}_{T'} \mathbf{u}_t^+ . x)\, \mathrm{dt} > \delta.$$

Note

$$\int_{-0.5}^{0.5} \mathbf{1}_{B_{y_0}(\varepsilon_0)}(\mathbf{a}_{T'} \mathbf{u}_t^+ . x)\, \mathrm{dt} = \mathrm{Leb}\left\{ t \in [-0.5, 0.5] \,\middle|\, \mathbf{a}_{T'} \mathbf{u}_t^+ . x \in B_{y_0}(\varepsilon_0) \right\}.$$

3.4. **Preparations.** Firstly, by Lemma 3.2, we find $0 < \eta_1 < \eta_0$ such that $\mathcal{O}_{\eta_1} \subset B(\delta_0)$ for some $\delta_0 > 0$ that is smaller than $\mathrm{InjRad}(x)$ for all $x \in \mathscr{C}$. Thus, Lemma 3.4 is applicable to every $x \in \mathscr{C}$ and $\eta = \eta_1$.

Then choose $0 < \eta_2 < \min\{\eta_1, 0.1\}$ such that $\mathcal{O}_{\eta_2} \subset B(0.5\varepsilon_0)$. This has the effect that

Lemma 3.6. *For every $x \in \mathrm{X}_2$, $T \geq 0$ and $r, s \in (-\eta_2, \eta_2)$, one has the following implication:*

$$\mathbf{a}_T.(\mathbf{a}_r \mathbf{u}_s^- . x) \in B_{y_0}(0.5\varepsilon_0) \implies \mathbf{a}_T . x \in B_{y_0}(\varepsilon_0).$$

Proof. Indeed, given $\mathbf{a}_T.(\mathbf{a}_r \mathbf{u}_s^- . x) \in B_{y_0}(0.5\varepsilon_0)$, we have

$$\begin{aligned} d^{\mathrm{X}_2}(\mathbf{a}_T.x, y_0) &\leq d^{\mathrm{X}_2}(\mathbf{a}_T.x, \mathbf{a}_T.\mathbf{a}_r\mathbf{u}_s^-.x) + d^{\mathrm{X}_2}(\mathbf{a}_T.\mathbf{a}_r\mathbf{u}_s^-.x, y_0) \\ &< d^{\mathrm{X}_2}(\mathbf{a}_T.x, \mathbf{a}_r \mathbf{u}^-_{e^{-2T}s} \mathbf{a}_T.x) + 0.5\varepsilon_0 \\ &\leq d(\mathrm{I}_2, \mathbf{a}_r \mathbf{u}^-_{e^{-2T}s}) + 0.5\varepsilon_0 \\ (\because \mathcal{O}_{\eta_2} \subset B(0.5\varepsilon_0)\ ) &< 0.5\varepsilon_0 + 0.5\varepsilon_0 = \varepsilon_0. \end{aligned}$$

□

Next we choose $0 < \eta_3 < \eta_2$ satisfying the following:

Lemma 3.7. *There exists $0 < \eta < \eta_2$ such that for every $x, y \in \mathscr{C}$, the following implication holds:*

$$x \in \mathcal{O}_\eta . y \implies \mathcal{O}_\eta . y \subset \mathcal{O}_{\eta_2} . x.$$

Proof. Choose $0 < \theta < \delta_0$ (the uniform injectivity radius) such that $B(\theta) \subset \mathcal{O}_{\eta_2}$. Then choose $0 < \eta < \eta_2$ such that $\mathcal{O}_\eta \subset B(0.5\theta)$. So

$$x \in \mathcal{O}_\eta . y \implies x \in B(0.5\theta).y \implies y \in B(\theta).x \subset \mathcal{O}_{\eta_2}.x.$$

This completes the proof. □

3.5. **Proof of Theorem 3.5.** Find $M \in 2\mathbb{Z}^+$ large such that $\eta_2^{-1} - 2 \leq M \leq \eta_2^{-1}$. By compactness, find finitely many $\{x_1, ..., x_l\} \subset \mathscr{C}$ such that

$$\mathscr{C} \subset \bigcup_{i=1}^{l} \mathcal{O}_{\eta_3} . x_i.$$

By mixing (Theorem 1.22 from Lecture 2), for each $i = 1, .., l$, we find $T_i > 0$ such that for every $T > T_i$,

$$\begin{aligned} \mathrm{m}_{\mathrm{X}_2}\left(\mathcal{O}_{\eta_3}.x_i \cap \mathbf{a}_T^{-1} B_{y_0}(0.5\varepsilon_0)\right) &> 0.5 \mathrm{m}_{\mathrm{X}_2}\left(\mathcal{O}_{\eta_3}.x_i\right) \mathrm{m}_{\mathrm{X}_2}\left(B_{y_0}(0.5\varepsilon_0)\right) \\ &= 0.5 \mathrm{m}_{\mathbf{SL}_2(\mathbb{R})}\left(\mathcal{O}_{\eta_3}\right) \mathrm{m}_{\mathrm{X}_2}\left(B_{y_0}(0.5\varepsilon_0)\right). \end{aligned}$$

Let $T := \max\{T_i\}$ and c_1 denote the right hand side. Also, let

$$\delta := \frac{c_1}{C_1 4 (\eta_2)^2}.$$

Now take $x \in \mathscr{C}$ and $T' > T$ and let us prove the conclusion.

Find i such that $x \in \mathcal{O}_{\eta_3}.x_i$. By Lemma 3.7, we have $\mathcal{O}_{\eta_3}.x_i \subset \mathcal{O}_{\eta_2}.x$.

So

$$\begin{aligned}
c_1 &< \mathrm{m}_{\mathrm{X}_2}\left(\mathcal{O}_{\eta_3}.x_i \cap \mathbf{a}_{T'}^{-1} B_{y_0}(0.5\varepsilon_0)\right) \\
&< \mathrm{m}_{\mathrm{X}_2}\left(\mathcal{O}_{\eta_2}.x \cap \mathbf{a}_{T'}^{-1} B_{y_0}(0.5\varepsilon_0)\right) \\
&= \int_{\mathcal{O}_{\eta_2}.x} \mathbf{1}_{B_{y_0}(0.5\varepsilon_0)}(\mathbf{a}_{T'} z)\,\mathrm{m}_{\mathrm{X}_2}(z) \\
\text{(local integration lemma 3.4)} &= \int_{-\eta_2}^{\eta_2}\int_{-\eta_2}^{\eta_2}\int_{-\eta_2}^{\eta_2} \mathbf{1}_{B_{y_0}(0.5\varepsilon_0)}(\mathbf{a}_{T'}\mathbf{a}_r\mathbf{u}^-_s\mathbf{u}^+_t.x)\phi_{\eta_0}(r,s,t)\mathrm{drdsdt} \\
\text{(boundedness of density function)} &\le C_1 \int_{-\eta_2}^{\eta_2}\int_{-\eta_2}^{\eta_2}\int_{-\eta_2}^{\eta_2} \mathbf{1}_{B_{y_0}(0.5\varepsilon_0)}(\mathbf{a}_{T'}\mathbf{a}_r\mathbf{u}^-_s\mathbf{u}^+_t.x)\mathrm{drdsdt} \\
\text{(Lemma 3.6)} &\le C_1 \int_{-\eta_2}^{\eta_2}\int_{-\eta_2}^{\eta_2}\int_{-\eta_2}^{\eta_2} \mathbf{1}_{B_{y_0}(\varepsilon_0)}(\mathbf{a}_{T'}\mathbf{u}^+_t.x)\mathrm{drdsdt} \\
&= C_1 4\eta_2^2 \int_{-\eta_2}^{\eta_2} \mathbf{1}_{B_{y_0}(\varepsilon_0)}(\mathbf{a}_{T'}\mathbf{u}^+_t.x)\mathrm{dt} \\
&< C_1 4\eta_2^2 \int_{-0.5}^{0.5} \mathbf{1}_{B_{y_0}(\varepsilon_0)}(\mathbf{a}_{T'}\mathbf{u}^+_t.x)\mathrm{dt}
\end{aligned}$$

Finally we have

$$\int_{-0.5}^{0.5} \mathbf{1}_{B_{y_0}(\varepsilon_0)}(\mathbf{a}_{T'}\mathbf{u}^+_t.x)\mathrm{dt} > \frac{c_1}{C_1 4(\eta_2)^2} = \delta.$$

3.6. **Exercises.**

Exercise A. *Let* $\mathbf{a}_t := \begin{bmatrix} e^t & 0 \\ 0 & e^{-t} \end{bmatrix}$. *Assume* $x \in \mathrm{X}_2$ *is such that*

$$\{\mathbf{a}_t.x \mid t \ge 0\} \subset \mathrm{X}_2$$

is bounded. Take another $y \in \mathrm{X}_2$. *Show that the following two are equivalent*

(1) $\lim_{t\to+\infty} d(\mathbf{a}_t.x, \mathbf{a}_t.y)$ *exists and is* 0*;*

(2) $y = \begin{bmatrix} 1 & 0 \\ r & 1 \end{bmatrix}.x$ *for some* $r \in \mathbb{R}$.

Recall: $d^{\mathbf{SL}_2(\mathbb{R})}(\cdot,\cdot)$ (or just write d) denotes a right $\mathbf{SL}_2(\mathbb{R})$-invariant metric on $\mathbf{SL}_2(\mathbb{R})$ that is compatible with the topology on $\mathbf{SL}_2(\mathbb{R})$ and d^{X_2} denotes the quotient metric on $\mathrm{X}_2 \cong \mathbf{SL}_2(\mathbb{R})/\mathbf{SL}_2(\mathbb{Z})$. Using this we defined

$$\mathrm{InjRad}(x) := \frac{1}{10}\inf\left\{d^{\mathrm{X}_2}(g,h) \;\middle|\; g \ne h \in \mathbf{SL}_2(\mathbb{R}),\ g.\mathbb{Z}^2 = h.\mathbb{Z}^2 = x\right\}.$$

Exercise B. *Show that a sequence* $(x_n) \subset \mathrm{X}_2$ *diverges iff* $\mathrm{InjRad}(x_n) \to 0$.

Hint: Find nilpotent integer matrices (N_n) such that $(g_n N_n g_n^{-1})$ converges to the zero matrix.

Exercise C. *Show that for* $x \in \mathrm{X}_2$, *if* $\left\{\begin{bmatrix} 1 & r \\ 0 & 1 \end{bmatrix}.x \;\middle|\; r \in \mathbb{R}\right\}$ *is compact, then* $\left(\begin{bmatrix} e^{t_n} & 0 \\ 0 & e^{-t_n} \end{bmatrix}.x\right)$ *diverges in* X_2 *whenever* $t_n \to -\infty$.

Hint: use the exercise above.

4. Lecture 3, Hausdorff Dimension of BAD

Reference: See [Mat95] for more on Hausdorff dimensions. The main result in this lecture follows [KM96]. Lemma 4.3 is taken from [McM87].

4.1. **Prelude.** When a set has Lebesgue measure zero, there is a more refined way of measuring its size: Hausdorff dimension. A Lebesgue-null subset of $[0,1)$ could have dimension from 0 to 1. The classical Cantor's middle third set has Hausdorff dimension $\frac{\log 2}{\log 3}$. In this lecture we will show that the set of badly approximable numbers, which is small in terms of Lebesgue measure, is big in terms of Hausdorff dimension. Its Hausdorff dimension is equal to 1, proved by Jarnik. We are going to follow the proof by Kleinbock–Margulis, using the mixing property of geodesic flow to construct a Cantor-like set in **BAD** with large Hausdorff dimension.

4.2. **Hausdorff dimension.** Let $E \subset [0,1)$. For $s > 0$ and $\varepsilon > 0$, define

$$\mathcal{H}^s_\varepsilon := \inf \left\{ \sum \operatorname{diam}(I_i)^s \,\middle|\, E \subset \bigcup I_i \text{ countable union of intervals, } \operatorname{diam}(I_i) < \varepsilon,\ \forall\, i \right\}.$$

For $s > 0$, define

$$\mathcal{H}^s(E) := \lim_{\varepsilon \to 0} \mathcal{H}^s_\varepsilon(E).$$

Note that such a limit indeed exists (possibly $+\infty$) since $\mathcal{H}^s_\varepsilon(E)$ is non-decreasing as ε decreases to 0.

The Hausdorff dimension is defined by

$$\dim_{\mathrm{H}}(E) := \inf \{ s \geq 0 \mid \mathcal{H}^s(E) = 0 \}. \tag{8}$$

If non-empty (otws, $\dim_{\mathrm{H}}(E) = 0$.), then one can directly check that

$$\dim_{\mathrm{H}}(E) := \sup \{ s \geq 0 \mid \mathcal{H}^s(E) = +\infty \}.$$

From the definition, one sees that

Lemma 4.1. *Let $\alpha \in [0,1]$ and E be a subset of $[0,1)$. If there exist $C, \varepsilon > 0$ such that for every covering of E by countably many intervals (I_i) with $\operatorname{diam}(I_i) < \varepsilon$, one has $\sum \operatorname{diam}(I_i)^\alpha > C$, then $\dim_{\mathrm{H}}(E) \geq \alpha$.*

The main goal of this lecture is to prove that

Theorem 4.2. *The Hausdorff dimension of* **BAD** *is equal to* 1.

4.3. **Lower bound of Hausdorff dimension.** For $N \in \mathbb{Z}^+$, let $\mathscr{I}_N$ denote the collection of intervals

$$\mathscr{I}_N := \left\{ [\frac{i}{N}, \frac{i+1}{N}) \,\middle|\, i = 0, 1, ..., N-1 \right\}.$$

Lemma 4.3. *Fix $N \in \mathbb{Z}^+$ and $\delta \in (0,1)$. Suppose that for each $k \in \mathbb{Z}^+$, we have a subset $\mathscr{E}_k$ of $\mathscr{I}_{N^k}$ (by default, also set $\mathscr{E}_0 := \{[0,1)\}$) satisfying*

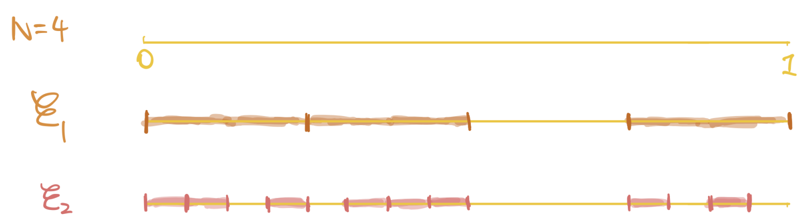
(1) *for every $k \in \mathbb{Z}^+$ and $E \in \mathscr{E}_k$ there exists $E' \in \mathscr{E}_{k-1}$ containing E;*
(2) *for every $k \in \mathbb{Z}^+$ and $E \in \mathscr{E}_{k-1}$,*

$$\frac{\#\{F \in \mathscr{E}_k \mid F \subset E\}}{N} \geq 1 - \delta.$$

Let $E_\infty := \bigcap_{k=1}^\infty \bigcup_{E \in \mathscr{E}_k} E$. Then

$$\dim_{\mathrm{H}}(E_\infty) \geq 1 - \frac{\log((1-\delta)^{-1})}{\log N}.$$

Here is a picture of Cantor like sets...



4.4. **Convergence of measures.** We claim that the "natural" probability measures supported on $\cup_{E \in \mathscr{E}_k} E$ converges as $k \to +\infty$ under the weak$*$ topology[9]. By Riesz's representation theorem (See Rudin's book, real and complex analysis, Theorem 2.14.), we may and do specify a measure by integrating compactly supported continuous functions.

Let f be a continuous function on $[0,1]$, define

$$L_1(f) := \frac{1}{\#\mathscr{E}_1} \sum_{E_1 \in \mathscr{E}_1} N \cdot \int_{E_1} f(x) \mathrm{d}x.$$

[9] A sequence of measures (μ_n) converges to μ under the **weak$*$ topology** iff $\int f(x)\, \mu_n(x)$ converges to $\int f(x)\, \mu(x)$ for every continuous function f on $[0,1]$.

So this is integrating f against the normalized probability measure supported on $\bigsqcup_{E\in\mathscr{E}_1} E$. Then one "refines" this measure by

$$L_2(f) := \frac{1}{\#\mathscr{E}_1}\sum_{E_1\in\mathscr{E}_1}\frac{1}{\#\{E_2\in\mathscr{E}_2, E_2\subset E_1\}}\sum_{E_2\in\mathscr{E}_2,E_2\subset E_1} N^2\cdot\int_{E_2} f(x)\mathrm{dx}.$$

In general, for $E\in\mathscr{E}_k$, let

$$\mathscr{E}_{k+1}(E) := \{F\in\mathscr{E}_{k+1} \mid F\subset E\}.$$

Also, given $k\in\mathbb{Z}^+$ and $E\in\mathscr{E}_k$, define

$$E =: E^k\subset E^{k-1}\subset ...\subset E^1\subset E^0 := [0,1) \tag{9}$$

by requiring E_{i-1} to be the unique element in $\mathscr{E}_{i-1}$ containing $E_i\in\mathscr{E}_i$ for $i = k, k-1, ..., 1$. Further define the weight for E by

$$\mathbf{w}_E := \frac{1}{\#\mathscr{E}_k(E^{k-1})\cdot\#\mathscr{E}_{k-1}(E^{k-2})\cdot ...\cdot\#\mathscr{E}_2(E^1)\cdot\#\mathscr{E}_1}.$$

Now for general $k\in\mathbb{Z}^+$, a positive linear functional L_k is defined for $f\in C[0,1]$[10] by

$$L_k(f) := \sum_{E\in\mathscr{E}_k}\mathbf{w}_E\cdot N^k\cdot\int_E f(x)\mathrm{dx}.$$

Using the fact that f's are uniformly continuous, one can check that

Lemma 4.4. *For every $f\in C[0,1]$, the limit $L_\infty(f) := \lim_{k\to\infty} L_k(f)$ exists and $f\mapsto L_\infty(f)$ is a bounded positive linear functional on $C[0,1]$ mapping the constant one function to 1. Consequently, there exists a probability measure μ such that $L_\infty(f) = \int_0^1 f(x)\mu(x)$.*

We reserve μ for such a measure till the end of the proof of Lemma 4.3.

4.5. **Proof of Lemma 4.3.** For $E\in\mathscr{E}_k$, find $E =: E^k\subset E^{k-1}\subset ...\subset E^1\subset E^0 := [0,1)$ as in Equa.(9). Then $\mu(E) = \mathbf{w}_E$. By assumption, each $\#\mathscr{E}_i(E^{i-1})\ge(1-\delta)N$. Therefore,

$$\mu(E)\le\frac{1}{(1-\delta)^k N^k}.$$

Fix a covering of $E\subset\bigcup I_i$ by countably many intervals. For each i, let k_i be the unique positive integer such that

$$\frac{1}{N^{k_i+1}}\le\operatorname{diam}(I_i) < \frac{1}{N^{k_i}}.$$

Let $\mathscr{E}_{k_i}(I_i)$ collect intervals E in $\mathscr{E}_{k_i}$ with $E\cap I_i\neq\emptyset$. We note that

Lemma 4.5. $\#\mathscr{E}_{k_i}(I_i)\le 2$.

On the other hand, for $E\in\mathscr{E}_{k_i}(I_i)$ and $\alpha\in[0,1]$,

$$\mu(E)\le\frac{N}{(1-\delta)^{k_i}N^{k_i+1}}\le\frac{N\operatorname{diam}(I_i)^\alpha}{(1-\delta)^{k_i}N^{(k_i+1)(1-\alpha)}} = \frac{1}{((1-\delta)N^{1-\alpha})^{k_i}}N^\alpha\operatorname{diam}(I_i)^\alpha,$$

which is $\le N^\alpha\operatorname{diam}(I_i)^\alpha$ provided $(1-\delta)N^{1-\alpha}\ge 1$, or equivalently,

$$\alpha\le 1-\frac{\log((1-\delta)^{-1})}{\log N}.$$

Therefore, for α satisfying the inequality above,

$$1 = \mu(E)\le\sum\mu(I_i)\le\sum_i\sum_{E\in\mathscr{E}_{k_i}(I_i)}\mu(E)\le\sum_i 2\cdot N^\alpha\operatorname{diam}(I_i)^\alpha$$
$$\implies\sum\operatorname{diam}(I_i)^\alpha\ge\frac{1}{2N^\alpha}.$$

As $0.5N^{-\alpha}$ is a positive constant independent of the covering (I_i) chosen, this completes the proof of Lemma 4.3 by Lemma 4.1.

[10] For a topological space X, let $C(X)$ denote the Banach space of continuous functions on X equipped with the sup-norm.

4.6. **A remark.** One could have rewritten the above proof into two steps (let $\alpha < 1 - \frac{\log((1-\delta)^{-1})}{\log N}$):

(1) Construct a probability measure μ on E_∞ with the property that for some $C > 0$, for every r small enough, $\mu((x-r, x+r)) < Cr^\alpha$ holds for μ almost all x;
(2) Show that whenever a set has positive μ-measure, it must has Hausdorff dimension at least α.

It turns out that the second step has a converse to it. Namely, if a set E has Hausdorff dimension $> \alpha$, then one can find a probability measure μ supported on E (meaning, $\mu(E) = 1$) such that for some $C > 0$, for every r small enough, $\mu((x-r, x+r)) < Cr^\alpha$ holds for μ almost all x. This is called the Frostman Lemma.

4.7. **Construct Cantor-like sets in BAD.** We give a construction of $(\mathscr{E}_k)$, whose intersections E_∞ lies inside **BAD**. For this we start with a positive integer N and a compact subset $\mathscr{C}$ of X_2. Define $\mathscr{C}' := \mathbf{u}^+_{[-1,1]}.\mathscr{C}$.

For $x \in \mathrm{X}_2$ and $I = [a_I, b_I) \in \mathscr{I}_N$, define

$$\phi_I(x) := \mathbf{a}_{\frac{1}{2}\log N}\mathbf{u}^+_{a_I}.x = \begin{bmatrix} N^{\frac{1}{2}} & 0 \\ 0 & N^{-\frac{1}{2}} \end{bmatrix} \begin{bmatrix} 1 & a_I \\ 0 & 1 \end{bmatrix}.x.$$

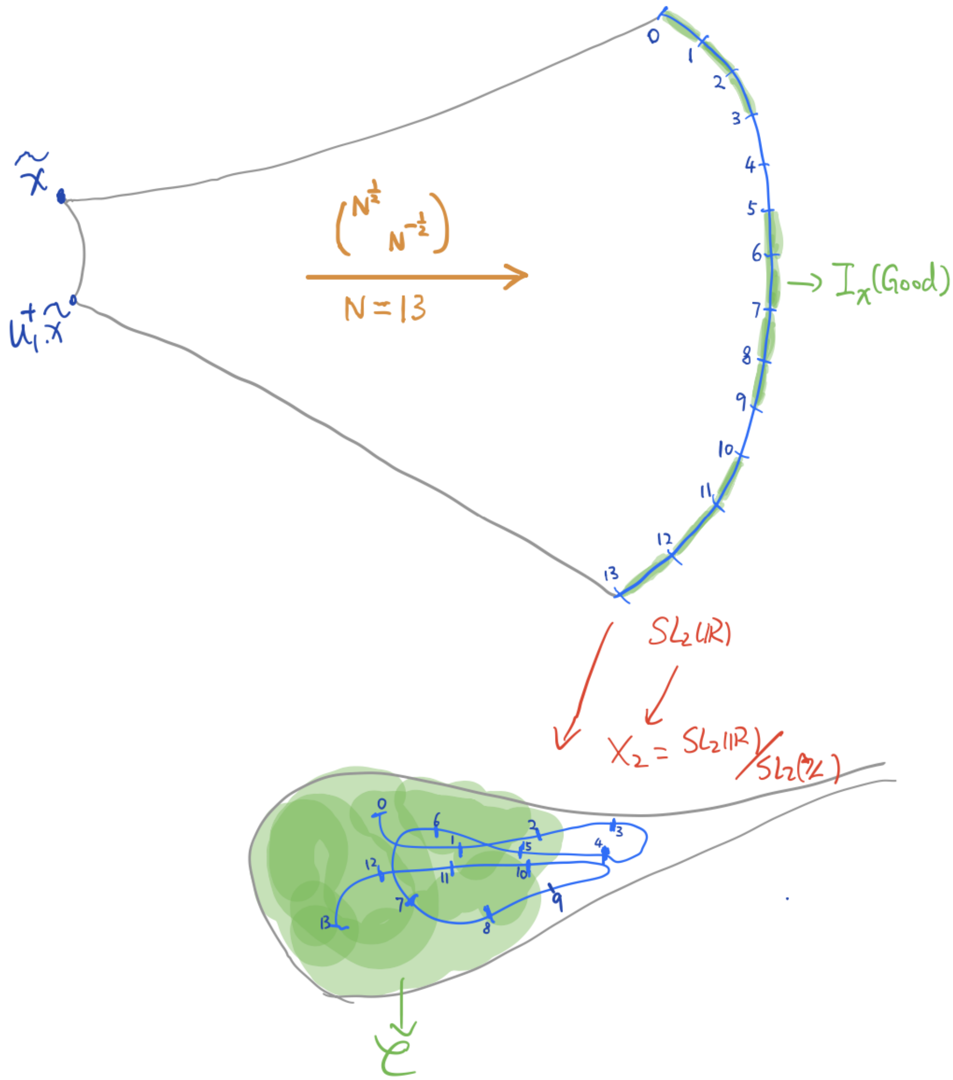
Hence

$$\mathbf{a}_{\frac{1}{2}\log N}\mathbf{u}^+_I.x = \mathbf{u}^+_{[0,1)}.\phi_I(x).$$

For $x \in \mathrm{X}_2$, define a subset $\mathscr{I}_x(\mathrm{Good}) \subset \mathscr{I}_N$ by

$$I \in \mathscr{I}_x(\mathrm{Good}) \iff \mathbf{a}_{\frac{1}{2}\log N}\mathbf{u}^+_I.x \cap \mathscr{C} \neq \emptyset \iff \mathbf{u}^+_{[0,1)}.\phi_I(x) \cap \mathscr{C} \neq \emptyset.$$

Thus $\phi_I(x) \in \mathscr{C}'$ if $I \in \mathscr{I}_x(\mathrm{Good})$.



4.7.1. *Initial steps.* Let $x_0 := \mathbb{Z}^2$, we define $\mathscr{E}_1 := \mathscr{I}_{x_0}(\text{Good})$.

Let

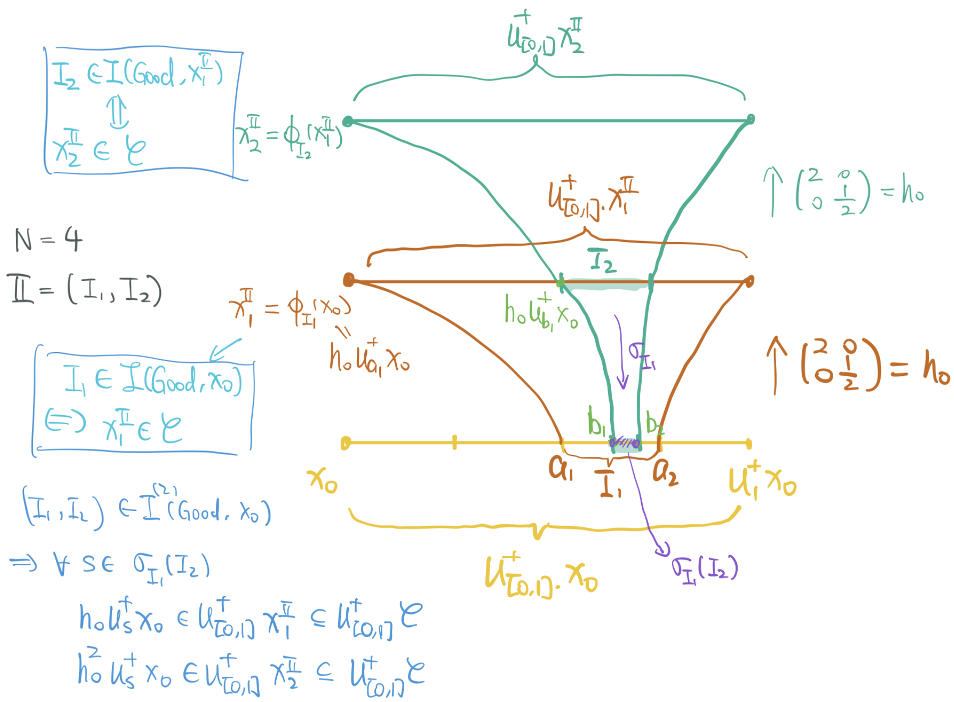
$$\mathscr{I}^2_{x_0}(\text{Good}) := \left\{ (I_1, I_2) \;\middle|\; I_1 \in \mathscr{I}_{x_0}(\text{Good}),\; I_2 \in \mathscr{I}_{\phi_{I_1}(x_0)}(\text{Good}) \right\}$$

For an interval $I = [a_I, b_I)$, let σ_I be the unique (orientation-preserving) affine transformation sending $[0, 1)$ to $[a_I, b_I)$, namely,

$$\begin{aligned} \sigma_I : [0,1) &\to I = [a_I, b_I) \\ t &\mapsto t a_I + (1-t) b_I. \end{aligned}$$

Define

$$\mathscr{E}_2 := \left\{ \sigma_{I_1}(I_2) \;\middle|\; (I_1, I_2) \in \mathscr{I}^2_{x_0}(\text{Good}) \right\}.$$



4.7.2. *In general....* Given a finite sequence $\mathbb{I} := (I_1, I_2, ..., I_k)$ of elements in $\mathscr{I}_N$, for $i \in \{1, ..., k\}$, let

$$x_i^{\mathbb{I}} := \phi_{I_i} \circ \phi_{I_{i-1}} \circ ... \circ \phi_{I_1}(x_0).$$

By default, set $x_0^{\mathbb{I}} := x_0$. Define, for $k \in \mathbb{Z}^+$,

$$\mathscr{I}^k_{x_0}(\mathrm{Good}) := \left\{ \mathbb{I} = (I_1, ..., I_k) \;\middle|\; I_i \in \mathscr{I}_{x^{\mathbb{I}}_{i-1}}(\mathrm{Good}),\ \forall\, i = 1, ..., N \right\},$$

and

$$\mathscr{E}_k := \left\{ \sigma_{I_1} \circ ... \circ \sigma_{I_{k-1}}(I_k) \;\middle|\; (I_1, ..., I_k) \in \mathscr{I}^k_{x_0}(\mathrm{Good}) \right\}.$$

Lemma 4.6. *Given $N \in \mathbb{Z}^+$ and a compact subset $\mathscr{C}$ of X_2. The family of sets $(\mathscr{E}_k)$ constructed above satisfy condition (1) in Lemma 4.3. And $E_\infty := \cap_{k=1}^{\infty} \cup_{E \in \mathscr{E}_k} E$ is contained in* **BAD**.

Proof. The first part follows from the construction. Turn to the second part. By definition, for $s \in \mathscr{E}_k$ associated with $\mathbb{I} = (I_1, ..., I_k)$, and for each $i = 1, .., k$, $x_i^{\mathbb{I}}$ is contained in $\mathscr{C}$. Also, $\mathbf{a}^i_{\frac{1}{2}\log N} \mathbf{u}_s^+ . x_0 \subset \mathbf{u}^+_{[0,1]} . x_i^{\mathbb{I}}$. Thus $\mathbf{a}^k_{\frac{1}{2}\log N} \mathbf{u}_s^+ . x_0 \subset \mathscr{C}'$.

Now take $s \in E_\infty$, we have seen that $\mathbf{a}^k_{\frac{1}{2}\log N} \mathbf{u}_s^+ . x_0 \subset \mathscr{C}'$ for every $k \in \mathbb{Z}^+$. Thus $\mathbf{a}_{\geq 0} \mathbf{u}_s^+ . x_0 \subset \mathbf{a}_{[0, \frac{1}{2}\log N]} . \mathscr{C}'$ is bounded. By Dani correspondence, s is badly approximable. $\square$

Note that actually **BAD** can be written as a countable union of such E_∞'s. And we need to choose $N, \mathscr{C}$ such that condition (2) from Lemma 4.3 holds for δ fixed but N tends to infinity.

4.8. **Consequence of mixing.** The following is Theorem 3.5 in the Appendix to Lecture 2.

Theorem 4.7. *Fix $y_0 \in \mathrm{X}_2$, $\varepsilon_0 \in (0, 1)$ and a compact subset $\mathscr{C}$ of X_2. There exist $\delta, T > 0$ and $M \in 2\mathbb{Z}^+$ such that for every $x \in \mathscr{C}$ and $T' > T$,*

$$\int_{-0.5}^{0.5} \mathbf{1}_{B_{y_0}(\varepsilon_0)}(\mathbf{a}_{T'} \mathbf{u}_t^+ . x)\, \mathrm{dt} > \delta.$$

Fix some $y_0 \in \mathrm{X}_2$ and $\varepsilon_0 \in (0, 1)$. Let $\mathscr{C}_3 := \mathbf{u}^+_{[-3,3]} . \overline{B_{y_0}(\varepsilon_0)}$. By Theorem 4.7, we find $\delta_0 > 0$, $T_0 > 0$ such that for every $T > T_0$ and $x \in \mathscr{C}_3$,

$$\mathrm{Leb}\left\{ t \in [-0.5, 0.5] \;\middle|\; \mathbf{a}_T \mathbf{u}_t^+ . x \in B_{y_0}(\varepsilon_0) \right\} > \delta_0.$$

For $x \in \mathscr{C}_2 := \mathbf{u}^+_{[-2,2]}.\overline{B_{y_0}(\varepsilon_0)}$, apply the above to $\mathbf{u}^+_{0.5}.x \in \mathscr{C}_3$, we get

$$\mathrm{Leb}\left\{t \in [0,1] \;\middle|\; \mathbf{a}_T\mathbf{u}^+_t.x \in B_{y_0}(\varepsilon_0)\right\} > \delta_0, \quad \forall T > T_0. \tag{10}$$

Apply the Cantor-like set construction to N with $\frac{1}{2}\log(N) > T_0$ and $\mathscr{C} := \mathscr{C}_1 := \mathbf{u}^+_{[-1,1]}.\overline{B_{y_0}(\varepsilon_0)}$. For simplicity write $h_N := \mathbf{a}_{\frac{1}{2}\log(N)}$.

Take $E \in \mathscr{E}_k$, we need to bound

$$\frac{\#\left\{F \in \mathscr{E}_{k+1} \mid F \subset E\right\}}{N}$$

from below. Recall that E is of the form $\sigma_{I_1} \circ ... \circ \sigma_{I_{k-1}}(I_k)$ for some $\mathbb{I} = (I_i)_{i=1}^k \subset \mathscr{I}_N$. And if $I_i = [a_i, b_i)$, we have defined

$$\mathscr{C}_1 \ni x^{\mathbb{I}}_k = h_N\mathbf{u}^+_{a_k}.x^{\mathbb{I}}_{k-1} = ... = (h_N\mathbf{u}^+_{a_k}) \cdot (h_N\mathbf{u}^+_{a_{k-1}}) \cdot ... \cdot (h_N\mathbf{u}^+_{a_1}).x_0$$

where $x_0 = \mathbb{Z}^2$ is the identity coset. Moreover, we have a bijection

$$\begin{aligned} \mathscr{I}_{x^{\mathbb{I}}_k}(\mathrm{Good}) &\to \{F \in \mathscr{E}_{k+1} \mid F \subset E\} \\ I &\mapsto \sigma_{I_1} \circ ... \circ \sigma_{I_k}(I). \end{aligned}$$

Recall an interval $I = [a_I, b_I) \in \mathscr{I}_N$ is contained in $\mathscr{I}_{x^{\mathbb{I}}_k}(\mathrm{Good})$ iff $h_N\mathbf{u}^+_{a_I}.x^{\mathbb{I}}_k \in \mathscr{C}_1$. As

$$\begin{aligned} I \notin \mathscr{I}_{x^{\mathbb{I}}_k}(\mathrm{Good}) &\implies h_N\mathbf{u}^+_{a_I}.x^{\mathbb{I}}_k \notin \mathscr{C}_1 \\ &\implies \mathbf{u}^+_{[0,1]}h_N\mathbf{u}^+_{a_I}.x^{\mathbb{I}}_k \cap B_{y_0}(\varepsilon_0) = \emptyset \\ &\iff h_N\mathbf{u}^+_{I}.x^{\mathbb{I}}_k \cap B_{y_0}(\varepsilon_0) = \emptyset. \end{aligned}$$

Thus,

$$\mathrm{Leb}\left\{t \in [0,1] \;\middle|\; \mathbf{a}_T\mathbf{u}^+_t.x \notin B_{y_0}(\varepsilon_0)\right\} > 1 - \frac{\#\mathscr{I}_{x^{\mathbb{I}}_k}(\mathrm{Good})}{N}.$$

Combined with Equa.(10) (note that $\frac{1}{2}\log(N) > T_0$),

$$\begin{aligned} &1 - \delta_0 > 1 - \frac{\#\mathscr{I}_{x^{\mathbb{I}}_k}(\mathrm{Good})}{N} \\ \implies &\frac{\#\left\{F \in \mathscr{E}_{k+1} \mid F \subset E\right\}}{N} = \frac{\#\mathscr{I}_{x^{\mathbb{I}}_k}(\mathrm{Good})}{N} > \delta_0. \end{aligned}$$

By Lemma 4.3 and Lemma 4.6, we have

$$\dim_H(\mathbf{BAD}) \geq \dim_H(E_\infty) \geq 1 - \frac{\log(\delta_0^{-1})}{\log(N)}.$$

Letting $N \to +\infty$, we get

$$\dim_H(\mathbf{BAD}) \geq 1.$$

Remark 4.8. *You can also show that* $\dim_H(E_\infty)$ *is strictly smaller than* 1.

4.9. **Exercises.** Let $X := \{0,1\}^{\mathbb{Z}_{\geq 0}}$. Equipped with the product topology, X is compact. In concrete terms, a sequence $(x^k_\star)$ converges to $x_\star$ iff for every i, (x^k_i) converges to x_i as k tends to infinity. We usually denote an element of X by $x_\star$ or $(x_0x_1x_2....)$. Define a metric on X by

$$d(x_\star, y_\star) := 2^{-\inf\left\{n \in \mathbb{Z}_{\geq 0} \;\middle|\; x_n \neq y_n\right\}}.$$

For instance $d(0101..., 1000....) = 2^{-0} = 1$. This metric is compatible with the product topology. Given $\mathbf{a} := (a_0, a_1, ..., a_{l-1}) \in \{0,1\}^l$ (we will refer to such things as **words of length** l), let

$$C_{\mathbf{a}} := \{x_\star \in X \mid x_i = a_i, \text{ for } i = 0, ..., l-1.\}$$

For instance, $C_{(0,1)} = \{x_\star \in X \mid x_0 = 0, x_1 = 1\}$.

Finally, define a continuous map $\sigma : X \to X$ by $\sigma(x)_i := x_{i+1}$.

Exercise A. *Let X be equipped with the metric defined above,* $\dim_H(X) = 1$.

Exercise B. *Let X be equipped with the metric defined above,* $\dim_\square(X) = 1$.

5. Lecture 4, Space of Lattices in $\mathbb{R}^3$, Rigidity Conjectures and an Isolation Principle

5.1. **Prelude.** The connection between Littlewood conjecture (and Oppenheim conjecture) and subgroup action on the space of lattices has been (at least implicitly) noted in Cassels–Swinnerton-Dyer's paper [CSD55] in 1950s. Whereas Oppenheim conjecture is now a theorem of Margulis, Littlewood conjecture remains unsolved. One contribution of the CSD paper is an "isolation principle". This can be used to establish implications between several (unknown) conjectures. Also, it can be used to show that Littlewood conjecture holds for pairs of numbers contained in the same cubic number field. One may also consult [Mar97] and [LW01].

5.2. **Space of unimodular lattices.**

5.2.1. *The definition.* Just as before, we say that a discrete subgroup Λ of $(\mathbb{R}^3,+)$ is a lattice iff $\Lambda = \mathbb{Z}\mathbf{u}+\mathbb{Z}\mathbf{v}+\mathbf{w}$ for some linearly independent $\mathbf{u},\mathbf{v},\mathbf{w}\in\mathbb{R}^3$. For such a lattice $\Lambda = \mathbb{Z}\mathbf{u}+\mathbb{Z}\mathbf{v}+\mathbb{Z}\mathbf{w}$, define $\|\Lambda\| := |\det((\mathbf{u},\mathbf{v},\mathbf{w}))|$. A lattice Λ is said to be unimodular iff $\|\Lambda\|=1$.

Let X_3 denote the set of all unimodular lattices of $\mathbb{R}^3$ equipped with the Chabauty topology. Under this topology, a sequence of unimodular lattices (Λ_n) converges to $\Lambda\in\mathrm{X}_3$ iff we can write $\Lambda_n = \mathbb{Z}\mathbf{u}_n+\mathbb{Z}\mathbf{v}_n+\mathbb{Z}\mathbf{w}_n$ and $\Lambda = \mathbb{Z}\mathbf{u}+\mathbb{Z}\mathbf{v}+\mathbb{Z}\mathbf{w}$ such that $\|\mathbf{u}_n-\mathbf{u}\|, \|\mathbf{v}_n-\mathbf{v}\|, \|\mathbf{w}_n-\mathbf{w}\|\to 0$.

5.2.2. *Mahler's criterion.* Define

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

Theorem 5.1. *The function* $\mathrm{sys}:\mathrm{X}_3\to(0,+\infty)$ *is bounded, continuous and proper.*

5.2.3. *Group action and local coordinates.* The group $\mathbf{SL}_3(\mathbb{R})$, consisting of all 3-by-3 real matrices of determinant one, acts on X_3 naturally. This action is continuous and transitive. The map $g\mapsto g.\mathbb{Z}^3$ induces a homeomorphism $\mathbf{SL}_3(\mathbb{R})/\mathbf{SL}_3(\mathbb{Z})\cong\mathrm{X}_3$.

We define several subgroups of $\mathbf{SL}_3(\mathbb{R})$:

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

$$\mathrm{U}^{++} := \left\{ \begin{bmatrix} 1 & u_{12} & u_{13}\\ 0 & 1 & u_{23}\\ 0 & 0 & 1\end{bmatrix} \middle| u_{ij}\in\mathbb{R}\right\};\quad \mathrm{U}^{--} := \left\{ \begin{bmatrix} 1 & 0 & 0\\ u_{21} & 1 & 0\\ u_{31} & u_{32} & 1\end{bmatrix} \middle| u_{ij}\in\mathbb{R}\right\}$$

For $s\in\mathbb{R}$ and $i\neq j=1,2,3$, let E_{ij} be the matrix whose (i,j)-th entry is 1 and is zero elsewhere. Let

$$\mathrm{U}_{ij} := \{\mathrm{I}_3 + sE_{ij} \mid s\in\mathbb{R}\},$$

a subgroup of $\mathbf{SL}_3(\mathbb{R})$ isomorphic to $(\mathbb{R},+)$.

For $\varepsilon>0$, let

$$\mathrm{A}(\varepsilon) := \left\{ \begin{bmatrix} e^{t_1} & 0 & 0\\ 0 & e^{t_2} & 0\\ 0 & 0 & e^{t_3}\end{bmatrix} \middle| \sum t_i = 0,\ t_1,t_2\in(-\varepsilon,\varepsilon)\right\};$$

$$\mathrm{U}^{++}(\varepsilon) := \left\{ \begin{bmatrix} 1 & u_{12} & u_{13}\\ 0 & 1 & u_{23}\\ 0 & 0 & 1\end{bmatrix} \middle| u_{ij}\in(-\varepsilon,\varepsilon)\right\};$$

$$\mathrm{U}^{--}(\varepsilon) := \left\{ \begin{bmatrix} 1 & 0 & 0\\ u_{21} & 1 & 0\\ u_{31} & u_{32} & 1\end{bmatrix} \middle| u_{ij}\in(-\varepsilon,\varepsilon)\right\}.$$

5.2.4. *Local coordinates.* For $x \in \mathrm{X}_3$, let $\mathrm{Obt}_x : \mathbf{SL}_3(\mathbb{R}) \to \mathrm{X}_3$ be the orbit map $g \mapsto g.x$. For every compact subset $\mathscr{C} \subset \mathrm{X}_3$, there exists $\varepsilon > 0$ such that for every $x \in \mathscr{C}$,

$$\begin{aligned} \mathrm{A}(\varepsilon) \times \mathrm{U}^{--}(\varepsilon) \times \mathrm{U}^{++}(\varepsilon) &\to \mathrm{X}_3 \\ (a, v, u) &\mapsto \mathrm{Obt}(a \cdot v \cdot u).x \end{aligned} \tag{11}$$

is a homeomorphism onto an open neighborhood, termed $\mathcal{N}_x^{AU}(\varepsilon)$, of $x \in \mathrm{X}_3$. Likewise, for $\varepsilon > 0$ small enough, we define $\mathcal{N}_x^{UA}(\varepsilon)$ by using $\mathrm{Obt}_x(v \cdot u \cdot a) = vua.x$ for $u \in \mathrm{U}^{--}$, $v \in \mathrm{U}^{++}$ and $a \in \mathrm{A}$.

5.2.5. *A metric.* One can define a right-invariant metric[11] on $\mathbf{SL}_3(\mathbb{R})$ by

$$d^{\mathbf{SL}_3(\mathbb{R})}(g, h) := \log\left(1 + \left\|gh^{-1}\right\|_{\mathrm{op}} + \left\|hg^{-1}\right\|_{\mathrm{op}}\right).$$

The metric topology is the usual topology on $\mathbf{SL}_3(\mathbb{R})$ (namely the subspace topology induced from $\mathbb{R}^9$). It induces a metric on $\mathbf{SL}_3(\mathbb{R})/\,\mathbf{SL}_3(\mathbb{Z}) \cong \mathrm{X}_3$ by

$$d^{\mathrm{X}_3}(g\mathbb{Z}^3, h\mathbb{Z}^3) := \inf_{\gamma_1, \gamma_2 \in \mathbf{SL}_3(\mathbb{Z})} d^{\mathbf{SL}_3(\mathbb{R})}(g\gamma_1, h\gamma_2).$$

This metric is compatible with the topology given. For ε small enough depending on some compact set $\mathscr{C}$, the orbit map $g \mapsto g.x$ is an isometry (and in particular, a homeomorphism) from $B(\varepsilon) := \{g,\ d(g, \mathrm{I}_3) < \varepsilon\}$ to its image for every $x \in \mathscr{C}$.

5.2.6. *The invariant measure.* The group $\mathbf{SL}_3(\mathbb{R})$ has a bi-invariant locally finite measure $\mathrm{m}_{\mathbf{SL}_3(\mathbb{R})}$. After being normalized by a positive scalar, it induces an $\mathbf{SL}_3(\mathbb{R})$-invariant probability measure $\mathrm{m}_{\mathrm{X}_3}$ on X_3. For $\varepsilon > 0$ small enough, the orbit map $g \mapsto g.x$ identifies the measure $\mathrm{m}_{\mathbf{SL}_3(\mathbb{R})}$ restricted to $B(\varepsilon)$ with $\mathrm{m}_{\mathrm{X}_3}$ restricted to $B(\varepsilon).x$.

5.3. **Two problems in Diophantine approximations.** For a function $f : \mathbb{R}^3 \to \mathbb{R}$, let

$$m^*(f) := \inf\left\{|f(x)| \;\middle|\; x \in \mathbb{Z}^3,\ x \neq \mathbf{0}\right\}.$$

First we consider real quadratic forms in three variables. Let $Q : \mathbb{R}^3 \to \mathbb{R}$ be such a form. So there are real numbers $(q_{ij})_{i,j=1,2,3}$ with $q_{ij} = q_{ji}$ such that $Q(x_1, ..., x_n) = \sum q_{ij} x_i x_j$.

Theorem 5.2. *Assume Q is non-degenerate (that is, $\det(q_{ij}) \neq 0$). If Q is indefinite and is not a scalar multiple of one with $\mathbb{Q}$-coefficients, then $m^*(f) = 0$.*

Remark 5.3. *Analogous statement for quadratic forms in two-variable is false. It is true if the number of variable is greater than three, which can be reduced to the above case.*

Let $\phi : \mathbb{R}^3 \to \mathbb{R}$ be a product of three real linear forms. Namely, there exist three L_1, L_2, L_3 linear functionals on $\mathbb{R}^3$ such that $\phi(x) = L_1(x)L_2(x)L_3(x)$.

Conjecture 5.4. *Assume that ϕ is non-degenerate (namely, L_1, L_2, L_3 are linearly independent) and is not a scalar multiple of one with $\mathbb{Q}$-coefficients. Then $m^*(\phi) = 0$.*

5.4. **Linear symmetry.** Let Q, ϕ be as in the last section. Let

$$H_Q := \mathbf{SO}_Q(\mathbb{R}) := \left\{g \in \mathbf{SL}_3(\mathbb{R}) \;\middle|\; Q(g.x) = Q(x),\ \forall\, x \in \mathbb{R}^3\right\}.$$

$$H_{|\phi|} := \left\{g \in \mathbf{SL}_3(\mathbb{R}) \;\middle|\; \phi(g.x) = \phi(x),\ \forall\, x \in \mathbb{R}^3\right\}.$$

Lemma 5.5. *We have*

$$H_Q.\mathbb{Z}^3 \text{ is unbounded in } \mathrm{X}_3 \implies m^*(Q) = 0,$$

$$H_{|\phi|}.\mathbb{Z}^3 \text{ is unbounded in } \mathrm{X}_3 \implies m^*(\phi) = 0.$$

With some Galois theory, it can be shown that

Lemma 5.6. *We have*

$$H_Q.\mathbb{Z}^3 \text{ is compact in } \mathrm{X}_3 \implies \text{up to a scalar, } Q \text{ has rational coefficients},$$

$$H_{|\phi|}.\mathbb{Z}^3 \text{ is compact in } \mathrm{X}_3 \implies \text{up to a scalar, } \phi \text{ has rational coefficients}.$$

Theorem 5.7. *Assume Q is non-degenerate and indefinite (over $\mathbb{R}$). Every bounded orbit of H_Q on X_3 is closed (and hence compact).*

Conjecture 5.8. *Every bounded orbit of A (which is of finite index in $H_{|xyz|}$) on X_3 is closed (and hence compact).*

[11] the sup-index on $d^{\mathbf{SL}_3(\mathbb{R})}$ is dropped when it is clear which space we are referring to.

By the lemmas above, we have

Corollary 5.9. *Conjecture 5.8 $\implies$ Conjecture 5.4. And Theorem 5.7 $\implies$ Theorem 5.2.*

5.5. **Measure rigidity.** How to prove Theorem 5.7? A crucial fact is that the symmetry group $\mathbf{SO}_Q(\mathbb{R})$, locally isomorphic to $\mathbf{SL}_2(\mathbb{R})$, is generated by unipotent matrices. Though the original proof of Theorem 5.7 does not involve any measures, it is possible to decompose the proof of Theorem 5.7 into two steps:

1. Classification of unipotent-invariant ergodic measures: they are all homogeneous;
2. Deduce Theorem 5.7 from this.

Regarding A-action, the measure classification is unknown:

Conjecture 5.10. *Every* A*-invariant probability measure on* X_3 *is a convex combination of those supported on compact* A*-orbits and* $\mathrm{m}_{\mathrm{X}_3}$*.*

Conjecture 5.11. *Every* A*-invariant compact subset of* X_3 *is a union of finitely many compact* A*-orbits.*

Conjecture 5.12. *Every bounded subset of* X_3 *contains only finitely many compact* A*-orbits.*

We do know the following implications

Theorem 5.13. *Conjecture 5.10 $\implies$ Conjecture 5.8 $\implies$ Conjecture 5.11 $\implies$ Conjecture 5.12.*

Theorem 5.14. *Conjecture 5.11 $\implies$ Littlewood conjecture.*

The proof of these implications is based on the following "isolation principle".

Theorem 5.15. *Given a compact* A*-orbit* $\mathrm{A}.y$ *on* X_3*. For every compact subset* $\mathscr{C} \subset \mathrm{X}_3$*, there exists* $\varepsilon > 0$ *such that*

$$\mathrm{dist}(x, y) < \varepsilon \implies \mathrm{A}.x \not\subseteq \mathscr{C}$$

In particular, if the orbit closure of some A*-orbit* $\mathrm{A}.x$ *contains a compact* A*-orbit, then* $\mathrm{A}.x$ *is either compact or unbounded.*

Remark 5.16. *This (and all the conjectures above) is wrong on* X_2 *where* A*, isomorphic to* $(\mathbb{R}, +)$*, has "rank one".*

Conjecture 5.10 seems to be partly motivated by a question of Furstenberg [Fur67]. Let $T_p : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ defined by $x + \mathbb{Z} \mapsto px + \mathbb{Z}$. Note that there are many irrational numbers such that $\{T_p^n(\alpha + \mathbb{Z}),\ n \in \mathbb{Z}^+\}$ is not dense in $\mathbb{R}/\mathbb{Z}$.

Theorem 5.17. *If* α *is irrational, then*

$$\left\{T_2^n T_3^m.\alpha \;\middle|\; n, m \in \mathbb{Z}^+\right\}$$

is dense in $\mathbb{R}/\mathbb{Z}$*.*

Conjecture 5.18. *Let* μ *be a probability measure on* $\mathbb{R}/\mathbb{Z}$ *invariant under* T_2 *and* T_3*, then* μ *is a convex combinations of those supported on certain finite sets and the Lebesgue measure.*

What we know about Conjecture 5.10 is

Theorem 5.19. *Let* μ *be an* A*-invariant probability measure with compact support, then* $h_\mu(a) = 0$ *for every* $a \in \mathrm{A}$*.*

This may be compared with (see [Rud90])

Theorem 5.20. *Let* μ *be an ergodic probability measure on* $\mathbb{R}/\mathbb{Z}$ *invariant under* T_2 *and* T_3 *and* $h_\mu(T_2) > 0$*, then* μ *is the Lebesgue measure.*

Applications of measure rigidity theorems can be found in the survey [Ein10] or [Lin22].

5.6. **Compact A-orbits.** In this section we give a more explicit description of compact A-orbits.

Lemma 5.21. *Let* $\mathrm{A}g\mathbb{Z}^3$ *be a compact* A *orbit. Then there exists a cubic number field (i.e. field extension of* $\mathbb{Q}$ *of degree three)* K*,* $(x,y,z)\in K^3$ *and* $\lambda\in\mathbb{R}$ *such that* $\mathrm{A}g\mathbb{Z}^3=\mathrm{A}M\mathbb{Z}^3$ *for*

$$M=\lambda\cdot\begin{bmatrix} x & y & z \\ \sigma_2(x) & \sigma_2(y) & \sigma_2(z) \\ \sigma_3(x) & \sigma_3(y) & \sigma_3(z)\end{bmatrix}\in\mathbf{SL}_3(\mathbb{R}) \tag{12}$$

where $\{\mathrm{id},\sigma_2,\sigma_3\}$ *denotes the three field embeddings of* K *into* $\mathbb{C}$*.*

Lemma 5.22. *Assume* $\gamma\in\mathbf{SL}_3(\mathbb{Z})$ *is diagonalizable and none of the eigenvalues are equal to* ±1*. Then its characteristic polynomial is irreducible in* $\mathbb{Q}[x]$*.*

Proof. Let $p(x)\in\mathbb{Z}[X]:=\det(x\mathrm{I}_3-\gamma)$ be the characteristic polynomial of γ. It suffices to show that $p(x)$ is irreducible in $\mathbb{Z}[x]$ as it is monic (Gauss' lemma?). Otherwise,

$$p(x)=(x^2+ax+b)(x+c),\quad \exists\, a,b,c\in\mathbb{Z}$$

Since $\det(\gamma)=1$, $bc=1$. So $c=\pm1$, a contradiction. $\square$

Proof of Lemma 5.21. By assumption, $\mathrm{A}g\,\mathbf{SL}_3(\mathbb{Z})/\,\mathbf{SL}_3(\mathbb{Z})$ is compact. In other words, $\mathrm{A}\cap g\,\mathbf{SL}_3(\mathbb{Z})g^{-1}$ is a lattice in A. Therefore, we can find $\gamma\in g^{-1}\mathrm{A}g\cap\mathbf{SL}_3(\mathbb{Z})$ with three distinct eigenvalues and none of which is equal to ±1. Let $p(x)$ be the characteristic polynomial of γ, then $p(x)$ is irreducible by lemma above. Let θ be one of its root. Then $K:=\mathbb{Q}(x)$, isomorphic to $\mathbb{Q}[x]/(p(x))$, has dimension three as a $\mathbb{Q}$-vector space. So there exists exactly three different embeddings $\{\mathrm{id},\sigma_2,\sigma_3\}$ of K into $\mathbb{C}$. By linear algebra, one can find $(x,y,z)\in K^3$ with

$$(x,y,z)\cdot\gamma=\theta(x,y,z)$$

By applying the other two embeddings, we get

$$\begin{bmatrix} x & y & z \\ \sigma_2(x) & \sigma_2(y) & \sigma_2(z) \\ \sigma_3(x) & \sigma_3(y) & \sigma_3(z)\end{bmatrix}\cdot\gamma=\begin{bmatrix} \theta & 0 & 0 \\ 0 & \sigma_2(\theta) & 0 \\ 0 & 0 & \sigma_3(\theta)\end{bmatrix}\cdot\begin{bmatrix} x & y & z \\ \sigma_2(x) & \sigma_2(y) & \sigma_2(z) \\ \sigma_3(x) & \sigma_3(y) & \sigma_3(z)\end{bmatrix}$$

Define M as in Equa.(12) where λ is chosen such that M has determinant one. Then $M\gamma M^{-1}$, as well as $g\gamma g^{-1}$, belongs to A. Replacing θ by $\sigma_i(\theta)$ and K by $\sigma_i(K)$ if necessary, we assume that

$$M\gamma M^{-1}=g\gamma g^{-1}.$$

Consequently, gM^{-1} commutes with $M\gamma M^{-1}$ and is therefore diagonal. In particular, $\mathrm{A}g\mathbb{Z}^3=\mathrm{A}M.\mathbb{Z}^3$. This finishes the proof. $\square$

5.7. **An equivalent form of Littlewood conjecture.** Let

$$\mathrm{A}^+:=\left\{\begin{bmatrix} e^{t_1} & 0 & 0 \\ 0 & e^{t_2} & 0 \\ 0 & 0 & e^{t_3}\end{bmatrix}\;\middle|\;\sum t_i=0,\ t_1,t_2>0\right\}$$

be a sub-semigroup of A.

For a pair of real numbers $(\alpha,\beta)\in\mathbb{R}^2$, let

$$\Lambda_{\alpha,\beta}:=\begin{bmatrix} 1 & 0 & \alpha \\ 0 & 1 & \beta \\ 0 & 0 & 1\end{bmatrix}.\mathbb{Z}^3=\mathbb{Z}\begin{bmatrix}1\\0\\0\end{bmatrix}+\mathbb{Z}\begin{bmatrix}0\\1\\0\end{bmatrix}+\mathbb{Z}\begin{bmatrix}\alpha\\\beta\\1\end{bmatrix}.$$

Lemma 5.23. *Let* $(\alpha,\beta)\in\mathbb{R}^2$*. The following two are equivalent*

(1) $\mathrm{A}^+.\Lambda_{\alpha,\beta}$ *is unbounded in* X_3*;*
(2) (α,β) *satisfies Littlewood conjecture.*

Proof. From definition we have

$$\begin{bmatrix} e^{t_1} & 0 & 0 \\ 0 & e^{t_2} & 0 \\ 0 & 0 & e^{t_3}\end{bmatrix}\Lambda_{\alpha,\beta}=\left\{\begin{bmatrix} e^{t_1}(l+n\alpha) \\ e^{t_2}(m+n\beta) \\ e^{-t_1-t_2}n\end{bmatrix}\;\middle|\; l,m,n\in\mathbb{Z}\right\}$$

Take $\varepsilon\in(0,1)$.

If $(t_1, t_2) \in (\mathbb{R}^+)^2$ is such that $\mathrm{sys}(\mathbf{a}_{t_1,t_2}\Lambda_{\alpha,\beta}) < \varepsilon$, then we can find $(l, m, n) \in \mathbb{Z}^3 \setminus \{\mathbf{0}\}$ such that

$$\left.\begin{array}{r} |e^{t_1}(l+n\alpha)| < \varepsilon \\ |e^{t_2}(m+n\beta)| < \varepsilon \\ |e^{-t_1-t_2}n| < \varepsilon \end{array}\right\} \implies \left\{\begin{array}{l} |n|\,|l+n\alpha|\,|m+n\beta| < \varepsilon^3, \\ n \neq 0 \end{array}\right.$$

Hence $n \neq 0$ and $|n|\,\langle n\alpha\rangle\langle n\beta\rangle < \varepsilon^3$.

Conversely, let $n \in \mathbb{Z}_{\neq 0}$ be such that $|n|\,\langle n\alpha\rangle\langle n\beta\rangle < \varepsilon^3$. Then one finds l, m such that $\langle n\alpha\rangle = |l + n\alpha|$ and $\langle n\beta\rangle = |m + n\beta|$. Assume $l + n\alpha \neq 0$ and $m + n\beta \neq 0$ (the remaining cases are left to the reader). We wish to set $t_1, t_2 \in \mathbb{R}$ such that

$$e^{t_1} = \frac{\varepsilon}{|l+n\alpha|},\ e^{t_2} = \frac{\varepsilon}{|m+n\alpha|}. \tag{13}$$

But there is no guarantee that $t_1, t_2 > 0$, which happens exactly when one of $\langle n\alpha\rangle$ or $\langle n\alpha\rangle$ is larger than ε. This can be remedied as follows:

Say $\langle n\beta\rangle > \varepsilon$. By Dirichlet theorem, we can find $n_2 < \lceil \varepsilon^{-1}\rceil$ such that

$$\langle n_2 n\beta\rangle < \lceil \varepsilon^{-1}\rceil^{-1} < \varepsilon.$$

On the other hand,

$$|n|\,\langle n_2 n\alpha\rangle \leq |nn_2|\,\langle n\alpha\rangle < (\varepsilon^{-1}+1)\varepsilon^2 < 2\varepsilon.$$

Thus, if replacing n by $n' := nn_2$ and ε by $\varepsilon' := \sqrt[3]{2\varepsilon^2}$, we would have t_1, t_2 as defined by Equa.(13) are both positive. One has

$$\left|e^{t_1}(l+n\alpha)\right| = \varepsilon',\ \left|e^{t_2}(m+n\beta)\right| = \varepsilon',\ \left|e^{-t_1-t_2}n\right| < \varepsilon'.$$

And the proof is complete. □

5.8. **Conjecture 5.11 implies Littlewood.** By Lemma 5.23, it suffices to show that $\mathrm{A}^+.\Lambda_{\alpha,\beta}$ is not bounded. So let us assume that it is and seek for a contradiction.

Define

$$Y := \left\{y \in \mathrm{X}_3 \;\middle|\; y = \lim \mathbf{a}_{(s_n,t_n)}.\Lambda_{\alpha,\beta},\ \exists\, s_n, t_n \to +\infty\right\}.$$

Then Y is A-invariant and bounded. Let $\overline{Y}$ be its closure, which is also A-invariant. By Conjecture 5.23, $\overline{Y}$ is a finite union of compact A-orbits. Therefore, Y is also a finite union of compact A-orbits, say

$$Y = \mathrm{A}y_1 \sqcup \mathrm{A}y_2 \sqcup ... \sqcup \mathrm{A}y_k.$$

Choose $\varepsilon > 0$ small enough such that $\mathcal{N}_{\mathrm{A}y_i}(\varepsilon)$ for $i = 1, ..., k$ are disjoint from each other. On the other hand, by the definition of Y, there exists $T(\varepsilon) \in \mathbb{R}^+$ such that

$$Y_N := \left\{\mathbf{a}_{(s,t)} \;\middle|\; s, t > T(\varepsilon)\right\} \subset \bigsqcup_{i=1}^{k} \mathcal{N}_{\mathrm{A}y_i}(\varepsilon).$$

But Y_N is connected, it has to be contained in a unique $\mathcal{N}_{\mathrm{A}y_i}(\varepsilon)$. In other words, $k = 1$ and $Y = \mathrm{A}.y_1$.

Using local coordinates, one shows that

Lemma 5.24. *For $\varepsilon > 0$ small enough, the map*

$$\begin{aligned} \mathrm{U}^{--}(\varepsilon) \times \mathrm{U}^{++}(\varepsilon) \times \mathrm{A}.y_1 &\to \mathrm{X}_3 \\ (v, u, a.y_1) &\mapsto vua.y_1 \end{aligned}$$

is a homeomorphism onto an open subset, called $\mathcal{N}^{UA}_{\mathrm{A}.y_1}(\varepsilon)$.

Choose $\varepsilon > 0$ small enough according to this lemma and find N large enough such that $Y_N \subset \mathcal{N}^{UA}_{\mathrm{A}.y_1}(0.5\varepsilon)$. Note that Y_N is A^+-invariant, so we can analyze Y_N under the action of A^+ using these local coordinates. For $z = \mathbf{u}^{--}(z)\mathbf{u}^{++}(z)y_z \in Y_N$ for some $y_z \in \mathrm{A}.y_1$ and $a \in \mathrm{A}^+$,

$$a.z = (a\mathbf{u}^{--}(z)a^{-1}) \cdot (a\mathbf{u}^{++}(z)a^{-1}) \cdot a.y_z.$$

If $\mathbf{u}^{++}(z) \neq \mathrm{I}_3$, then one can find $a \in \mathrm{A}^+$ such that $a\mathbf{u}^{++}(z)a^{-1} \in \mathrm{U}_\varepsilon \setminus \mathrm{U}_{0.5\varepsilon}$. This is a contradiction. Likewise, we also have that the $(2,1)$-entry of $\mathbf{u}^{--}(z)$ is zero. Combined with Lemma 5.21, we get

$$\begin{bmatrix} 1 & \alpha & \beta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} e^{t_1} & 0 & 0 \\ 0 & e^{t_2} & 0 \\ 0 & 0 & e^{t_3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ r_1 & r_2 & 1 \end{bmatrix} \begin{bmatrix} x & y & z \\ \sigma_2(x) & \sigma_2(y) & \sigma_2(z) \\ \sigma_3(x) & \sigma_3(y) & \sigma_3(z) \end{bmatrix} \cdot \gamma$$

for some

$$\gamma \in \mathbf{SL}_3(\mathbb{Z}),\ t_1, t_2, t_3, r_1, r_2 \in \mathbb{R},\ x, y, z \in \text{some cubic number field } K.$$

Hence

$$\begin{bmatrix} 1 & \alpha & \beta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} e^{t_1} & 0 & 0 \\ 0 & e^{t_2} & 0 \\ 0 & 0 & e^{t_3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ r_1 & r_2 & 1 \end{bmatrix} \begin{bmatrix} x' & y' & z' \\ \sigma_2(x') & \sigma_2(y') & \sigma_2(z') \\ \sigma_3(x') & \sigma_3(y') & \sigma_3(z') \end{bmatrix}$$

for some possibly different $x', y', z' \in K$. By comparing the second row of both sides, one sees that $x' = z' = 0$, which is a contradiction.

5.9. **Conjecture 5.8 implies Conjecture 5.11.** Assume otherwise, then we can find infinitely many distinct compact A-orbits $\mathrm{A}.y_1, \mathrm{A}y_2, ...$ contained in some fixed compact subset $\mathscr{C} \subset \mathrm{X}_3$. Let y be a limit point of $(\mathrm{A}.y_i)_i$. Then $\mathrm{A}.y$ is contained in $\mathscr{C}$. By Conjecture 5.8, $\mathrm{A}.y$ is closed. By Theorem 5.15, for $z \in \mathrm{X}_3$ that is close enough to y, $\mathrm{A}.z$ compact implies that it can not be contained in $\mathscr{C}$. This is a contradiction.

5.10. **Ergodic decomposition.** Let μ be a Borel probability measure on X_3. We say that μ is A-ergodic iff every A-invariant Borel subset has μ-measure zero or one.

Lemma 5.25. *Let μ be a Borel probability measure on X_3. The following three are equivalent:*

1. *μ is A-ergodic;*
2. *every A-invariant L^1-function is constant almost everywhere;*
3. *If $\mu = \nu_1 + (1-\lambda)\nu_2$ for some $\lambda \in [0,1]$ and ν_1, ν_2 are A-invariant probability measure, then $\lambda = 0$ or 1.*

Let $\mathrm{Prob}(\mathrm{X}_3)^{\mathrm{A}}$ be the set of A-invariant Borel probability measures on X_3 equipped with the weak-$*$ topology. And let $\mathrm{Prob}(\mathrm{X}_3)^{\mathrm{A,erg}}$ be those ergodic ones.

Theorem 5.26 (Ergodic decomposition)**.** *For every $\mu \in \mathrm{Prob}(\mathrm{X}_3)^{\mathrm{A}}$, there exists a probability measure λ_μ on $\mathrm{Prob}(\mathrm{X}_3)^{\mathrm{A}}$ with $\lambda_\mu(\mathrm{Prob}(\mathrm{X}_3)^{\mathrm{A,erg}}) = 1$ such that*

$$\mu = \int_{\mathrm{Prob}(\mathrm{X}_3)^{\mathrm{A}}} \nu \, \lambda_\mu(\nu).$$

More explicitly, for a compactly supported continuous function $f : \mathrm{X} \to \mathbb{R}$, let φ_f be the continuous function on $\mathrm{Prob}(\mathrm{X}_3)^{\mathrm{A}}$ defined by $\varphi(\nu) = \int f(x)\nu(x)$. Then

$$\int f(x)\mu(x) = \int_{\mathrm{Prob}(\mathrm{X}_3)^{\mathrm{A}}} \varphi_f(\nu) \, \lambda_\mu(\nu).$$

Remark 5.27. *This can be deduced from Choquet's theorem. A quick proof for the case needed can be found in* [Phe01].

5.11. **Conjecture 5.10 implies Conjecture 5.8.** So take $\mathrm{A}.x$ to be a bounded A-orbit. For $T > 0$, define

$$\mu_T := \frac{1}{4T^2} \int_{-T}^{T} \int_{-T}^{T} (\mathbf{a}_{(s,t)})_* \delta_x \, \mathrm{ds} \, \mathrm{dt} \in \mathrm{Prob}(\mathrm{X}_3).$$

Since $\mathrm{A}.x$ is bounded, by passing to a subsequence, we assume $\lim_n \mu_{T_n}$ exists in $\mathrm{Prob}(\mathrm{X}_3)$. Let μ denote this limit. Then μ is A-invariant. By ergodic decomposition

$$\mu = \int_{\mathrm{Prob}(\mathrm{X}_3)^{\mathrm{A,erg}}} \nu \, \lambda_\mu(\nu).$$

Now Conjecture 5.10 says that

$$\mathrm{Prob}(\mathrm{X}_3)^{\mathrm{A,erg}} = \{\mathrm{m}_{\mathrm{A}.y},\ \mathrm{A}.y \text{ compact }\} \sqcup \{\mathrm{m}_{\mathrm{X}_3}\}.$$

As $\mathrm{m}_{\mathrm{X}_3}$ has unbounded support, λ_μ must put positive mass on certain $\mathrm{m}_{\mathrm{A}.y}$ with $\mathrm{A}.y$ compact. In particular, $\mathrm{A}.x$ contains some compact A-orbit in its closure. By Theorem 5.15, $\mathrm{A}.x$, being bounded, must be compact.

5.12. **Proof of Theorem 5.15.** Assume otherwise, namely, there exist a compact subset $\mathscr{C} \subset \mathrm{X}_3$ and a sequence $(x_n) \subset \mathrm{X}_3$ converging to $y \in \mathrm{X}_3$ such that $\mathrm{A}.x_n$ is contained in $\mathscr{C}$ for every n, $\mathrm{A}.y$ is compact and $\mathrm{A}.x_n \neq \mathrm{A}.y$ for every n.

5.12.1. *Exponential "blow-up".* Fix $\varepsilon_0 > 0$ such that the conclusion of Lemma 5.24 holds. For n large enough such that $x_n \in \mathcal{N}^{UA}_{\mathrm{A}.y}(0.5\varepsilon_0)$,

$$x_n = \mathbf{u}^{--}(x_n)\mathbf{u}^{++}(x_n).y(x), \quad \mathbf{u}^{--}(x_n) \in \mathrm{U}^{--}(0.5\varepsilon_0),\ \mathbf{u}^{++}(x_n) \in \mathrm{U}^{++}(0.5\varepsilon_0),\ y(x) \in \mathrm{A}.y.$$

Now we look at

$$\max \left\{ \begin{array}{l} |(\mathbf{u}^{--}(x_n))_{21}|,\ |(\mathbf{u}^{--}(x_n))_{31}|,\ |(\mathbf{u}^{--}(x_n))_{32}|, \\ |(\mathbf{u}^{--}(x_n))_{12}|,\ |(\mathbf{u}^{--}(x_n))_{13}|\ ,\ |(\mathbf{u}^{--}(x_n))_{23}| \end{array} \right\} \tag{14}$$

Without loss of generality, we are going to assume, by passing to a subsequence, that the maximum above is taken by $|(\mathbf{u}^{--}(x_n))_{12}|$ for all n and that $(\mathbf{u}^{--}(x_n))_{12} > 0$ for all n.

Let

$$\beta_t := \begin{bmatrix} t & 0 & 0 \\ 0 & t^{-1} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Choose $t_n > 0$ such that

$$t_n^2 \cdot \left(\mathbf{u}^{--}(x_n)\right)_{12} = 0.5\varepsilon_0$$

Then $\beta_{t_n}.x_n$ stays inside the neighborhood $\mathcal{N}^{UA}_{\mathrm{A}.y}(0.5\varepsilon_0)$. Let $\varepsilon_n = \frac{0.5\varepsilon_0}{M_n^2}$ be the maximum appearing in Equa.(14). About the size of $(\mathbf{u}^{--}(\beta_{t_n}.x_n))_{ij}$ $(i \neq j)$, we have

$$\begin{bmatrix} & 0.5\varepsilon_0 & \leq \frac{0.5\varepsilon_0}{M_n} \\ \leq \frac{0.5\varepsilon_0}{M_n^4} & & \leq \frac{0.5\varepsilon_0}{M_n^3} \\ \leq \frac{0.5\varepsilon_0}{M_n^3} & \leq \frac{0.5\varepsilon_0}{M_n} & \end{bmatrix}$$

By passing to a further subsequence, assume $\beta_{t_n}.x_n$ converges to x_∞ and $\beta_{t_n}.y(x_n)$ converges to y_∞. Then we have

$$x_\infty = u_{12}(0.5\varepsilon_0).y_\infty.$$

By definition, $\mathrm{A}.x_\infty$ is contained in $\mathscr{C}$.

5.12.2. *Promotion.* Now we use a one-parameter subgroup of A that commutes with $u_{12}(\mathbb{R})$. Define

$$\alpha_t := \begin{bmatrix} t & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & t^{-2} \end{bmatrix}$$

Then

$$\alpha_t.x_\infty = u_{12}(0.5\varepsilon_0)\alpha_t.y_\infty. \tag{15}$$

Lemma 5.28. *$\{\alpha_t.y_\infty,\ t \in \mathbb{R}\}$ is dense in* $\mathrm{A}.y_\infty = \mathrm{A}.y$.

By Lemma 5.28 and Equa.(15),

$$\overline{\mathrm{A}.x_\infty} \supset u_{12}(0.5\varepsilon_0)\mathrm{A}.y$$

Using the A-invariance of the LHS, we get

$$\overline{\mathrm{A}.x_\infty} \supset u_{12}(\mathbb{R}^+)\mathrm{A}.y.$$

To get a contradiction, it suffices to show that $u_{12}(\mathbb{R}^+)\mathrm{A}.y$ (as lattices) contains arbitrarily small non-zero vectors.

Given $\varepsilon > 0$, one can find $(u,v,w)^{\mathrm{tr}} \in y$ with $u < 0$, $v > 0$. Take $t > 0$ large enough such that $|e^{-t}v| < \varepsilon$ and $|e^{-t}w| < \varepsilon$. Then take $r := \frac{e^{2t}u}{-e^{-t}v}$. One has:

$$\begin{bmatrix} 1 & r & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} e^{2t} & 0 & 0 \\ 0 & e^{-t} & 0 \\ 0 & 0 & e^{-t} \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ e^{-t}v \\ e^{-t}w \end{bmatrix},$$

which is a vector contained in the lattice $u_{12}(r)a.y$ for some $a \in \mathrm{A}$ and $r > 0$. This shows that the sys$(\cdot)$ of elements in $\overline{\mathrm{A}.x_\infty}$ could tend to 0. By the continuity of sys$(\cdot)$, $\overline{\mathrm{A}.x_\infty}$ is non-compact, a contradiction.

5.13. **Littlewood conjecture for cubic numbers.** Using a variant of the isolation principle presented above, one can show that

Theorem 5.29. *Let K be a cubic totally real number field and $\alpha, \beta \in K$. Then (α, β) satisfies the Littlewood conjecture.*

By taking transpose inverse $(\cdot)^{-\operatorname{tr}}$, one sees that $\mathrm{A}^{+}.\Lambda_{\alpha,\beta}$ is unbounded iff $\mathrm{A}^{-}.\Lambda^{'}_{\alpha,\beta}$ is unbounded where

$$\Lambda^{'}_{\alpha,\beta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\alpha & -\beta & 1 \end{bmatrix} .\mathbb{Z}^3.$$

Let $\{\sigma_1 = \mathrm{id}, \sigma_2, \sigma_3\}$ denote the three different embedding of $K \hookrightarrow \mathbb{R}$. Let

$$M_0 := \begin{bmatrix} -\sigma_3(\alpha) & -\sigma_3(\beta) & 1 \\ -\sigma_2(\alpha) & -\sigma_2(\beta) & 1 \\ -\alpha & -\beta & 1 \end{bmatrix}$$

Let $\lambda_0 \in \mathbb{R}$ such that $\det(\lambda_0 \cdot M_0) = 1$.

Lemma 5.30. $\mathrm{A}.(\lambda_0 M_0).\mathbb{Z}^3$ *is compact.*

Proof. Dirichlet's unit theorem and commensurability of lattices. □

Note that

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\alpha & -\beta & 1 \end{bmatrix} = \begin{bmatrix} t_1 & 0 & 0 \\ 0 & t_2 & 0 \\ 0 & 0 & t_3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ u_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{bmatrix} \cdot M_0$$

for some real numbers t_i, u_{ij}. Thus

$$\alpha_s.\Lambda'_{\alpha,\beta} = \begin{bmatrix} t_1 & 0 & 0 \\ 0 & t_2 & 0 \\ 0 & 0 & t_3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ u_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & u_{12} & s^3 u_{13} \\ 0 & 1 & s^3 u_{23} \\ 0 & 0 & 1 \end{bmatrix} \cdot \alpha_s M_0.\mathbb{Z}^3$$

Take some sequence $s_n \to 0$ such that $\lim \alpha_{s_n}(\lambda_0 M_0).\mathbb{Z}^3$ exists)and is equal to y_1. Then

$$x_1 := \lim \alpha_{s_n}\Lambda'_{\alpha,\beta} = \begin{bmatrix} t'_1 & 0 & 0 \\ 0 & t'_2 & 0 \\ 0 & 0 & t'_3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ u_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & u_{12} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} .y_1$$

Using α_s again,

$$\alpha_s.x_1 = \begin{bmatrix} t'_1 & 0 & 0 \\ 0 & t'_2 & 0 \\ 0 & 0 & t'_3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ u_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & u_{12} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \alpha_s.y_1$$

By Lemma 5.28,

$$\begin{aligned} \overline{\{\alpha_s.x_1,\ s \in \mathbb{R}_{<0}\}} \supset & \begin{bmatrix} t'_1 & 0 & 0 \\ 0 & t'_2 & 0 \\ 0 & 0 & t'_3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ u_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & u_{12} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathrm{A}.y_1 \\ = & \begin{bmatrix} 1 & 0 & 0 \\ u'_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & u'_{12} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathrm{A}.y_1 \end{aligned}$$

Therefore

$$\overline{\mathrm{A}^{-}.\Lambda'_{\alpha,\beta}} \supset \left\{ \begin{bmatrix} 1 & 0 & 0 \\ su'_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & s^{-1}u'_{12} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathrm{A}.y_1 \;\middle|\; s \in \mathbb{R}^{+} \right\}$$

Note that u_{12} and hence u'_{12} is non-zero. Thus, for non-zero $(l, m, n)^{\operatorname{tr}} \in \mathrm{A}.y_1$ (certainly $l \neq 0$!), by taking

$$s := \frac{mu_{12}}{l}$$

we get

$$\begin{bmatrix} 0 \\ mu_{12}u_{21} + m + mu_{21}u_{12} \\ n \end{bmatrix} \in \overline{\mathrm{A}^{-}.\Lambda'_{\alpha,\beta}}$$

Now we choose $(l,m,n) \in \mathrm{A}.y_1$ such that

$$l < 0,\ mu_{12} > 0,\ m,n \text{ very small}$$

Then invoke the A^--action on such a vector. This shows that $\mathrm{sys}(\mathrm{A}^-.\Lambda'_{\alpha,\beta})$ can not be bounded away from 0.

5.14. **Exercises.** From Exercises A to J, let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function with $f(\mathbf{0}) = 0$. You are allowed to take $n = 2$ (or $n = 3$) when working on these exercises since we did not define X_n for $n \geq 4$ in the class. Define

$$H_f := \{h \in \mathbf{SL}_n(\mathbb{R}) \mid f(h.\mathbf{v}) = f(\mathbf{v}),\ \forall\, \mathbf{v} \in \mathbb{R}^n\}.$$

Also,

$$m^*(f) := \inf\{|f(\mathbf{v})| \mid \mathbf{v} \in \mathbb{Z}^n \setminus \{\mathbf{0}\}\}$$
$$m^{**}(f) := \inf\{|f(\mathbf{v})| \mid \mathbf{v} \in \mathbb{Z}^n,\ f(\mathbf{v}) \neq 0\}.$$

Exercise A. *If $H_f.\mathbb{Z}^n$ is unbounded in X_n, then $m^*(f) = 0$.*

Exercise B. *If $f : \mathbb{R}^n \to \mathbb{R}$ is surjective and $H_f.\mathbb{Z}^n$ is dense in X_n, then $f(\mathbb{Z}^n)$ is dense in $\mathbb{R}$.*

For

$$\sigma \in S_3 := \{\text{bijections from } \{1,2,3\} \text{ to } \{1,2,3\}\},$$

let $M(\sigma) \in \mathbf{SL}_3(\mathbb{R})$ be defined by specifying the elements on the i-th row and j-th column:

$$M(\sigma)_{ij} := \begin{cases} \mathrm{sgn}(\sigma) & \text{if } \sigma(j) = i; \\ 0 & \text{otherwise} \end{cases}$$

where $\mathrm{sgn}(\sigma) \in \{1,-1\}$ denotes the usual signature of a permutation as you learned in linear algebra. Alternatively, $M(\sigma)\mathbf{e}_i = \mathrm{sgn}(\sigma)\mathbf{e}_{\sigma(i)}$ for $i = 1,2,3$ where $(\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3)$ denotes the standard basis of $\mathbb{R}^3$. Note that

$$W_3 := \{M(\sigma) \mid \sigma \in S_3\}$$

is a subgroup of $\mathbf{SL}_3(\mathbb{R})$ normalizing

$$D := \left\{ \begin{bmatrix} t_1 & 0 & 0 \\ 0 & t_2 & 0 \\ 0 & 0 & t_3 \end{bmatrix} \middle|\ t_1,t_2,t_3 \in \mathbb{R},\ t_1 \cdot t_2 \cdot t_3 = 1 \right\}.$$

Thus $W_3 \cdot D = D \cdot W_3$ is a subgroup of $\mathbf{SL}_3(\mathbb{R})$. Let $\mathrm{Nm} : \mathbb{R}^3 \to \mathbb{R}$ be defined by $\mathrm{Nm}(x,y,z) := |xyz|$.

Exercise C. *Show that $H_{\mathrm{Nm}} = W_3 \cdot D$.*

Exercise D. *Show that $H_{\mathrm{Nm}}.\mathbb{Z}^3$ is unbounded in X_3, yet $m^{**}(\mathrm{Nm}) \neq 0$.*

We say that

(S1) f has property (S1) iff for every $\varepsilon > 0$, there exists $\delta > 0$ such that for every $\mathbf{v} \in \mathbb{R}^n$ with $0 < |f(\mathbf{v})| < \delta$, there exists $h \in H_f$ such that $\|h.x\| < \varepsilon$;

(S2) f has property (S2) iff for every $\varepsilon > 0$, there exists $\delta > 0$ such that for every $\mathbf{v} \in \mathbb{R}^n$ with $0 \leq |f(\mathbf{v})| < \delta$, there exists $h \in H_f$ such that $\|h.x\| < \varepsilon$.

Exercise E. *Assume f has property* (S2)*, then $m^*(f) = 0$ implies that $H_f.\mathbb{Z}^n$ is unbounded in X_n.*

Exercise F. *Assume f has property* (S1)*, then $m^{**}(f) = 0$ implies that $H_f.\mathbb{Z}^n$ is unbounded in X_n.*

Let $\pi : \mathbb{R}^2 \to \mathbb{R}$ be defined by $\pi(x,y) := x$.

Exercise G. *Show that*

$$H_\pi = \left\{ \begin{bmatrix} 1 & 0 \\ r & 1 \end{bmatrix} \middle|\ r \in \mathbb{R} \right\}.$$

Exercise H. *The function π defined above has property* (S1) *but not* (S2)*. Also, $m^*(\pi) = 0$ but $H_\pi.\mathbb{Z}^2$ is bounded in X_2.*

Let $Q_0 : \mathbb{R}^3 \to \mathbb{R}$ defined by $Q_0(x,y,z) := x^2 + y^2 - z^2$.

Exercise I. *The function Q_0 satisfies property* (S2).

Exercise J. *Let $f(x,y) := (x-\sqrt{2}y)(x+(\sqrt{2}+1)y)$. Show that $m^*(f) \neq 0$*[12].

5.14.1. *Badly approximated vectors.* For $\lambda \in [0,1]$, define

$$\mathbf{BAD}(\lambda, 1-\lambda) := \left\{ (\alpha,\beta) \in \mathbb{R}^2 \;\middle|\; \exists c > 0, \forall q \in \mathbb{Z}^+, \max\left\{ q\langle q\alpha\rangle^{\frac{1}{\lambda}}, q\langle q\beta\rangle^{\frac{1}{1-\lambda}} \right\} > c \right\}$$

where by convention we set $\langle\cdot\rangle^{\frac{1}{0}} := 0$. Thus $\mathbf{BAD}(1,0)$ is just $\mathbf{BAD} \times [0,1)$.

Exercise K. *Show that if (α,β) is a counter-example to Littlewood conjecture, then there exists $\lambda \in [0,1]$ such that $(\alpha,\beta) \in \mathbf{BAD}(\lambda, 1-\lambda)$.*

Remark 5.31. *It seems unknown whether $\bigcap_{\lambda\in[0,1]} \mathbf{BAD}(\lambda,1-\lambda)$ is empty or not. It would have full dimension if the intersection were taken over a countable subset.*

Recall:

$$\Lambda_{\alpha,\beta} := \begin{bmatrix} 1 & 0 & \alpha \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{bmatrix} .\mathbb{Z}^3 \in \mathrm{X}_3, \quad \mathbf{u}^+_{\alpha,\beta} := \begin{bmatrix} 1 & 0 & \alpha \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{bmatrix}.$$

Exercise L. *Given $\lambda \in [0,1]$ and $(\alpha,\beta) \in \mathbb{R}^2$, show that the following two are equivalent:*

1. $(\alpha,\beta) \in \mathbf{BAD}(\lambda, 1-\lambda)$*;*
2. $$\left\{ \begin{bmatrix} e^{\lambda t} & 0 & 0 \\ 0 & e^{(1-\lambda)t} & 0 \\ 0 & 0 & e^{-t} \end{bmatrix} .\Lambda_{\alpha,\beta} \;\middle|\; t \geq 0 \right\} \text{ is bounded in } \mathrm{X}_3.$$

Given $\lambda \in [0,1]$, we say that a linear functional $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is λ-**bad** iff

$$\exists\, c > 0,\ \forall\, (l,m) \in \mathbb{Z}^2 \setminus \{(0,0)\},\ \langle \varphi(l,m) \rangle \max\{ l^{\frac{1}{\lambda}}, m^{\frac{1}{1-\lambda}} \} > c.$$

By convention : $l^{\frac{1}{\lambda}} := +\infty$ if $l \neq 0, \lambda = 0$ and $l^{\frac{1}{\lambda}} := 0$ if $l = 0, \lambda = 0$. The convention for $m^{\frac{1}{1-\lambda}}$ is similar. Define

$$\Lambda'_{\alpha,\beta} := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \alpha & \beta & 1 \end{bmatrix} .\mathbb{Z}^3 \in \mathrm{X}_3$$

Exercise M. *Given $\lambda \in [0,1]$ and $(\alpha,\beta) \in \mathbb{R}^2$, show that the following two are equivalent:*

1. *The linear functional $\varphi_{\alpha,\beta}(x,y) := \alpha x + \beta y$ is λ-bad;*
2. $$\left\{ \begin{bmatrix} e^{-\lambda t} & 0 & 0 \\ 0 & e^{-(1-\lambda)t} & 0 \\ 0 & 0 & e^{t} \end{bmatrix} .\Lambda'_{\alpha,\beta} \;\middle|\; t \geq 0 \right\} \text{ is bounded in } \mathrm{X}_3.$$

Exercise N. *Given $\lambda \in [0,1]$ and $(\alpha,\beta) \in \mathbb{R}^2$, show that the following two are equivalent:*

1. $(\alpha,\beta) \in \mathbf{BAD}(\lambda, 1-\lambda)$*;*
2. *The linear functional $\varphi_{\alpha,\beta}(x,y) := \alpha x + \beta y$ is λ-bad.*

Hint: show that the map $g\mathbb{Z}^3 \mapsto g^{-\operatorname{tr}}\mathbb{Z}^3$ on X_3 is well-defined and continuous.

Exercise O. *Show that if $(\alpha,\beta) \in \mathbb{R}^2$ are linearly dependent over $\mathbb{Q}$, then $(\alpha,\beta) \notin \mathbf{BAD}(\lambda, 1-\lambda)$ for every $\lambda \in (0,1)$.*

Exercise P. *Assume that α,β are two numbers contained in the same totally real number field. Moreover assume α,β are linearly independent over $\mathbb{Q}$. Then (α,β) is contained in $\mathbf{BAD}(1/2,1/2)$.*

Hint: In the proof, you may assume the following facts, which can be proved quickly if you know a little Galois theory. There exist real numbers t_i's, u_{ij}'s and $M_0 \in \mathbf{SL}_3(\mathbb{R})$ such that $\prod_{i=1}^3 t_i = 1$, the A-orbit $\mathrm{A}.(M_0.\mathbb{Z}^3)$ is compact and

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\alpha & -\beta & 1 \end{bmatrix} = \begin{bmatrix} t_1 & 0 & 0 \\ 0 & t_2 & 0 \\ 0 & 0 & t_3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ u_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{bmatrix} \cdot M_0$$

[12] This shows that Oppenheim conjecture fails in two variables.

Indeed, if $\{\sigma_1 = \mathrm{id}, \sigma_2, \sigma_3\}$ denote the three distinct embeddings of the totally real field into $\mathbb{R}$, then M_0 can be chosen to be a scalar multiple of

$$\begin{bmatrix} -\sigma_3(\alpha) & -\sigma_3(\beta) & 1 \\ -\sigma_2(\alpha) & -\sigma_2(\beta) & 1 \\ -\alpha & -\beta & 1 \end{bmatrix}.$$

5.14.2. *Furstenberg's times 2 times three theorem.*

Exercise Q. *Find an irrational number α and $\varepsilon_0 > 0$ such that $\langle 2^n\alpha\rangle > \varepsilon_0$ for all $n \in \mathbb{Z}_{\geq 0}$.*

On the other hand, Furstenberg's theorem asserts that if α is an irrational number, then $\{2^n 3^m \alpha,\ n, m \in \mathbb{Z}_{\geq 0}\}$ is dense modulo $\mathbb{Z}$. The following few exercises is to walk you through a proof of this fact[13].

Order the set $\{2^n 3^m,\ n, m \in \mathbb{Z}_{\geq 0}\}$ as

$$\{2^n 3^m,\ n, m \in \mathbb{Z}_{\geq 0}\} = \{a_1 < a_2 < a_3 < ...\}$$

Exercise R. *Let α be an irrational number. Show that for every $\varepsilon > 0$, there exists $n, m \in \mathbb{Z}^+$ such that*

$$0 < n\alpha - m < \varepsilon.$$

Exercise S. *Show that $\lim_n \frac{a_{n+1}}{a_n} = 1$.*

Hint: apply the proceeding exercise to $\log(2)/\log(3)$ and $\log(3)/\log(2)$.

For an integer p, let $T_p : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be defined by $x + \mathbb{Z} \mapsto px + \mathbb{Z}$. For an irrational number α, let C_α be the closure of the set

$$\{T_2^n T_3^m(\alpha + \mathbb{Z}) \mid n, m \in \mathbb{Z}_{\geq 0}\} = \{2^n 3^m \alpha + \mathbb{Z} \mid n, m \in \mathbb{Z}_{\geq 0}\} \subset \mathbb{R}/\mathbb{Z}.$$

Exercise T. *Take α, an irrational number. Show that there exists $(\varepsilon_n) \subset C_\alpha - C_\alpha$, $\varepsilon_n \neq 0 + \mathbb{Z}$ such that $\lim_n \varepsilon_n = 0 + \mathbb{Z}$.*

Exercise U. *Take α, an irrational number. Show that $C_\alpha - C_\alpha = \mathbb{R}/\mathbb{Z}$. Namely, for every $x + \mathbb{Z} \in \mathbb{R}/\mathbb{Z}$, there exists $a + \mathbb{Z}, b + \mathbb{Z} \in C_\alpha$ such that $a - b \equiv x \pmod{\mathbb{Z}}$.*

Hint: Use the Exercise T and S.

For $l \in \mathbb{Z}^+$ that is coprime to $2, 3$, define

$$\Sigma_l := \{2^n 3^m \mid n, m \in \mathbb{Z}_{\geq 0},\ 2^n 3^m \equiv 1 \pmod{l}\}.$$

And order Σ_l as

$$\Sigma_l = \{a_1^{(l)} < a_2^{(l)} < a_3^{(l)} < ...\}$$

Exercise V. *For $l \in \mathbb{Z}^+$ that is coprime to $2, 3$, $\lim_n \frac{a_{n+1}^{(l)}}{a_n^{(l)}} = 1$.*

Exercise W. *Take an irrational number α. If C_α contains a rational number modulo $\mathbb{Z}$, then $C_\alpha = \mathbb{R}/\mathbb{Z}$.*

Hint: similar to the proof of Exercise U. Maybe you also need Exercise V in the place of Exercise S.

Exercise X. *Let α be an irrational number. Show that C_α must contain a rational point modulo $\mathbb{Z}$.*

Hint: If C_α does not contain any rational number, then for every $l \in \mathbb{Z}^+$, show that[14]

$$C_\alpha \cap \left(C_\alpha - \frac{1}{5^l}\right) \cap \left(C_\alpha - \frac{2}{5^l}\right) \cap .... \cap \left(C_\alpha - \frac{5^l - 1}{5^l}\right) \neq \emptyset. \tag{16}$$

How to show this? Certainly, $C_\alpha \cap \left(C_\alpha - \frac{1}{5^l}\right)$ is non-empty by Exercise U. Call $D_1 := C_\alpha \cap (C_\alpha - \frac{1}{5^l})$. Note that

$$C_\alpha \cap (C_\alpha - \frac{1}{5^l}) \cap (C_\alpha - \frac{2}{5^l}) = D_1 \cap \left(D_1 - \frac{1}{5^l}\right).$$

On the other hand D_1 is invariant under $\{T_n,\ n \in \Sigma_{5^l}\}$. If it does not contain any rational point, then you can show that, just as Exercise U, $D_1 - D_1 = \mathbb{R}/\mathbb{Z}$. In particular $D_1 \cap \left(D_1 - \frac{1}{5^l}\right)$ is non-empty. Equa.(16) can be verified by repeating this process.

[13] if you manage to give a proof without doing any of the following exercises, you would still earn the full credit.

[14] Here "$-\frac{1}{5^l}$" should be understood as modulo $\mathbb{Z}$.

6. Lecture 5, A Naive Explanation of the Low and High Entropy Method

6.1. **Prelude.** In this lecture, we present the key idea of the EKL paper: the high and low entropy method. We are going to make some (too strong) assumptions under which the idea of these methods shall be explained.

The key are unipotent matrices and their interplay with diagonal matrices. Unipotent matrices could be sources of being unbounded. For instance, if Γ is a discrete subgroup of $G = \mathbf{SL}_2(\mathbb{R})$ (or any other semisimple linear Lie group) that contains some non-trivial unipotent matrix, then G/Γ is non-compact.

6.2. **Notation.**

$$\mathrm{A} = \left\{ \begin{bmatrix} e^{t_1} & 0 & 0 \\ 0 & e^{t_2} & 0 \\ 0 & 0 & e^{t_3} \end{bmatrix} \middle| \sum t_i = 0 \right\}$$

$$\mathrm{A}^+ = \left\{ \begin{bmatrix} e^{t_1} & 0 & 0 \\ 0 & e^{t_2} & 0 \\ 0 & 0 & e^{t_3} \end{bmatrix} \in \mathrm{A} \middle| t_1, t_2 > 0 \right\}.$$

For $i \neq j$, let E_{ij} be the matrix whose (i,j)-entry is one and is zero elsewhere. Let $\mathbf{u}_{ij}(r) := \mathrm{I}_3 + rE_{ij}$ and $\mathrm{U}_{ij} := \{\mathbf{u}_{ij}(r),\ r \in \mathbb{R}\}$. For instance,

$$\mathrm{U}_{12} := \left\{ \mathbf{u}_{12}(s) = \begin{bmatrix} 1 & s & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \middle| s \in \mathbb{R} \right\}, \quad \mathrm{U}_{13} := \left\{ \mathbf{u}_{13}(s) = \begin{bmatrix} 1 & 0 & s \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \middle| s \in \mathbb{R} \right\},$$

$$\mathrm{U}_{23} := \left\{ \mathbf{u}_{23}(s) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & s \\ 0 & 0 & 1 \end{bmatrix} \middle| s \in \mathbb{R} \right\}.$$

Also for (i,j,k), an ordering of $\{1,2,3\}$, let $\mathrm{U}_{ijk} := \mathrm{U}_{ij}\mathrm{U}_{ik}\mathrm{U}_{jk}$, $\mathrm{U}_{ij,ik} := \mathrm{U}_{ij}\mathrm{U}_{ik}$ and $\mathrm{U}_{ik,jk} := \mathrm{U}_{ik}\mathrm{U}_{jk}$. These are subgroups. For instance:

$$\mathrm{U}_{123} := \left\{ \begin{bmatrix} 1 & r & s \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix} \middle| r, s, t \in \mathbb{R} \right\}, \quad \mathrm{U}_{12,13} := \left\{ \begin{bmatrix} 1 & r & s \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \middle| r, s \in \mathbb{R} \right\}.$$

It is also useful to note that the centralizer of U_{13} is

$$\mathrm{Z}_{\mathbf{SL}_3}(\mathrm{U}_{13}) = \left\{ \begin{bmatrix} t & u_{12} & u_{13} \\ 0 & t^{-2} & u_{23} \\ 0 & 0 & t \end{bmatrix} \right\}.$$

6.3. **Recurrence leaf.** [15]

Recall that for $(\alpha, \beta) \in \mathbb{R}^2$, we let

$$\Lambda_{\alpha,\beta} := \begin{bmatrix} 1 & 0 & \alpha \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{bmatrix} .\mathbb{Z}^3.$$

For $E \subset [0,1)^2$,

$$\mathscr{C}_E := \left\{ x \mid x = \lim a_n.\Lambda_{\alpha,\beta},\ \exists\, (\alpha,\beta) \in E \text{ and divergent } (a_n) \subset \mathrm{A}^+ \right\}.$$

For $x \in \mathrm{X}_3$ and an ordering (ijk) of $\{1,2,3\}$, let

$$\mathscr{C}_x^{ij} := \{u \in \mathrm{U}_{ij} \mid u.x \in \mathscr{C}_E\}$$
$$\mathscr{C}_x^{ij,ik} := \{u \in \mathrm{U}_{ij,ik} \mid u.x \in \mathscr{C}_E\},\ \mathscr{C}_x^{ik,jk} := \{u \in \mathrm{U}_{ik,jk} \mid u.x \in \mathscr{C}_E\}$$
$$\mathscr{C}_x^{ijk} := \{u \in \mathrm{U}_{ijk} \mid u.x \in \mathscr{C}_E\}$$

Lemma 6.1. *These sets satisfy certain formal properties such as*

1. *for* $u \in \mathrm{U}_{ij}$, $\mathscr{C}_{u.x}^{ij} \cdot u = \mathscr{C}_x^{ij}$;
2. *for* $a = \mathrm{diag}(a_1, a_2, a_3) \in \mathrm{A}$, $\mathscr{C}_{a.x}^{ij} = a\mathscr{C}_x^{ij}a^{-1}$.

Now fix some $E \subset [0,1)^2$ such that $\mathrm{A}^+.\Lambda_E$ is contained in some compact subset of X_3. From now on, we make the following (too strong) assumptions[16]:

[15] Maybe the correct name should be recurrence set on leaves?

[16] The key assumption is the first one on continuity.

Assumption 6.2.
- *The map $x \mapsto \mathscr{C}_x^\star$ is continuous from $\mathscr{C}_E$ to the set of closed subsets*[17] *of some $\mathrm{U}_\star$ for any $\star = (ij)$, (ij, ik), (ik, jk) or (ijk);*
- *For every ordering (ijk) of $\{1,2,3\}$, $\mathscr{C}_x^{ijk}$ being infinite for every $x \in \mathscr{C}_E$ is equivalent to $\mathscr{C}_x^{kji}$ being infinite for every $x \in \mathscr{C}_E$.*
- *for every $i \neq j$, we have the following dichotomy: either $\mathscr{C}_x^{ij}$ is a singleton $\{\mathrm{I}_3\}$ for every $x \in \mathscr{C}_E$ or $\mathscr{C}_x^{ij}$ is infinite for every $\mathscr{C}_E$;*
- *there exists $i \neq j$ such that $\mathscr{C}_x^{ij}$ is infinite*[18] *for every $x \in \mathscr{C}_E$.*

Corollary 6.3. *Under the above assumptions, if $\mathscr{C}_x^{ij}$ is infinite for every $x \in \mathscr{C}_E$, then $\mathscr{C}_x^{ij}$ contains arbitrarily small non-identity elements for every $x \in \mathscr{C}_E$;*

Proof. If for some $x \in \mathscr{C}_E$, one can find $\rho > 0$ with $\mathbf{u}_{ij}((-\rho,\rho)) \cap \mathscr{C}_x^{ij} = \{\mathrm{I}_3\}$, then

$$\mathbf{u}_{ij}((-a_i a_j^{-1}\rho, a_i a_j^{-1}\rho)) \cap \mathscr{C}_x^{ij} = \{\mathrm{I}_3\}, \quad \forall\, a = \mathrm{diag}(a_1,a_2,a_3) \in \mathrm{A}.$$

Choose $a(n) = \mathrm{diag}(a(n)_1, a(n)_2, a(n)_3)$ such that $a(n)_i/a(n)_j \to +\infty$. And let y be any limit point of $a(n).x$. Then by continuity, $\mathscr{C}_y^{ij} = \{0\}$. This is a contradiction. $\square$

From now on assume E is non-empty and $\mathscr{C}_E$ is compact[19]. We would like to derive a contradiction. Let us actually make a statement in case it seems too vague to you.

Theorem 6.4. *Let $\mathscr{C}_E$ (the subscript E means nothing here) be an A-invariant compact subset of X_3 satisfying Assumption 6.2. Then $\mathscr{C}_E$ is empty.*

Anticipating the proof, we shall exhibit a $\mathrm{U}_{ij}^+ := \mathbf{u}_{ij}(\mathbb{R}_{\geq 0})$ or $\mathrm{U}_{ij}^- := \mathbf{u}_{ij}(\mathbb{R}_{\leq 0})$ orbit inside $\mathscr{C}_E$ for some $i \neq j$. But this would contradict against the following:

Lemma 6.5. *For each $i \neq j$ and $\star = +, -$, every orbit of the semigroup $\mathrm{A} \cdot \mathrm{U}_{ij}^\star$ on X_3 is unbounded.*

Proof. Without loss of generality assume $(i,j) = (2,3)$ and $\star = +$.

Take $\Lambda \in \mathrm{X}_3$ and $\mathbf{v} = (v_1, v_2, v_3) \in \Lambda$ with $v_3 < 0$ (every lattice would contain such a vector). By choosing suitable r, the lattice

$$u.\Lambda := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & r \\ 0 & 0 & 1 \end{bmatrix}.\Lambda$$

contains some vector $\mathbf{w} = (w_1, 0, w_2)$. Then one can find $a_n \in \mathrm{A}$ such that $a_n.\mathbf{w} \to \mathbf{0}$. By continuity of the systole function, $\{a_n u.\Lambda\}$ is unbounded in X_3. $\square$

6.4. **Product structure.** Roughly speaking, the lemma below says that "Recurrence leaf in the central direction is unchanged along unstable leaves".

Lemma 6.6. *Take $x \in \mathscr{C}_E$ and $u \in \mathrm{U}_{12}, v \in \mathrm{U}_{13,23}$ such that $y := uv.x \in \mathscr{C}_E$, then*

$$u \cdot \mathscr{C}_y^{12} = \mathscr{C}_x^{12}.$$

Proof. Take a sequence $(a_n) \subset \mathrm{A}$ such that conjugating by a_n contracts $\mathrm{U}_{13,23}$ and a_n commutes with U_{12} (e.g., take $a_n := \mathrm{diag}(n^{-1}, n^{-1}, n^2)$). Passing to a subsequence, assume that $a_n.x$ converges to $x_\infty \in \mathscr{C}_E$. By continuity of $\mathscr{C}_\bullet^{12}$,

$$\mathscr{C}_x^{12} = \mathscr{C}_{a_n.x}^{12} \to \mathscr{C}_{x_\infty}^{12} = u \cdot \mathscr{C}_{u.x_\infty}^{12} \leftarrow u \cdot \mathscr{C}_{a_n uv.x}^{12} = u \cdot \mathscr{C}_y^{12}.$$

$\square$

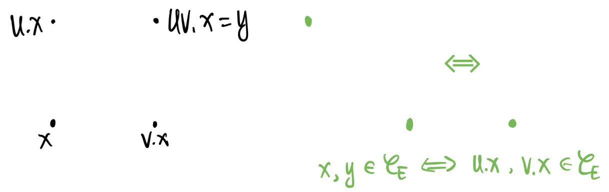
Corollary 6.7. *The product map $(g,h) \mapsto g \cdot h$ induces a bijection $\mathscr{C}_x^{12} \times \mathscr{C}_x^{13,23} \cong \mathscr{C}_x^{123}$ for every $x \in \mathscr{C}_E$.*

[17] A sequence of closed subsets (E_n) of $\mathbb{R}^n$ converges to E iff for every bounded open subset $O \subset \mathbb{R}^n$ the Hausdorff distance between $E_n \cap O$ and $E \cap O$ decreases to zero.

[18] One expects that this is likely to hold if $\dim E > 0$

[19] Recall that for (α,β) that fails Littlewood, we have that $\mathscr{C}_{(\alpha,\beta)}$ is compact

By similar arguments, $\mathscr{C}_x^{23} \times \mathscr{C}_x^{12,13} \cong \mathscr{C}_x^{123}$, $\mathscr{C}_x^{12} \times \mathscr{C}_x^{13} \cong \mathscr{C}_x^{12,13}$.... Soon we will see that these different decomposition of $\mathscr{C}_x^{123}$ lead to additional invariance.



Proof. Let $u \in \mathrm{U}_{12}, v \in \mathrm{U}_{13,23}$ and $x \in \mathscr{C}_E$. We need to show that

$$uv.x \in \mathscr{C}_E \iff u.x,\, v.x \in \mathscr{C}_E.$$

First we do the " $\Longrightarrow$ " direction. Indeed, by Lemma 6.6,

$$uv.x \in \mathscr{C}_E \implies \mathrm{id} \in \mathscr{C}_{uv.x}^{12} = u^{-1} \cdot \mathscr{C}_x^{12} \implies u \in \mathscr{C}_x^{12} \implies u.x \in \mathscr{C}_E.$$

Similarly,

$$x \in \mathscr{C}_E \implies \mathrm{id} \in \mathscr{C}_x^{12} = u \cdot \mathscr{C}_{uv.x}^{12} \implies u^{-1} \in \mathscr{C}_{uv.x}^{12} \implies v.x \in \mathscr{C}_E.$$

For the reverse implication "$\Longleftarrow$", by Lemma 6.6,

$$x \in \mathscr{C}_E, v.x \in \mathscr{C}_E \implies \mathscr{C}_x^{12} = \mathscr{C}_{v.x}^{12}.$$

Therefore,

$$u.x \in \mathscr{C}_E \implies u \in \mathscr{C}_x^{12} = \mathscr{C}_{v.x}^{12} \implies uv.x \in \mathscr{C}_E.$$

□

6.5. **Product structure vs. non-commutativity of the Heisenberg group.** Let us calculate the commutator $[u, v]$ for $u \in \mathrm{U}_{12}$ and $v \in \mathrm{U}_{23}$:

$$\begin{bmatrix} 1 & s & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -s & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -t \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & st \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{17}$$

Namely,

$$\mathbf{u}_{12}(s)\mathbf{u}_{23}(t)\mathbf{u}_{12}(s)^{-1}\mathbf{u}_{23}(t)^{-1} = \mathbf{u}_{13}(st).$$

Lemma 6.8. *Take $x \in \mathscr{C}_E$. Assume that both $\mathscr{C}_x^{12}$ and $\mathscr{C}_x^{23}$ contain non-identity elements arbitrarily close to the identity, then $\mathscr{C}_x^{13}$ contains $\mathbf{u}_{13}(\mathbb{R}_{\geq 0})$ or $\mathbf{u}_{13}(\mathbb{R}_{\leq 0})$.*

Proof. Take non-zero $s_n \to 0$ and $t_n \to 0$ such that $\mathbf{u}_{12}(s_n) \in \mathscr{C}_x^{12}$ and $\mathbf{u}_{23}(t_n) \in \mathscr{C}_x^{23}$. Without loss of generality assume $s_n, t_n > 0$.

We are going to show that $\mathscr{C}_x^{13}$, when identified with a subset of $\mathbb{R}$, is invariant under addition by $s_n t_n$. Namely, taking $z \in \mathrm{U}_{13}$ with $z.x \in \mathscr{C}_E$, we need to show that $\mathbf{u}_{13}(s_n t_n) z.x \in \mathscr{C}_E$. Once this is done, a continuity argument shows that $\mathbb{R}_{\geq 0} + \mathscr{C}_x^{13} \subset \mathscr{C}_x^{13}$. In particular, $\mathscr{C}_x^{13}$ contains $\mathbb{R}_{\geq 0}$.

Note that $x \in \mathscr{C}_E$. By Corollary 6.7 (applied to $\mathrm{U}_{23} \times \mathrm{U}_{13}$),

$$\mathbf{u}_{23}(t_n).x,\, z.x \in \mathscr{C}_E \implies \mathbf{u}_{23}(t_n) z.x \in \mathscr{C}_E.$$

By Corollary 6.7 again (applied to $\mathrm{U}_{12} \times \mathrm{U}_{13,23}$),

$$\mathbf{u}_{12}(s_n).x,\, \mathbf{u}_{23}(t_n) z.x \in \mathscr{C}_E \implies \mathbf{u}_{12}(s_n)\mathbf{u}_{23}(t_n) z.x \in \mathscr{C}_E.$$

By Equa.(17), this is equivalent to

$$\mathbf{u}_{13}(s_n t_n).(\mathbf{u}_{23}(t_n)\mathbf{u}_{12}(s_n) z.x) = (\mathbf{u}_{23}(t_n)) \cdot (\mathbf{u}_{13}(s_n t_n)\mathbf{u}_{12}(s_n)).(z.x) \in \mathscr{C}_E.$$

By Corollary 6.7 and $z.x \in \mathscr{C}_E$,

$$\begin{aligned} & (\mathbf{u}_{23}(t_n)) \cdot (\mathbf{u}_{13}(s_n t_n)\mathbf{u}_{12}(s_n)).(z.x) \in \mathscr{C}_E \\ (\mathrm{U}_{23} \times \mathrm{U}_{12,13}) & \implies \mathbf{u}_{13}(s_n t_n)\mathbf{u}_{23}(t_n).(z.x) \in \mathscr{C}_E \\ (\mathrm{U}_{13} \times \mathrm{U}_{12}) & \implies \mathbf{u}_{13}(s_n t_n).(z.x) \in \mathscr{C}_E. \end{aligned}$$

So we are done. □

6.6. **Conclusion of the high entropy method.**

Lemma 6.9. *For every $x \in \mathscr{C}_E$, at most one of $\mathscr{C}_x^{12}, \mathscr{C}_x^{13}$ and $\mathscr{C}_x^{23}$ is infinite.*

Similarly at most one of $\mathscr{C}_x^{21}, \mathscr{C}_x^{23}$ and $\mathscr{C}_x^{13}$ is infinite.
There are essentially two cases to consider.

6.6.1. *Case I.* Assume that $\mathscr{C}_x^{12}$ and $\mathscr{C}_x^{23}$ are infinite. By Corollary 1.3 (2), $\mathscr{C}_x^{12}$ and $\mathscr{C}_x^{23}$ contains non-identity elements arbitrarily close to id. By Lemma 6.8, we may assume $\mathbf{u}_{13}(\mathbb{R}_{\geq 0})$ (the other case is similar) belongs to $\mathscr{C}_x^{13}$. So we have

$$\left\{ \begin{bmatrix} e^{t_1} & 0 & r_2 \\ 0 & e^{t_2} & 0 \\ 0 & 0 & e^{t_3} \end{bmatrix} \middle| \sum t_i = 0, r_2 \geq 0 \right\} .x \subset \mathscr{C}_E \text{ is bounded,} \tag{18}$$

which is impossible by Lemma 6.5.

6.6.2. *Case II.* Assume that $\mathscr{C}_x^{12}$ and $\mathscr{C}_x^{13}$ are infinite but $\mathscr{C}_x^{23}$ is finite for every x. By part (2) of the Assumption 6.2, $\mathscr{C}_x^{321}$ is infinite and hence by product structure at least one of $\mathscr{C}_x^{21}, \mathscr{C}_x^{31}, \mathscr{C}_x^{32}$ is infinite. Then similar arguments as in case I would lead to a contradiction against Lemma 6.5.

6.6.3. *One can avoid the use of the assumption here...* **We did not do this in the class. One can skip ahead to the Lemma below.**

We claim that at least one of $\mathscr{C}_x^{21}, \mathscr{C}_x^{31}, \mathscr{C}_x^{32}$ is infinite holds without invoking the part (2) of the assumption.

Now assume they are all finite. For $\eta > 0$, let[20]

$$H_\eta := \left\{ \begin{bmatrix} e^{t_1} & r_1 & r_2 \\ 0 & e^{t_2} & r_3 \\ 0 & 0 & e^{-t_1-t_2} \end{bmatrix} \middle| \, |t_1|, |t_2|, |r_1|, |r_2|, |r_3| < \eta \right\}$$

and

$$\theta_t := \begin{bmatrix} e^{-t} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^t \end{bmatrix}$$

Then by Assumption 6.2, there exists $\delta_1, \eta_1 > 0$ small enough such that for every $x, x' \in \mathscr{C}_E$ with $d(x, x') < \delta_1 \implies x' \in H(\eta_1).x$.

Indeed, if this were not the case, by the "exponential blow-up" (see Lecture 4), we can construct $z \neq z' \in \mathscr{C}_E$ such that

$$z' = \begin{bmatrix} 1 & 0 & 0 \\ r_1 & 1 & 0 \\ r_2 & r_3 & 1 \end{bmatrix} .z$$

with r_1, r_2, r_3 arbitrarily close to 0. By Corollary 6.7 together with our assumption that these leaves are finite, this would contradict against Assumption 6.2.

Now, we can cover $\mathscr{C}_E$ by finitely many $\{H(\eta_1).x_i, \ i = 1, .., l\}$. Choose $t_n \to +\infty$ such that $z_i := \lim \theta_{t_n}.x_i$ exists for every i. Then[21]

$$\mathscr{C}_E = \theta_{t_n}.\mathscr{C}_E \subset \bigcup \theta_{t_n}.H(\eta_1).x_i \to \bigcup \mathrm{A}.z_i \subset \mathscr{C}_E.$$

Therefore, $\mathscr{C}_E$ is a finite union of A-orbits. So each of them is compact. This contradicts against our assumption[22].

Corollary 6.10. *If $\mathscr{C}_x^{23}$ is infinite for every $x \in \mathscr{C}_E$, then $\mathscr{C}_x^{21}$, $\mathscr{C}_x^{31}$, $\mathscr{C}_x^{12}$ and $\mathscr{C}_x^{13}$ are finite for every $x \in \mathscr{C}_E$.*

[20] one can also impose $r_3 = 0$

[21] make sense the these implications!

[22] Imagine two compact A-orbit are linked by a unipotent, then suitable $a_n \in \mathrm{A}$ would bring these two tori closer and closer, which is impossible

6.7. **A "doubling" property.** Henceforth, we assume that $\mathscr{C}_x^{23}$ is infinite and $\mathscr{C}_x^{12}$, $\mathscr{C}_x^{13}$ are finite (for every $x \in \mathscr{C}_E$). The proof for the remaining cases is similar.

Lemma 6.11. *There exists $\rho_0 \in (0,1)$ such that for every $x \in \mathscr{C}_E$, there exists $\rho_x \in I_0 := (-1,-\rho_0) \cup (\rho_0, 1)$ such that $\mathbf{u}_{23}(\rho_x) \in \mathscr{C}_x^{23}$.*

Proof. If not, using the continuity of $x \mapsto \mathscr{C}_x^{12}$, one can show that $\mathscr{C}_x^{12} = \{\mathrm{I}_3\}$ for some $x \in \mathscr{C}_E$. A contradiction. $\square$

Fix such a ρ_0 and I_0. Using the A-action, one gets

Corollary 6.12. *For every $x \in \mathscr{C}_E$ and every $\lambda > 0$, there exists $\rho_x(\lambda) \in \lambda I_0$ such that $\mathbf{u}_{23}(\rho_x(\lambda)) \in \mathscr{C}_x^{23}$.*

Without loss of generality, assume that $\mathscr{C}_x^{23}$ is infinity for every $x \in \mathscr{C}_E$ and $I_0 = (\rho_0, 1)$.

6.8. **Unipotent blowup/Low entropy method.** The following calculation is the key to the low entropy method. Its use in dynamics can be traced back to the work of Ratner on joinings of unipotent flows.

$$\begin{aligned}
&\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & r \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -r \\ 0 & 0 & 1 \end{bmatrix} \\
= &\begin{bmatrix} g_{11} & g_{12} & g_{13} - rg_{12} \\ g_{21} + rg_{31} & g_{22} + rg_{32} & g_{23} - r^2 g_{32} + r(g_{33} - g_{22}) \\ g_{31} & g_{32} & g_{33} - rg_{32} \end{bmatrix}
\end{aligned}$$

For simplicity, let $r_0 := \mathrm{InjRad}(\mathscr{C}_E) > 0$.

For a pair of points $x, x' \in \mathscr{C}_E$ with $d(x,x') < r_0$, there exists a unique $g = g(x,x') \in B(r_0)$ such that $x' = g.x$. We let

$$\varepsilon(x,x') := \|\mathrm{I}_3 - g(x,x')\|_{\sup}.$$

For $\delta > 0$, we let

$$r_\delta(x,x') := \min\left\{ \frac{\delta}{|g_{12}|}, \frac{\delta}{|g_{31}|}, \frac{\sqrt{\delta}}{\sqrt{|g_{32}|}}, \frac{\delta}{|g_{33} - g_{22}|} \right\}.$$

If some denominator is zero, we think of the corresponding term as being $+\infty$. So $r_\delta(x,x') \in (0, +\infty]$. From the above matrix calculation, we see that

Lemma 6.13. *For $x, x' \in \mathscr{C}_E$ with $d(x,x') < r_0$, If $g(x,x') \notin \mathrm{Z}_{\mathbf{SL}_3}(\mathrm{U}_{23})$, then $r_\delta(x,x') < +\infty$.*

Assume $r_\delta(x,x') < +\infty$, we let $n = n_\delta(x,x')$ be the unique integer such that

$$r_\delta(x,x') \in [\rho_0^{-n}, \rho_0^{-(n+1)}).$$

$n_\delta(x,x')$ is large if $\varepsilon(x,x')$ is much smaller compared to δ.

By Corollary 6.12, we can find

$$\begin{aligned}
r' &= r'_\delta(x,x') \in [\rho_0^{-n}, \rho_0^{-(n+1)}) \text{ such that } \mathbf{u}_{23}(r') \in \mathscr{C}_x^{23}, \\
r'' &= r''_\delta(x,x') \in [\rho_0^{-(n+2)}, \rho_0^{-(n+3)}) \text{ such that } \mathbf{u}_{23}(r'') \in \mathscr{C}_x^{23}.
\end{aligned}$$

Also let

$$\lambda_\delta(x,x') := \frac{r''_\delta(x,x')}{r'_\delta(x,x')} \in (\rho_0^{-1}, \rho_0^{-3}).$$

For $s \in \mathbb{R}$, let

$$\begin{aligned}
g(s) := \mathbf{u}_{23}(s) g \mathbf{u}_{23}(s)^{-1} &= \begin{bmatrix} g(s)_{11} & g(s)_{12} & g(s)_{13} \\ g(s)_{21} & g(s)_{22} & g(s)_{23} \\ g(s)_{31} & g(s)_{32} & g(s)_{33} \end{bmatrix} \\
&= \begin{bmatrix} g_{11} & g_{12} & g_{13} - sg_{12} \\ g_{21} + sg_{31} & g_{22} + sg_{32} & g_{23} - s^2 g_{32} + s(g_{33} - g_{22}) \\ g_{31} & g_{32} & g_{33} - sg_{32} \end{bmatrix}
\end{aligned}$$

Lemma 6.14. *Fix $\delta \in (0,1)$. Take $x, x' \in \mathscr{C}_E$ with $d(x,x') < r_0$ and $r_\delta(x,x') < +\infty$. Assume further that $\varepsilon(x,x') < \frac{\rho_0(1-\rho_0)}{4}\delta < \frac{1}{4}\rho_0\delta$. Then there is $s := s_\delta(x,x') \in \{r'_\delta(x,x'), r''_\delta(x,x')\}$ such that*

$$3\delta > \max\{|g(s)_{21}|, |g(s)_{13}|, |g(s)_{23}|\} \geq \rho_1\delta$$

where $\rho_1 := \frac{\rho_0(1-\rho_0)}{4}$.

Proof. The "$3\delta >$" part is easy. Let us focus on the other inequality.

If $r_\delta(x,x') = \delta |g_{12}|^{-1}$, then take $s := r'_\delta(x,x')$. We have

$$|g(s)_{13}| = |g_{13} - sg_{12}| \geq \rho_0\delta - \varepsilon(x,x') \geq \rho_1\delta.$$

Similarly, if $r = r_\delta(x,x') = \delta |g_{31}|^{-1}$, then

$$|g(s)_{21}| = |g_{21} + sg_{31}| \geq \rho_0\delta - \varepsilon(x,x') \geq \rho_1\delta.$$

where $s := r'_\delta(x,x')$.

Now assume that $r = r_\delta(x,x') = \min\left\{\delta |g_{32}|^{-\frac{1}{2}}, \delta |g_{33} - g_{22}|^{-1}\right\}$, then

$$\max\left\{(r')^2 |g_{32}|, r' |g_{33} - g_{22}|\right\} \geq \rho_0\delta.$$

where $r' := r'_\delta(x,x')$ Write $\lambda := \lambda_\delta(x,x')$ and note that

$$\begin{aligned}
&\begin{bmatrix} 1 & 1 \\ \lambda^2 & \lambda \end{bmatrix} \begin{bmatrix} -r^2 g_{32} \\ r(g_{33} - g_{22}) \end{bmatrix} = \begin{bmatrix} -r^2 g_{32} + r(g_{33} - g_{22}) \\ -(\lambda r)^2 g_{32} + (\lambda r)(g_{33} - g_{22}) \end{bmatrix} \\
\implies &\begin{bmatrix} -r^2 g_{32} \\ r(g_{33} - g_{22}) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \lambda^2 & \lambda \end{bmatrix}^{-1} \begin{bmatrix} -r^2 g_{32} + r(g_{33} - g_{22}) \\ -(\lambda r)^2 g_{32} + (\lambda r)(g_{33} - g_{22}) \end{bmatrix} \\
\implies &\left\| \begin{bmatrix} -r^2 g_{32} \\ r(g_{33} - g_{22}) \end{bmatrix} \right\|_{\sup} \leq 2 \left\| \begin{bmatrix} 1 & 1 \\ \lambda^2 & \lambda \end{bmatrix}^{-1} \right\|_{\sup} \left\| \begin{bmatrix} -r^2 g_{32} + r(g_{33} - g_{22}) \\ -(\lambda r)^2 g_{32} + (\lambda r)(g_{33} - g_{22}) \end{bmatrix} \right\|_{\sup}
\end{aligned}$$

But

$$\begin{aligned}
&\begin{bmatrix} 1 & 1 \\ \lambda^2 & \lambda \end{bmatrix}^{-1} = \frac{1}{\lambda - \lambda^2} \begin{bmatrix} \lambda & -1 \\ -\lambda^2 & 1 \end{bmatrix}^{-1} \\
\implies &\left\| \begin{bmatrix} 1 & 1 \\ \lambda^2 & \lambda \end{bmatrix}^{-1} \right\|_{\sup} = \frac{\lambda^2}{\lambda^2 - \lambda} \leq \frac{1}{1 - \rho_0}.
\end{aligned}$$

So we have

$$\left\| \begin{bmatrix} -r^2 g_{32} + r(g_{33} - g_{22}) \\ -(\lambda r)^2 g_{32} + (\lambda r)(g_{33} - g_{22}) \end{bmatrix} \right\|_{\sup} \geq \frac{\rho_0(1 - \rho_0)}{2}\delta.$$

Therefore,

$$\max\{|g(r)_{23}|, |g(\lambda r)_{23}|\} \geq \frac{\rho_0(1-\rho_0)}{2}\delta - \varepsilon(x,x') \geq \rho_1\delta.$$

□

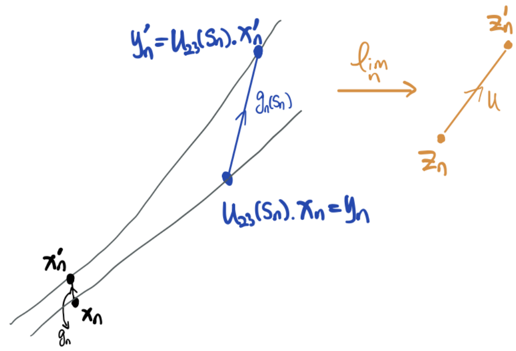
On the other hand, it is direct to verify that:

Lemma 6.15. *Assumption as in last lemma. Also, $s = s_\delta(x,x')$ same as there. Then*

$$\begin{aligned}
&\max\{|g(s)_{11} - 1|, |g(s)_{12}|, |g(s)_{31}|, |g(s)_{32}|\} \leq \varepsilon(x,x'); \\
&\max\{|g(s)_{22} - 1|, |g(s)_{33} - 1|\} \leq 2\sqrt{\varepsilon(x,x')}.
\end{aligned}$$

If one imagines that when $d(x,x')$ (and hence $\varepsilon(x,x')$) is extremely small (compared to δ, ρ_0,...), the matrix $g(s)$ would look like a unipotent matrix in the centralizer of U_{23}:

$$\begin{bmatrix} \approx 1 & \approx 0 & g(s)_{13} \\ g(s)_{21} & \approx 1 & g(s)_{23} \\ \approx 0 & \approx 0 & \approx 1 \end{bmatrix}$$



6.9. **Take the limit.** For $t \in \mathbb{R}$, let

$$\beta_t := \begin{bmatrix} e^{2t} & 0 & 0 \\ 0 & e^{-t} & 0 \\ 0 & 0 & e^{-t} \end{bmatrix},$$

which commutes with U_{23}.

We first find, by pigeon-hole principle, a sequence of pairs $(x_n, x_n') \in \mathscr{C}_E \times \mathscr{C}_E$ with

$$\begin{aligned} &x_n' = \beta_t . x_n \text{ for some } t \in \mathbb{R} \text{ with } t \geq 100; \\ &d(x_n, x_n') < \mathrm{InjRad}(\mathscr{C}_E) \text{ converges to } 0. \end{aligned} \tag{19}$$

Fix some point $x_0 \in \mathscr{C}_E$. For $n \in \mathbb{Z}^+$, cover $\mathscr{C}_E$ by finitely many balls of diameter $< \min\{n^{-1}, \mathrm{InjRad}(\mathscr{C}_E)\}$, then we can find $m < m' \in \mathbb{Z}$ such that $x_n := \beta_{100m}.x_0$ and $x_n' := \beta_{100m'}.x_0$ lies in the same ball. This pair of (x_n, x_n') satisfies Equa.(19) above.

Thus, there exists a unique $g_n := g(x_n, x_n') \in \mathbf{SL}_3(\mathbb{R})$ with $x_n' = g_n.x_n$ and $d(x_n, x_n') = d(\mathrm{I}_3, g_n)$. Further assume that

$$g_n \notin \mathrm{Z}_{\mathbf{SL}_3}(\mathrm{U}_{23}) \tag{20}$$

Consequently, $r_\delta(x_n, x_n') \neq +\infty$. By Lemma 6.17 below, this assumption is satisfied as long as $d(x_n, x_n')$ is small enough.

Now choose $\delta > 0$. For n large enough (such that $\varepsilon(x_n, x_n') < 0.5(1-\rho_0)\delta$ and $g(x_n, x_n') \notin \mathrm{Z}_{\mathbf{SL}_3(\mathbb{R})}(\mathrm{U}_{23})$), apply the unipotent blowup as in the last subsection.

Define

$$y_{n,\delta} := \mathbf{u}_{23}(s_{n,\delta}).x_n, \; y_{n,\delta}' := \mathbf{u}_{23}(s_{n,\delta}).x_n'.$$

Then

$$y_{n,\delta}' = g_{s_{n,\delta}} . y_{n,\delta}$$

with (write $s = s_{n,\delta}$ and $\varepsilon := \varepsilon(x_n, x_n')$ for simplicity)

$$g(s_{n,\delta}) - \mathrm{I}_3 = \begin{bmatrix} \leq \varepsilon & \leq \varepsilon & g(s)_{13} \\ g(s)_{21} & \leq 2\sqrt{\varepsilon} & g(s)_{23} \\ \leq \varepsilon & \leq \varepsilon & \leq 2\sqrt{\varepsilon} \end{bmatrix}$$

with

$$\max\left\{\left|g(s_{n,\delta})_{13}\right|, \left|g(s_{n,\delta})_{21}\right|, \left|g(s_{n,\delta})_{23}\right|\right\} \geq \rho_1 \delta.$$

Passing to a subsequence, assume

$$\lim y_{n,\delta} = z_\delta, \; \lim y_{n,\delta}' = z_\delta', \; \lim g(s_{n,\delta}) = u(\delta)$$

exists. Then $z_\delta, z_\delta' \in \mathscr{C}_E$ and $z_\delta' = u(\delta).z_\delta$ where

$$u(\delta) = \begin{bmatrix} 1 & 0 & u(\delta)_{13} \\ u(\delta)_{21} & 1 & u(\delta)_{23} \\ 0 & 0 & 1 \end{bmatrix}$$

with

$$\rho_1 \delta \leq \max\left\{\left|u(\delta)_{13}\right|, \left|u(\delta)_{21}\right|, \left|u(\delta)_{23}\right|\right\} \leq 3\delta.$$

By Assumption 6.2, Corollary 6.7 and Lemma 6.9, $u(\delta)_{21} = u(\delta)_{13} = 0$.

By the continuity of $\mathscr{C}^{23}_{\bullet}$, we have (note that $x_n' = \beta_t.x_n$ and β_t centralizes U_{23})

$$\mathscr{C}^{23}_{x_n} = \mathscr{C}^{23}_{x_n'} \implies \mathscr{C}^{23}_{y_{n,\delta}} = \mathscr{C}^{23}_{y_{n,\delta}'} \implies \mathscr{C}^{23}_{z_\delta} = \mathscr{C}^{23}_{z_\delta'}.$$

But $u(\delta) \in \mathrm{U}_{23}$, hence

$$u(\delta) \cdot \mathscr{C}^{23}_{z_\delta'} = \mathscr{C}^{23}_{z_\delta}.$$

Thus $\mathscr{C}^{23}_{z_\delta}$ is invariant under translation by $u(\delta)$. By taking a limit point z of (z_δ) as $\delta \to 0$, we see that, by continuity, $\mathrm{U}_{23}.z \subset \mathscr{C}_E$. Hence $(\mathrm{AU}_{23}).z$ is bounded. But this is impossible.

6.10. **No exceptional returns.** We verify Equa.(20) from last subsection. To have slightly better-looking notation, we replace the index $(2,3)$ by $(1,3)$ and β_t is replaced by $\beta'_t := \mathrm{diag}(e^{-t}, e^{2t}, e^{-t})$ accordingly.

Lemma 6.16. *If $M \in \mathbf{SL}_3(\mathbb{Z})$ only has two different eigenvalues, then all eigenvalues of M are ± 1.*

Proof. Let $p(x) := \det(x\mathrm{I}_3 - M) \in \mathbb{Z}[x]$ be the characteristic polynomial of M. By assumption, at least two roots of $p(x)$ are the same. Then $p(x)$ is reducible in $\mathbb{Q}[x]$. If you have not learned Galois theory, then here is a direct way of seeing this. Write $p(x) = (x-\alpha)^2(x-\beta) = x^3 + Ax^2 + Bx + C$ for some $\alpha, \beta \in \mathbb{R}$, $A, B, C \in \mathbb{Q}$. By comparing the coefficients, we see that

$$A = -x_2 - 2x_1,\ B = x_1^2 + 2x_1x_2.$$

The first one implies that $2Ax_1 = -2x_1x_2 - 4x_1^2$, combined with the second one, we get

$$x_1^2 + 2/3Ax_1 + 1/3B = 0.$$

By Euclidean algorithm, the polynomial $q(x) := x^2 + 2/3Ax + 1/3B$ divides $p(x)$. In particular, $p(x)$ is reducible in $\mathbb{Q}[x]$.

Note that $p(x)$ is also reducible in $\mathbb{Z}[x]$ by Gauss lemma. Write $p(x) = (x^2+ax+b)(x-c)$ for some $a, b, c \in \mathbb{Z}$. Since $\det M = 1$, we have $bc = 1$. So $b = c = 1$ or $b = c = -1$. If $x^2 + ax + b$ is irreducible, then it would have two different non-rational roots. So all three roots of p are distinct, contradiction. Hence $p(x) = (x-x_1)(x-x_2)(x-x_3)$ for some $x_i \in \mathbb{Z}$ with $\prod x_i = 1$. So all $x_i = \pm 1$. $\square$

Lemma 6.17. *Take $x \in \mathrm{X}_3$ be such that $\mathrm{A}.x$ is bounded. Assume $\eta \in (0, \mathrm{InjRad}(x))$ is small enough such that*

$$d(\mathrm{I}_3, g) < \eta \implies \|\mathrm{I}_3 - g\|_{\sup} < 0.1.$$

Let $t \geq 100$ be such that $\beta'_t.x = g.x$ with $d(x, g.x) = d(\mathrm{I}_3, g) < \eta$. Then g is not contained in the centralizer of U_{13}. Namely, it is impossible for g to take the form

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \cdot \begin{bmatrix} e^{-s} & 0 & 0 \\ 0 & e^{2s} & 0 \\ 0 & 0 & e^{-s} \end{bmatrix} \cdot \begin{bmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} e^{-s} & 0 & 0 \\ 0 & e^{2s} & 0 \\ 0 & 0 & e^{-s} \end{bmatrix} \cdot \begin{bmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{bmatrix}.$$

Proof. Assume g does take this form and let us derive a contradiction.

By assumption on η, $\mathrm{diag}(1, -1, -1)$ is not allowed. So we have

$$\begin{bmatrix} e^{-s} & 0 & 0 \\ 0 & e^{2s} & 0 \\ 0 & 0 & e^{-s} \end{bmatrix} \begin{bmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{bmatrix} .x = \begin{bmatrix} e^{-t} & 0 & 0 \\ 0 & e^{2t} & 0 \\ 0 & 0 & e^{-t} \end{bmatrix} .x$$

Thus,

$$\begin{bmatrix} e^{-(s-t)} & 0 & 0 \\ 0 & e^{2(s-t)} & 0 \\ 0 & 0 & e^{-(s-t)} \end{bmatrix} \begin{bmatrix} 1 & u_{12} & u_{13} \\ 0 & 1 & u_{23} \\ 0 & 0 & 1 \end{bmatrix} \text{ is conjugate to some element in } \mathbf{SL}_3(\mathbb{Z}).$$

By Lemma 6.16, $s = t$. But our assumption implies that $t \geq 100 > \log(1.1) \geq s$. Contradiction. $\square$

6.11. **Exercises.** Missing!!

7. Lecture 6, Conditional Measures

7.1. **Prelude.** In probability theory, one often has a space, thought of as the collection of all possible "events" together with a probability measure, measuring which event is more likely to happen. Given these data, one can make predictions on "random variables". In mathematical terms, a random variable is just a (measurable) function and the "expectation" of this function is nothing but its integration.

Conditional expectations of a random variable means that we make predictions based on certain information. For instance, one might have another function g on this space and we have perfect knowledge of what the value of g is. So "conditional on" the value of g taken, we make more refined predictions on our random variable.

Conditional expectations, just like expectations, can also be written as integration of the random variable against certain probability measures, known as conditional measures.

From a different perspective, one may also view conditional measures as "Fubini-type theorem".

The material of this lecture is mostly taken from [EW11, chapter 5].

7.2. **Statement of the main theorem.** Let X be a compact metrizable topological space and $\mathscr{C}_X$ be its Borel σ-algebra. Let μ be a probability measure on $(X, \mathscr{B}_X)$. We refer the triple $(X, \mathscr{B}_X, \mu)$ as a compact Borel probability space.

Theorem 7.1. *Let $(X, \mathscr{B}_X, \mu)$ be a compact Borel probability space and $\mathscr{A} \subset \mathscr{B}_X$ be a σ-subalgebra. Then there exist a subset $X' \in \mathscr{A}$ of full μ-measure (i.e. $\mu(X \setminus X') = 0$) and a map $X' \to \mathrm{Prob}(X, \mathscr{B}_X)$, denoted by $x \mapsto \mu_x^{\mathscr{A}}$, satisfying:*

(1) *for every $f \in C(X)$, the map $x \mapsto \int_X f(\omega)\, \mu_x^{\mathscr{A}}(\omega)$ from X' to $\mathbb{R}$ is measurable (w.r.t. $\mathscr{A} \cap X'$) and*
$$\int_{A\cap X'} \left(\int_X f(\omega)\, \mu_x^{\mathscr{A}}(\omega) \right) \mu(x) = \int_A f(\omega)\, \mu(\omega), \quad \forall\, A \in \mathscr{A}.$$
(2) *for every $E \in \mathscr{B}_X$,*
$$x \mapsto \int_X \mathbf{1}_E(\omega)\, \mu_x^{\mathscr{A}}(\omega)$$
is measurable on $(X', \mathscr{A} \cap X')$ and
$$\int_{A\cap X'} \left(\int_X \mathbf{1}_E(\omega)\, \mu_x^{\mathscr{A}}(\omega) \right) \mu(x) = \int_A \mathbf{1}_E(\omega)\, \mu(\omega), \quad \forall\, A \in \mathscr{A}.$$
or in different terms,
$$\int_{A\cap X'} \mu_x^{\mathscr{A}}(E)\, \mu(x) = \mu(A \cap E), \quad \forall\, A \in \mathscr{A}.$$
(3) *If $Y \in \mathscr{A}$ is of full measure and $x \mapsto \nu_x^{\mathscr{A}}$ is another map from Y to $\mathrm{Prob}(X, \mathcal{B}_X)$ satisfying for every f in some dense subset of $C(X)$, the map $x \mapsto \int_X f(\omega)\, \mu_x^{\mathscr{A}}(\omega)$ from Y to $\mathbb{R}$ is measurable (w.r.t. $\mathscr{B}_X \cap Y$) and*
$$\int_{A\cap Y} \left(\int_X f(\omega)\, \nu_x^{\mathscr{A}}(\omega) \right) \mu(x) = \int_A f(\omega)\, \mu(\omega), \quad \forall\, A \in \mathscr{A},$$
then there exists $Y' \subset X' \cap Y$ in $\mathscr{A}$ of full measure such that $\mu_x^{\mathscr{A}} = \nu_x^{\mathscr{A}}$ for all $x \in Y'$;

(4) *If $\mathscr{A}$ is additionally assumed to be countably generated, then one can choose $X'' \subset X'$ in $\mathscr{A}$ of full measure such that $\mu_x^{\mathscr{A}}([x]_{\mathscr{A}}) = 1$* [23] *for every $x \in X''$ and $\mu_y^{\mathscr{A}} = \mu_x^{\mathscr{A}}$ whenever $[x]_{\mathscr{A}} = [y]_{\mathscr{A}} \subset X''$.*

(5) *If $\mathscr{A}_1 \subset \mathscr{A}_2 \subset \ldots$ is an increasing sequence of σ-subalgebras and $\mathscr{A}_\infty$ is the smallest σ-subalgebra containing all of them, then for every $E \in \mathscr{B}$, for μ-almost all x, the relevant conditional measures are defined and*
$$\lim_{n\to\infty} \mu_x^{\mathscr{A}_n}(E) = \mu_x^{\mathscr{A}_\infty}(E).$$

The family of measures $(\mu_x^{\mathscr{A}})$ satisfying condition (1) and (2) as in the theorem are referred to as **conditional measures**.

There are two examples when the conclusion of the theorem (which we leave to the reader to fill in) might be more familiar to the reader.

Example 7.2. *$\mathscr{A}$ is generated by a finite measurable partition $(P_1, P_2, \ldots, P_n)$ of X. If you prefer, you may even take X to be a finite set to see what happens.*

Example 7.3. *$X = [0,1]^2$ and $\mu = \phi(x,y)\mathrm{dxdy}$ where $\phi(x,y)$ is a measurable non-negative function with $\int \phi(x,y)\mathrm{dxdy} = 1$. And*
$$\mathscr{A} := \{A \times [0,1],\ A \text{ is Borel measurable in } [0,1]\}.$$

[23]See Section 7.7 for the definition of $[x]_{\mathscr{A}}$.

7.3. **The set X'.** We are going to construct the measure, thanks to Riesz representation theorem, by specifying the integrals of continuous functions.

First, we choose a countable dense subset $\mathscr{C} \subset C(X)$ containing the constant one function. Let

$$\mathscr{C}_{\mathbb{Q}} := \{ \text{ finite } \mathbb{Q}\text{-linear combinations of elements in } \mathscr{C} \}$$

Thus $\mathscr{C}_{\mathbb{Q}}$ is a countable dense $\mathbb{Q}$-linear subspace of $C(X)$. Let $\pi_{\mathscr{A}}$ denote the orthogonal projection from $L^2(X, \mathscr{B}_X, \mu) \to L^2(X, \mathscr{A}, \mu)$. For every $f \in \mathscr{C}_{\mathbb{Q}}$, choose some representative $f^{\mathscr{A}}$ of $\pi_{\mathscr{A}}([f])$ [24]. Without loss of generality, for the constant one function $\mathbf{1}_X$, which is $\mathscr{A}$-measurable, choose $\mathbf{1}_X^{\mathscr{A}} := \mathbf{1}_X$.

Lemma 7.4. *For every $f \in \mathscr{C}_{\mathbb{Q}}$,*

$$\mu \left\{ x \;\middle|\; \left|f^{\mathscr{A}}(x)\right| > \|f\|_{\sup} \right\} = 0.$$

Proof. The sets $A^\star := \left\{ x \;\middle|\; \star f^{\mathscr{A}}(x) > \|f\|_{\sup} \right\}$ for $\star = +$ or $-$ are in $\mathscr{A}$. Thus their characteristic functions $\mathbf{1}_{A^\star}$ are contained in $L^2(X, \mathscr{B}_X, \mu)$. So if $\mu(A^+) \neq 0$,

$$\|f\|_{\sup} \mu(A^+) \geq \langle [f], \mathbf{1}_{A^+} \rangle = \langle \pi_{\mathscr{A}}([f]), \mathbf{1}_{A^+} \rangle > \|f\|_{\sup} \mu(A^+),$$

a contradiction. So $\mu(A^+) = 0$. Similarly, $\mu(A^-) = 0$ and hence $\mu(A) = 0$. □

Similarly, one shows that

Lemma 7.5. *For every $f \in \mathscr{C}_{\mathbb{Q}}$ with $f \geq 0$, one has $\{x,\ f^{\mathscr{A}}(x) \geq 0\}$ is an element of $\mathscr{A}$ with full measure.*

Lemma 7.6. *For every finite collection $(f_0, f_1, ..., f_n) \subset \mathscr{C}_{\mathbb{Q}}$ and finitely many $(q_1, ..., q_n) \subset \mathbb{Q}$ such that*

$$f_0 = \sum q_i f_i,$$

the set

$$\left\{ x \;\middle|\; f_0^{\mathscr{A}}(x) = \sum q_i f_i^{\mathscr{A}}(x) \right\}$$

is $\mathscr{A}$-measurable and has full measure.

As there are only countably many data, we can find a $\mathscr{A}$-measurable set X' of full measure such that for every $x \in X'$,

(0) $\mathbf{1}_X^{\mathscr{A}}(x) = 1$;
(1) $\left|f^{\mathscr{A}}(x)\right| \leq \|f\|_{\sup}, \quad \forall f \in \mathscr{C}_{\mathbb{Q}}$;
(2) $f^{\mathscr{A}}(x) \geq 0, \quad \forall f \in \mathscr{C}_{\mathbb{Q}},\ f \geq 0$;
(3) $f_0^{\mathscr{A}}(x) = \sum_{i=1}^n q_i f_i^{\mathscr{A}}(x), \quad \forall (f_i)_{i=0}^n \subset \mathscr{C}_{\mathbb{Q}},\ (q_i) \subset \mathbb{Q}$ with $f_0 = \sum_{i=1}^n q_i f_i$.

7.4. **Construction of measures.** For every $x \in X'$ and $f \in C(X)$, find $(f_n) \subset \mathscr{C}_{\mathbb{Q}}$ converging to f in sup-norm. We define $\Lambda_x : C(X) \to \mathbb{R}$ by $\Lambda_x(f) := \lim_{n\to\infty} f_n^{\mathscr{A}}(x)$.

Lemma 7.7. *For $x \in X'$, $(f_n^{\mathscr{A}}(x))$ converges. Consequently, $\Lambda_x(f)$ is well-defined and independent of the choice of (f_n).*

Proof. Take $n, m \in \mathbb{Z}^+$ with $\|f_n - f_m\|_{\sup} \leq \varepsilon$. As $f_n - f_m \in \mathscr{C}$, we have

$$\left|f_n^{\mathscr{A}}(x) - f_m^{\mathscr{A}}(x)\right| = \left|(f_n - f_m)^{\mathscr{A}}(x)\right| \leq \|f_n - f_m\|_{\sup} \leq \varepsilon.$$

This shows that $(f_n^{\mathscr{A}}(x))$ is a Cauchy sequence. □

Also, one sees from the lemma that $\Lambda_x(f)$ is independent of the choice of (f_n). Moreover,

Lemma 7.8. *For $x \in X'$, Λ_x defines a positive bounded linear functional on $C(X)$ sending $\mathbf{1}$ to 1. Therefore, by Riesz representation theorem, there exists a unique Borel probability measure, denoted as $\mu_x^{\mathscr{A}}$, such that $\Lambda_x(f) = \int f(\omega)\mu_x^{\mathscr{A}}(\omega)$.*

Part (1) of Theorem 7.1 is automatically true for $f \in \mathscr{C}_{\mathbb{Q}}$, the general case follows by, say, dominated convergence theorem.

[24] In order to distinguish a genuine function f from its equivalence class up to measure zero, we write $[f]$ for the equivalence class.

7.5. **Extending to measurable functions.** Let O be an open subset of X, by Urysohn lemma (see e.g. 2.12 of [Rud87, Chapter 2]), there exists a sequence of continuous functions (f_n) that is uniformly bounded and converges to $\mathbf{1}_O$. Similarly, one can find a uniformly bounded sequence of continuous functions converging to the characteristic function of a closed subset.

By dominated convergence theorem, Part (2) of Theorem 7.1 holds for E being open or compact. Actually, the characteristic function of $E = O \cap C$, the intersection of some open subset and closed subset (for simplicity, we shall call such a set **locally closed**), can also be pointwisely approximated by a sequence of uniformly bounded continuous functions. Let

$$\mathcal{R} := \{\text{subsets that can be written as a finite disjoint union of locally closed subsets}\}.$$

Lemma 7.9. *$\mathcal{R}$ is an algebra in the sense that it is closed under taking complements, finite intersections and finite unions.*

Proof. For C_1, C_2 closed and O_1, O_2 open, we note that $(C_1 \cap O_1) \cup (C_2 \cap O_2)$ is a disjoint union of locally closed subsets:

$$\begin{aligned}
&(C_1 \cap O_1) \cup (C_2 \cap O_2) \\
=& ((C_1 \cap O_1) \cap (C_2 \cap O_2)) \sqcup (C_1 \cap O_1) \cap (C_2 \cap O_2)^c \\
=& ((C_1 \cap C_2) \cap (O_1 \cap O_2)) \sqcup ((C_1 \cap O_1) \cap (C_2^c \cup O_2^c)) \\
=& ((C_1 \cap C_2) \cap (O_1 \cap O_2)) \sqcup (C_1 \cap O_1 \cap C_2^c) \sqcup (C_1 \cap O_1 \cap O_2^c \cap C_2).
\end{aligned}$$

The rest follows from this. $\square$

On the other hand, let

$$\mathcal{M} := \{\text{subsets of } \mathscr{B}_X \text{ that satisfy part (2) of Theorem 7.1}\}$$

Lemma 7.10. *Let $E_1 \subset E_2 \subset \ldots$ be an increasing sequence of elements in $\mathcal{M}$, then $E_\infty := \bigcup E_i$ belongs to $\mathcal{M}$. If $E \in \mathcal{M}$, then $E^c \in \mathcal{M}$.*

Proof. This follows from dominated convergence theorem. $\square$

Let $\sigma(\mathcal{R})$ denote the smallest σ-subalgebra of $\mathcal{B}_X$ containing $\mathcal{R}$. We have shown that $\mathcal{R} \subset \mathcal{M}$. It is a general fact that if $\mathcal{M}$ is a subset of some σ-algebra satisfying the conclusion of Lemma 7.10 and contains some subalgebra $\mathcal{R}$ as in Lemma 7.9, then $\mathcal{M}$ contains $\sigma(\mathcal{R})$. In the current case, $\sigma(\mathcal{R})$ is $\mathcal{B}_X$, so they are equal.

Lemma 7.11. $\mathcal{M} = \sigma(\mathcal{R})$.

Proof. Let $\mathcal{M}_0$ be the smallest subset of $\mathcal{M}$ containing $\mathcal{R}$ such that the conclusion of Lemma 7.10 holds.

First take $E \in \mathcal{R} \subset \mathcal{M}_0$, consider

$$\mathcal{M}_E := \{F \in \mathcal{M}_0 \mid E \cap F,\, E \cup F,\, E^c \cap F,\, E^c \cup F \in \mathcal{M}_0\}$$

If $F_1 \subset F_2 \subset \ldots$ are contained in $\mathcal{M}_E$, then

$$\begin{aligned}
E \cap (\bigcup F_i) &= \bigcup (E \cap F_i), \quad E \cup (\bigcup F_i) = \bigcup (E \cup F_i), \\
E^c \cap (\bigcup F_i) &= \bigcup (E^c \cap F_i), \quad E^c \cup (\bigcup F_i) = \bigcup (E^c \cup F_i)
\end{aligned}$$

are all contained in $\mathcal{M}_0$. Hence $\cup F_i \in \mathcal{M}_E$. If $F \in \mathcal{M}_E$, then the complements of

$$E \cap F^c,\, E \cup F^c,\, E^c \cap F^c,\, E^c \cup F^c$$

are contained $\mathcal{M}_0$. Thus they are also contained in $\mathcal{M}_0$, implying that $F^c \in \mathcal{M}_E$.

So we have shown that $\mathcal{M}_E$ satisfies the conclusion of Lemma 7.10. On the other hand, $\mathcal{M}_E$ contains $\mathcal{R}$ by Lemma 7.9. By minimality of $\mathcal{M}_0$, we get $\mathcal{M}_E = \mathcal{M}_0$.

For general $E \in \mathcal{M}_0$, $\mathcal{M}_F = \mathcal{M}_0$, $\forall F \in \mathcal{R}$ implies that $\mathcal{R} \subset \mathcal{M}_E$. Same arguments as above show that $\mathcal{M}_E$ satisfies the conclusion of Lemma 7.10. Again by minimality, $\mathcal{M}_E = \mathcal{M}_0$.

Now that $\mathcal{M}_0$ is closed under taking finite unions and intersections, one can directly verify that $\mathcal{M}_0$ is a σ-algebra. This forces $\mathcal{M}_0 = \mathcal{M} = \sigma(\mathcal{R})$. $\square$

Remark 7.12. *Similar arguments are used to prove the π-λ theorem in measure theory.*

7.6. **Uniqueness.** Now we turn to part (3) of Theorem 7.1.

So we have a dense subset $\mathscr{C}$ of $C(X)$ such that for every $f \in \mathscr{C}$ and $A \in \mathscr{A}$

$$\int_{A\cap Y}\left(\int_X f(\omega)\,\nu_x^{\mathscr{A}}(\omega)\right)\mu(x) = \int_A f(\omega)\,\mu(\omega) = \int_{A\cap Y}\left(\int_X f(\omega)\,\mu_x^{\mathscr{A}}(\omega)\right)\mu(x). \tag{21}$$

Let $\mathscr{C}' \subset \mathscr{C}$ be a countable subset that is still dense in $C(X)$. For each $f \in \mathscr{C}'$, let

$$D_f^+ := \left\{x \in X' \cap Y \;\middle|\; \mu_x^{\mathscr{A}}(f) > \nu_x^{\mathscr{A}}(f)\right\}$$
$$D_f^- := \left\{x \in X' \cap Y \;\middle|\; \mu_x^{\mathscr{A}}(f) < \nu_x^{\mathscr{A}}(f)\right\}.$$

Applying Equa.(21) we see that both D_f^+ and D_f^- has measure zero. Let Y' be the complement of their unions as f varies in $\mathscr{C}'$. Then Y' is full in X. And for every $x \in Y'$ and every $f \in \mathscr{C}'$,

$$\int_X f(\omega)\,\mu_x^{\mathscr{A}}(\omega) = \int_X f(\omega)\,\nu_x^{\mathscr{A}}(\omega)$$

which extends to all $f \in \mathscr{C}'$ by dominated convergence theorem. So $\mu_x^{\mathscr{A}} = \nu_x^{\mathscr{A}}$ for all $x \in Y'$.

7.7. **Countably generated sigma-subalgebras.** Now let $\mathscr{A}$ be a countably generated σ-subalgebra of $\mathcal{B}_X$. For $x \in X$, define

$$[x]_{\mathscr{A}} := \bigcap_{x\in A\in\mathscr{A}} A.$$

We sometimes refer to $[x]_{\mathscr{A}}$ as the **atom** containing x.

Lemma 7.13. *Take $(A_1, A_2, ...)$ be such that $\mathscr{A}$ is the smallest σ-subalgebra of $\mathcal{B}_X$ containing all A_i's. Fix $x \in X$ and let*

$$B_i := \begin{cases} A_i, & \text{if } x \in A_i \\ A_i^c, & \text{if } x \notin A_i. \end{cases}$$

Then $[x]_{\mathscr{A}} = \bigcap_{i=1}^\infty B_i$. In particular, $[x]_{\mathscr{A}} \in \mathscr{A}$.

Proof. Fix some $x \in X$ and let $[x]' := \bigcap_{i=1}^\infty B_i$. Consider

$$\mathscr{A}_x' := \{A \in \mathscr{A} \mid \text{either } [x]' \subset A \text{ or } [x]' \subset A^c\}$$

Then one verifies that $\mathscr{A}_x'$ is a σ-algebra containing all A_i's. Thus it is equal to $\mathscr{A}$. This proves the lemma. $\square$

Fix a countable generator $(A_1, A_2, ...)$ of $\mathscr{A}$. Let $(B_1, B_2, ...)$ be obtained by including their complements.

By part (2) of the theorem, we can find $X_1 \in \mathscr{A}$ contained in X' (defined by intersection of all X'_{B_i}'s) of full measure such that

$$\int_{A\cap X_1}\left(\int_X \mathbf{1}_{B_i}(\omega)\,\mu_x^{\mathscr{A}}(\omega)\right)\mu(x) = \int_A \mathbf{1}_{B_i}(\omega)\,\mu(\omega), \quad \forall\, A \in \mathscr{A}.$$

Let $N_i \in \mathscr{A}$ defined by $\left\{x \in B_i \cap X_1 \;\middle|\; \mu_x^{\mathscr{A}}(B_i) \neq 1\right\}$. Then the equation above (with $A := N_i$) shows that $\mu(N_i) = 0$. Let $X'' := X_1 \setminus \bigcup N_i$.

Then for $x \in X''$ and $x \in B_i$, one has $\mu_x^{\mathscr{A}}(B_i) = 1$. Hence $\mu_x^{\mathscr{A}}([x]_{\mathscr{A}}) = 1$.

As for each $f \in \mathscr{C}$, the map $x \mapsto \mu_x^{\mathscr{A}}(f)$ is $\mathscr{A}$-measurable (on X'), it must be constant on each atom. So $\mu_x^{\mathscr{A}} = \mu_y^{\mathscr{A}}$ whenever $[x]_{\mathscr{A}} = [y]_{\mathscr{A}} \subset X'$.

7.8. **Pointwise convergence.** Choose $X' \subset X$ such that for $\star = \mathscr{A}_i$ or $\mathscr{A}_\infty$, conditional measures $\mu_x^\star$ are defined and (1) and (2) in this theorem hold.

By (2) of the theorem, for every $E \in \mathscr{A}_\infty$, (the equivalence class of) the function $x \mapsto \mu_x^\star(E)$ is just $\pi_\star([\mathbf{1}_E])$ for $\star = \mathscr{A}_?$ for $? = 1, 2, ...$ or ∞.

First note that

Lemma 7.14. *$\bigcup_{i=1}^\infty L^2(X, \mathscr{A}_i, \mu)$ is dense in $L^2(X, \mathscr{A}_\infty, \mu)$.*

Proof. It suffices to show that for every $E \in \mathscr{A}_\infty$ and $\varepsilon > 0$, there exists $F \in \bigcup_{i=1}^\infty \mathscr{A}_i$ such that $\mu(E\Delta F) < \varepsilon$. Let $\mathscr{A}'$ collect all $E \in \mathscr{A}_\infty$ such that this holds, then one checks that $\mathscr{A}'$ is a σ-subalgebra containing all the $\mathscr{A}_i$'s. Consequently, $\mathscr{A}' = \mathscr{A}_\infty$ and the proof is complete. $\square$

Thus the convergence in L^2-norm is guaranteed by general facts from Hilbert spaces. However, going from L^2-convergence to pointwise convergence requiring us to pay attention to the special choices of subspaces here (think about Carleson's difficult theorem on pointwise convergence of Fourier series). The key is the following "maximal inequalities":

Lemma 7.15. *For $f \in L^1(X, \mathscr{B}, \mu)$ (in application, $f = \mathbf{1}_E - \mathbf{1}_F$) and $\lambda > 0$, let*

$$E(\lambda) := \left\{ x \in X' \;\middle|\; \max_{n \in \mathbb{Z}^+} \mu_x^{\mathscr{A}_n}(f) > \lambda \right\}$$

Then

$$\mu\left(E(\lambda)\right) \leq \lambda^{-1} \left\| f \right\|_1 .$$

Proof. If all $\mathscr{A}_i$'s are the same, this is just Minkowski inequality. In general, let

$$\begin{aligned}
&F_1 := \left\{ x \in X' \mid \mu_x^{\mathscr{A}_1}(f) > \lambda \right\} \in \mathscr{A}_1; \\
&F_2 := \left\{ x \in X' \setminus F_1 \mid \mu_x^{\mathscr{A}_2}(f) > \lambda \right\} \in \mathscr{A}_2; \\
&F_3 := \left\{ x \in X' \setminus (F_1 \cup F_2) \mid \mu_x^{\mathscr{A}_3}(f) > \lambda \right\} \in \mathscr{A}_3; \\
&\ldots\ldots
\end{aligned}$$

Then $E(\lambda) = \bigsqcup_{k=1}^{\infty} F_k$. For every $k \in \mathbb{Z}^+$,

$$\lambda\mu(F_k) \leq \int_{F_k} \mu_x^{\mathscr{A}_k}(f)\mu(x) = \int_{F_k} f(\omega)\mu(\omega) \leq \int_{F_k} \left| f(\omega) \right| \mu(\omega).$$

Thus,

$$\mu(E(\lambda)) = \sum \mu(F_k) \leq \lambda^{-1} \sum \int_{F_k} \left| f(\omega) \right| \mu(\omega) \leq \lambda^{-1} \left\| f \right\|_1 .$$

□

Proof of (5) of Theorem 7.1. Fix $E \in \mathscr{A}_\infty$ and we would like to show that $\mu_x^{\mathscr{A}_n}(E)$ converges to $\mu_x^{\mathscr{A}_\infty}(E)$ almost surely. So for $\varepsilon > 0$, let

$$E(\varepsilon) := \left\{ x \mid \limsup \left| \mu_x^{\mathscr{A}_n}(E) - \mu_x^{\mathscr{A}_\infty}(E) \right| > \varepsilon \right\}.$$

It suffices to show that $\mu(E(\varepsilon)) \leq 4\varepsilon$ for every $\varepsilon > 0$.

Take $k = k(\varepsilon) \in \mathbb{Z}^+$ and $F \in \mathscr{A}_k$ such that

$$\left\| \mathbf{1}_E - \mathbf{1}_F \right\|_2 < \varepsilon^2.$$

Note that

$$\limsup \left| \mu_x^{\mathscr{A}_n}(E) - \mu_x^{\mathscr{A}_\infty}(E) \right| \leq \limsup \left| \mu_x^{\mathscr{A}_n}(E) - \mu_x^{\mathscr{A}_n}(F) \right| + \limsup \left| \mu_x^{\mathscr{A}_n}(F) - \mu_x^{\mathscr{A}_\infty}(E) \right|.$$

And for n larger than k, $\mu_x^{\mathscr{A}_n}(F) = \mu_x^{\mathscr{A}_\infty}(F) = \mathbf{1}_F(x)$ almost surely. So $E(\varepsilon) \subset F(\varepsilon) \cup G(\varepsilon)$ where

$$\begin{aligned}
F(\varepsilon) &:= \left\{ x \;\middle|\; \limsup_n \left| \mu_x^{\mathscr{A}_n}(E) - \mu_x^{\mathscr{A}_n}(F) \right| > 0.5\varepsilon \right\}, \\
G(\varepsilon) &:= \left\{ x \mid \left| \mu_x^{\mathscr{A}_\infty}(F) - \mu_x^{\mathscr{A}_\infty}(E) \right| > 0.5\varepsilon \right\}.
\end{aligned}$$

By Lemma 7.15, we have

$$\mu(F(\varepsilon)),\ \mu(G(\varepsilon)) \leq \frac{2}{\varepsilon} \left\| \mathbf{1}_E - \mathbf{1}_F \right\|_1 \leq 2\varepsilon.$$

Hence $\mu(E(\varepsilon)) \leq 4\varepsilon$ as desired. □

7.9. **Exercises.** Missing!!

8. Lecture 7, Dimension and Entropy

For more on dimension of metric spaces, see Mattilde's book [Mat95]. Hochman's notes[25] are recommended as an introduction to entropy in dynamics.

From this lecture on, we will loosely follow the EKL paper [EKL06] and EL's Pisa notes[26].

[25] Available here: http://math.huji.ac.il/~mhochman/courses/dynamics2014/notes.5.pdf

[26] Available here: https://people.math.ethz.ch/~einsiedl/Pisa-Ein-Lin.pdf.

8.1. **Upper Minkowski dimension.** Define a metric d on $[0,1)^2$ by

$$d(\mathbf{x},\mathbf{y}) := \inf\{\|\mathbf{x}-\mathbf{y}-\mathbf{v}\|_{\sup},\ \mathbf{v}\in\mathbb{Z}^2\}.$$

Replacing sup-norm by the usual Euclidean norm has no effect the definition of dimension below. But we find it slightly more convenient to work with the sup-norm. This metric is compatible with the topology defined by identifying $[0,1)^2$ with $\mathbb{R}^2/\mathbb{Z}^2$.

For a subset $E \subset [0,1)^2$, define for $s>0, \varepsilon>0$,

$$\mathcal{H}^s_\varepsilon(E) := \inf\left\{\sum \operatorname{diam}(B_i)^s \;\middle|\; (B_i) \text{ countable open balls covering } E \text{ of diameter} < \varepsilon\right\}$$
$$\mathcal{H}^s(E) := \lim_{\varepsilon\to 0}\mathcal{H}^s_\varepsilon(E).$$

The **Hausdorff dimension** of E is defined by

$$\dim_{\mathrm{H}}(E) := \inf\{s>0 \mid \mathcal{H}^s(E)=0\} = \inf\{s>0 \mid \mathcal{H}^s(E)<+\infty\}.$$

What is more relevant to us is the notion of upper Minkowski dimension (also called box dimension), which is larger than the Hausdorff dimension.

Definition 8.1. *Given a compact metric space (X,d) and $\varepsilon>0$, a subset $S\subset X$ is said to be an ε-**separating set** iff*

$$x,y\in S \text{ and } x\neq y \implies d(x,y)>\varepsilon.$$

*Let E be a subset of X and $\operatorname{Sep}(E,\varepsilon)$ denote the largest size of ε-separating sets contained in E. Then the **upper Minkowski dimension** is*

$$\dim_\square(E) := \limsup_{\varepsilon\to 0}\frac{\log(\operatorname{Sep}(E,\varepsilon))}{\log(\varepsilon^{-1})}$$

if $\operatorname{Sep}(E,\varepsilon)$ is finite for ε small enough. Otherwise $\dim_\square(E) := +\infty$.

Lemma 8.2. *Hausdorff dimension of a set is no greater than its upper Minkowski dimension.*

Proof. By definition, it suffices to show that

$$\begin{aligned}&\forall\, s>\dim_\square(E),\ \exists\, C>0,\ \forall\,\varepsilon>0,\\ &\exists \text{ covering of } E \text{ by countably many balls of radius} < \varepsilon\\ &\text{such that } \sum \operatorname{diam}(B_i)^s < C.\end{aligned}$$

Take such an s. By definition, for ε small enough, one has

$$\frac{\log \operatorname{Sep}(E,\varepsilon)}{\log(\varepsilon^{-1})} < s, \text{ equivalently, } \operatorname{Sep}(E,\varepsilon) < \varepsilon^{-s}.$$

Let $S=\{s_1,....,s_l\}\subset E$ be a ε-separated set with $l=\operatorname{Sep}(E,\varepsilon)$. Then it is a maximal ε-separating set. Let $B_i := B_{2\varepsilon}(s_i)$, the ball of radius 2ε centered at s_i. Then (B_i) forms a covering of E by balls of diameter 4ε. And

$$\sum \operatorname{diam}(B_i)^s = 4^s\varepsilon^s l < 4^s.$$

As RHS is independent of ε, we are done. □

8.2. **Measure entropy of finite partitions.** Let $(X,\mathscr{B})$ be a set equipped with a σ-algebra and μ be a probability measure on $(X,\mathscr{B})$. The triple $(X,\mathscr{B},\mu)$ is often referred to as a **probability space**.

In our examples, X is often the underlying set of a compact metrizable topological space and $\mathscr{B}$ is the Borel σ-algebra: the smallest σ-algebra containing all open and closed subsets of X. In this case the triple $(X,\mathscr{B},\mu)$ is referred to as a Borel probability space.

A **finite measurable partition**[27] is a set of measurable subsets $\mathcal{P}=\{P_1,...,P_l\}\subset\mathscr{B}$ of X such that

$$X = \bigsqcup_{i=1}^{l} P_i.$$

We define the entropy of a partition $\mathcal{P}=\{P_i\}$ by

$$H_\mu(\mathcal{P}) := \sum_{i=1}^{l} -\mu(P_i)\log(\mu(P_i)).$$

[27] Sometimes the word "measurable" is omitted.

where by convention,

$$-0 \cdot \log(0) := 0.$$

If $\phi(x) := -x\log(x)$ defined on $[0,1]$, then ϕ is strictly convex/concave in the sense that for every $\sum_{i=1}^{l} \lambda_i = 1$ with $\lambda_i > 0$, one has

$$\sum_{i=1}^{l} \lambda_i \phi(x_i) \leq \phi(\sum_{i=1}^{l} \lambda_i x_i), \quad \forall\, x_1, ..., x_l \in [0,1]$$

and "$=$" holds iff $x_1 = x_2 = ... = x_l$.

Entropy of a partition is a non-negative number and it is zero iff the partition consists of null ($\mu(P_i) = 0$) and co-null ($\mu(P_i^c) = 0$) sets.

Lemma 8.3. *Let $\mathcal{P} = (P_i)_{i=1}^{d}$ be a finite measurable partition, then*

- $H_\mu(\mathcal{P}) \leq \log d$;
- $H_\mu(\mathcal{P}) = \log d$ *iff* $\mu(P_i) = d^{-1}$ *for every* $i = 1, ..., d$.

Proof. This is a consequence of the convexity/concavity of $x \mapsto -x\log(x)$. □

Given two partitions $\mathcal{P} = (P_i)$ and $\mathcal{Q} = (Q_j)$, let $\mathcal{P} \vee \mathcal{Q}$ be the partition consisting of $\{P_i \cap Q_j\}$ as i, j vary. We define the entropy of $\mathcal{Q}$ **conditional on** $\mathcal{P}$ by

$$H_\mu(\mathcal{Q}|\mathcal{P}) := \sum_{i,j} -\mu(P_i \cap Q_j) \log \frac{\mu(P_i \cap Q_j)}{\mu(P_i)}.$$

If we let $\mu_i^{\mathcal{P}}$ denote the probability measure $\frac{1}{\mu(P_i)}\mu|_{P_i}$[28] whenever $\mu(P_i) \neq 0$, then

$$H_\mu(\mathcal{Q}|\mathcal{P}) = \sum_i \mu(P_i) H_{\mu_i^{\mathcal{P}}}(\mathcal{Q}).$$

Lemma 8.4. *Let $\mathcal{P} = (P_i)$ and $\mathcal{Q} = (Q_j)$ be two finite partitions. Then*

$$\max\{H_\mu(\mathcal{P}), H_\mu(\mathcal{Q})\} \leq H_\mu(\mathcal{P} \vee \mathcal{Q}) \leq H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q}).$$

Actually, we have

$$H_\mu(\mathcal{P} \vee \mathcal{Q}) = H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q}|\mathcal{P}) = H_\mu(\mathcal{Q}) + H_\mu(\mathcal{P}|\mathcal{Q})$$
$$0 \leq H_\mu(\mathcal{Q}|\mathcal{P}) \leq H_\mu(\mathcal{Q}),\ 0 \leq H_\mu(\mathcal{P}|\mathcal{Q}) \leq H_\mu(\mathcal{P}).$$

Proof. Firstly, a direct computation shows that

$$\sum_{i,j} -\mu(P_i \cap Q_j) \log(\mu(P_i \cap Q_j)) = \sum_{i,j} -\mu(P_i \cap Q_j) \log(\frac{\mu(P_i \cap Q_j)}{\mu(P_i)}) + \sum_i \mu(P_i) \log(\mu(P_i)).$$

So $H_\mu(\mathcal{P} \vee \mathcal{Q}) = H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q}|\mathcal{P})$. That $0 \leq H_\mu(\mathcal{Q}|\mathcal{P})$ follows from the definition. It only remains to show that $H_\mu(\mathcal{Q}|\mathcal{P}) \leq H_\mu(\mathcal{Q})$. By the convexity/concavity of $-x\log(x)$, we have for each fixed j,

$$\begin{aligned}
&\sum_i \mu(P_i) \left(-\frac{\mu(P_i \cap Q_j)}{\mu(P_i)} \log \left(\frac{\mu(P_i \cap Q_j)}{\mu(P_i)} \right) \right) \\
\leq& -\left(\sum_i \mu(P_i) \frac{\mu(P_i \cap Q_j)}{\mu(P_i)} \right) \cdot \log \left(\sum_i \mu(P_i) \frac{\mu(P_i \cap Q_j)}{\mu(P_i)} \right) \\
=& -\mu(Q_j) \log(\mu(Q_j)).
\end{aligned}$$

Summing over j completes the proof. □

8.3. **Dynamical entropy.** Let $T : (X, \mathscr{B}) \to (X, \mathscr{B})$ be a measurable map ($T^{-1}\mathscr{B} \subset \mathscr{B}$) preserving the measure μ. For a finite partition $\mathcal{P}$, define

$$h_\mu(T, \mathcal{P}) := \lim_{n \to +\infty} \frac{1}{n} H_\mu(\mathcal{P} \vee T^{-1}\mathcal{P} \vee ... \vee T^{-(n-1)}\mathcal{P}).$$

Lemma 8.5. *The limit indeed exists in* $[0, +\infty]$. *Also,*

$$h_\mu(T, \mathcal{P}) = \inf_{n \in \mathbb{Z}^+} \frac{1}{n} H_\mu(\mathcal{P} \vee T^{-1}\mathcal{P} \vee ... \vee T^{-(n-1)}\mathcal{P}).$$

[28]The notation $\mu|_{P_i}$ means the restriction of μ to P_i, namely, $\mu|_{P_i}(E) := \mu(P_i \cap E)$.

Proof. Fix $\mathcal{P}$, let $a_n := H_\mu(\mathcal{P} \vee T^{-1}\mathcal{P} \vee ... \vee T^{-(n-1)}\mathcal{P})$. Then the sequence (a_n) is non-negative and satisfies

$$a_{n+m} \leq a_n + a_m.$$

For any such sequence, similar conclusion holds. Indeed, we show that for every fixed $n \in \mathbb{Z}^+$ and $\varepsilon > 0$, there exists N_0 such that for every $N > N_0$,

$$\frac{a_N}{N} < \frac{a_n}{n} + \varepsilon.$$

Let $C_n := \max a_1, ..., a_n$. Write $N = dn + r$ with $d \in \mathbb{Z}_{\geq 0}$ and $r \in \{0, 1, ..., n-1\}$. Then $a_N \leq da_n + a_r$ and

$$\frac{a_N}{N} \leq \frac{da_n}{dn + r} + \frac{a_r}{N} \leq \frac{a_n}{n} + \frac{c_n}{N}.$$

So taking N_0 such that $c_n < \varepsilon N_0$ suffices. $\square$

Define the **measure entropy** of T with respect to μ as

$$h_\mu(T) := \sup \{h_\mu(T, \mathcal{P}) \mid \mathcal{P} \text{ is a finite partition }\}.$$

8.4. **Main theorem.** Recall:

$$\alpha_t := \begin{bmatrix} e^t & 0 & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^{-2t} \end{bmatrix} \in \mathrm{A}^+ := \left\{ \begin{bmatrix} e^{t_1} & 0 & 0 \\ 0 & e^{t_2} & 0 \\ 0 & 0 & e^{t_3} \end{bmatrix} \middle| \sum t_i = 0,\ t_1, t_2 > 0 \right\}$$

and for $\alpha, \beta \in [0, 1)$,

$$\Lambda_{\alpha,\beta} := \begin{bmatrix} 1 & 0 & \alpha \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{bmatrix} .\mathbb{Z}^3 \in \mathrm{X}_3, \quad \mathbf{u}^+_{\alpha,\beta} := \begin{bmatrix} 1 & 0 & \alpha \\ 0 & 1 & \beta \\ 0 & 0 & 1 \end{bmatrix}.$$

The map from $\mathbb{R}^2/\mathbb{Z}^2$ to X_3 induced by $(\alpha, \beta) \mapsto \Lambda_{\alpha,\beta}$ is continuous.

The main purpose of this lecture is to explain:

Theorem 8.6. *Let $\mathscr{C}$ be a compact subset of X_3. Let*

$$E_{\mathscr{C}} := \left\{(\alpha, \beta) \in [0, 1)^2 \,\middle|\, \mathrm{A}^+.\Lambda_{\alpha,\beta} \subset \mathscr{C}\right\}.$$

If $E_{\mathscr{C}}$ has positive upper Minkowski dimension, then $\mathscr{C}$ supports an A-invariant measure ν with $h_\nu(\alpha_1) > 0$.

Later it will be shown that such a measure can not exist, from which we deduce that $E_{\mathscr{C}}$ has zero upper Minkowski dimension and hence zero Hausdorff dimension. Consequently, the exceptional set to Littlewood conjecture is a countable union of sets with box dimension zero. In particular, it has Hausdorff dimension zero.

8.5. **Outline of the proof.** The proof of Theorem 8.6 consists of two steps

Step 1. Construct an α_1-invariant measure with positive entropy;

Step 2. Use an average process to promote it to an A-invariant measure. The point is that the entropy does not decrease when passing to the limit.

From now on, we fix such a compact subset as in Theorem 8.6 and call it $\mathscr{C}_1$ till the end of the proof. Also

1. fix some $\delta_1 \in (0, 1)$ such that $\dim_\square(E_{\mathscr{C}_1}) > \delta_1$;
2. fix some $\delta_2 \in (0, 1)$ such that $\mathrm{sys}(x) > \delta_2$ for every $x \in \mathscr{C}_1$.

Furthermore, choose $\delta_3, \delta_4 \in (0, 1)$ such that $B(\delta_4) \subset \mathcal{O}_{e^{-3}\delta_3} \subset \mathcal{O}_{\delta_3} \subset B(\delta_2)$. Consequently,

3. $e^{-3}\delta_3 \leq \|(s_1, s_2)\| \leq \delta_3 \implies \delta_4 < d^{\mathrm{X}_3}(\mathbf{u}^+_{s_1,s_2}.x, x) < \delta_2$ for every $x \in \mathscr{C}_1$.

By making $\delta_4 > 0$ even smaller, we assume

$$d(\mathbf{s}, \mathbf{t}) > e^{-3}\delta_3 \implies d(\Lambda_\mathbf{s}, \Lambda_\mathbf{t}) > \delta_4,\ \forall \mathbf{s}, \mathbf{t} \in [0, 1)^2 \cong \mathbb{R}^2/\mathbb{Z}^2.$$

Also we decompose $[0, 1)^2 = \bigcup_{i=1}^{l_0} \square_i$ into union of subsets of diameter smaller than δ_3. Hence for $\mathbf{s}, \mathbf{t}$ contained in the same $\square_i$, one has $d^{\mathrm{X}_3}(\Lambda_\mathbf{s}, \Lambda_\mathbf{t}) < \delta_2$.

8.6. **Step 1, construction of the measure.** By assumption, we can find a sequence of positive numbers (ε_n) decreasing to 0 such that

$$\frac{\log\left(\mathrm{Sep}(E_{\mathscr{C}_1}, \varepsilon_n)\right)}{\log(\varepsilon_n^{-1})} > \delta_1,$$

or equivalently,

$$\mathrm{Sep}(E_{\mathscr{C}_1}, \varepsilon_n) > \left(\frac{1}{\varepsilon_n}\right)^{\delta_1}.$$

Let $\mathcal{S}_n$ be an ε_n-separating set for $(E_{\mathscr{C}_1}, d^{\mathrm{X}_3})$ contained in some $\square_i$ whose size is at least $l_0^{-1}\,\mathrm{Sep}(E_{\mathscr{C}_1}, \varepsilon_n)$.

For a non-empty finite subset $F \subset \mathrm{X}_3$, let m_F denote the uniform probability measure supported on F, namely,

$$\mathrm{m}_F(E) := \frac{\# F \cap E}{\# F}.$$

For n large enough such that $\varepsilon_n < e^{-3}\delta_3$, choose $d_n \in \mathbb{Z}^+$ such that $\delta_3 < e^{3d_n}\varepsilon_n \leq e^3\delta_3$. Let

$$\mu_n := \frac{1}{d_n}\sum_{i=0}^{d_n-1}(\alpha_1)^i_*\mathrm{m}_{S_n} = \frac{1}{d_n}\sum_{i=0}^{d_n-1}(\alpha_i)_*\mathrm{m}_{S_n}.$$

By assumption, (μ_n) is a sequence of probability measures supported on $\mathscr{C}_1$. By the "diagonal argument", we can select a convergent subsequence (μ_{n_k}) under the weak$*$ topology. Let μ denote the limit measure.

Lemma 8.7. *The limit measure μ is α_1-invariant.*

Proof. Indeed, as $n \to \infty$,

$$(\alpha_1)_*\mu_n - \mu_n = \frac{1}{d_n}\left((\alpha_{d_n})_*\mathrm{m}_{S_n} - \mathrm{m}_{S_n}\right)$$

converges to 0. $\square$

8.7. **Separation properties under iterations.**

Lemma 8.8. *For every pair of distinct points $\mathbf{s}, \mathbf{t} \in \mathcal{S}_n$, there exists $j \in \{0, 1, ..., d_n - 1\}$ such that*

$$d(\alpha_j.\Lambda_{\mathbf{s}}, \alpha_j.\Lambda_{\mathbf{t}}) \geq \delta_4.$$

Proof. When $d(\mathbf{s}, \mathbf{t}) > e^{-3}\delta_3$, then the conclusion holds for $j = 0$.

Now assume $d(\mathbf{s}, \mathbf{t}) \leq e^{-3}\delta_3$ and let $\mathbf{t}' \in \mathbf{t} + \mathbb{Z}^2$ be such that $d(\mathbf{s}, \mathbf{t}) = \|\mathbf{s} - \mathbf{t}'\|_{\sup}$. By our choice of d_n, there exists $j \in \{0, 1, ..., d_n - 1\}$ such that

$$\left\|e^{3j}(\mathbf{s} - \mathbf{t}')\right\|_{\sup} = e^{3j}\left\|\mathbf{s} - \mathbf{t}'\right\|_{\sup} > e^{-3}\delta_3.$$

We choose j to be the smallest one with this property. Then

$$e^{-3}\delta_3 < \left\|e^{3j}\mathbf{s} - e^{3j}\mathbf{t}'\right\|_{\sup} \leq \delta_3, \text{ which implies } d(\mathbf{u}_{e^{3j}\mathbf{s}-e^{3j}\mathbf{t}'}.x, x) > \delta_4,\ \forall\, x \in \mathscr{C}_1. \quad (22)$$

Since

$$\alpha_j.\Lambda_{\mathbf{s}} = \mathbf{u}^+_{e^{3j}\mathbf{s}-e^{3j}\mathbf{t}'}.\alpha_j.\Lambda_{\mathbf{t}'} = \mathbf{u}^+_{e^{3j}\mathbf{s}-e^{3j}\mathbf{t}'}\alpha_j.\Lambda_{\mathbf{t}}$$

and

$$\alpha_j.\Lambda_{\mathbf{t}} \in \mathscr{C}_1,$$

we have by Equa.(22)

$$d\left(\alpha_j.\Lambda_{\mathbf{s}}, \alpha_j.\Lambda_{\mathbf{t}}\right) > \delta_4.$$

$\square$

8.8. **Test partitions.** Let X be a compact metrizable space. For a subset $E \subset X$, let $\mathrm{Int}(E)$ be its interior points, $\overline{E}$ its closure, E^c its complement and ∂E its boundary.

Lemma 8.9. *For every $\varepsilon > 0$, there exists a finite measurable partition $\mathcal{P}$ of $\mathscr{C}_1$ such that $\mu(\partial P) = 0$ and $\mathrm{diam}(P) < \varepsilon$ for every $P \in \mathcal{P}$.*

Proof. For every $x \in \mathscr{C}_1$, find $0 < r_x < 0.5\varepsilon$ such that $\mu(\partial B_x(r_x)) = 0$. Indeed, the sets

$$\partial B_x(r), \quad 0 < r < 0.5\varepsilon$$

form an uncountable family of disjoint measurable subsets. Thus one of them must have zero μ-measure. By compactness, we find $x_1, ..., x_k \in \mathscr{C}_1$ such that

$$\mathscr{C}_1 \subset \bigcup_{i=1}^{k} B_{x_i}(r_{x_i}).$$

Define

$$P_1 := B_{x_1}(r_{x_1}),\ P_2 := B_{x_2}(r_{x_2}) \setminus B_{x_1}(r_{x_1}),\ P_3 := B_{x_3}(r_{x_3}) \setminus \left(B_{x_1}(r_{x_1}) \cup B_{x_2}(r_{x_2})\right), ...$$

Note that $\partial(A \cap B) \subset \partial A \cup \partial B$ and $\partial(A^c) = \partial(A)$. Then

$$\partial P_j \subset \bigcup_{i \leq j} \partial B_{x_i}(r_{x_i})$$

has μ-measure zero. Thus $\mathcal{P} := (P_1, P_2, ..., P_k)$ is a desired partition. □

Lemma 8.10. *Let (ν_n) be a sequence of Borel probability measures converging to ν in weak* topology, then for every Borel measurable subset $E \subset X$ with $\nu(\partial E) = 0$, one has $\nu(E) = \lim_{n\to\infty} \nu_n(E)$.*

Proof. Without loss of generality, we assume E is bounded. Take an open bounded set F containing $\overline{E}$.

Choose a sequence of continuous functions (f_k) (resp. (g_k)) such that $f_k \leq \mathbf{1}_{\mathrm{Int}(E)}$ (resp. $g_k \leq \mathbf{1}_{F\setminus\overline{E}}$) and (f_k) converges to $\mathbf{1}_{\mathrm{Int}(E)}$ (resp. (g_k) converges to $\mathbf{1}_{F\setminus\overline{E}}$). Then

$$\begin{aligned}\nu(\mathrm{Int}(E)) &= \lim_{k\to\infty} \int f_k(x)\nu(x) \\ &= \lim_{k\to\infty}\lim_{n\to\infty} \int f_k(x)\nu_n(x) \\ &\leq \liminf_{n\to\infty} \nu_n(\mathrm{Int}(E)).\end{aligned}$$

Let

$$F_k(x) := \begin{cases} 1 - g_k(x) & x \in F \\ 0 & x \notin F \end{cases}.$$

Then (F_k) is a sequence of continuous functions such that $\mathbf{1}_{\overline{E}} \leq F_k \leq \mathbf{1}_F$ for every k and converges to $\mathbf{1}_{\overline{E}}$ pointwise. Therefore,

$$\begin{aligned}\nu(\overline{E}) &= \lim_{k\to\infty} \int F_k(x)\nu(x) \\ &= \lim_{k\to\infty}\lim_{n\to\infty} \int F_k(x)\nu_n(x) \\ &\geq \limsup_{n\to\infty} \nu_n(\overline{E}).\end{aligned}$$

Putting together we have

$$\nu(\mathrm{Int}(E)) \leq \liminf_{n\to\infty} \nu_n(\mathrm{Int}(E)) \leq \limsup_{n\to\infty} \nu_n(\overline{E}) \leq \nu(\overline{E}).$$

But $\nu(\partial E) = 0$, so the above inequalities are all equalities and we are done. □

8.9. **Completion of step 1.** In this subsection we complete step one, namely, we show

Lemma 8.11. *Let μ be as constructed in Section 8.6. Then $h_\mu(\alpha_1) \geq 3\delta_1$.*

We fix a finite measurable partition $\mathcal{P}$ as in Lemma 8.9 with $\varepsilon = \delta_4$. For every k and $P \in \mathcal{P} \vee \alpha_1^{-1}\mathcal{P} \vee ... \vee \alpha_1^{-(k-1)}\mathcal{P}$,

$$\partial P \subset \partial P_{i_0} \cup \partial\alpha_1^{-1}(P_{i_1}) \cup ... \cup \partial\alpha_1^{-(k-1)}(P_{i_{k-1}}) = \partial P_{i_0} \cup \alpha_1^{-1}(\partial P_{i_1}) \cup ... \cup \alpha_1^{-(k-1)}(\partial P_{i_{k-1}})$$

has μ-measure zero since μ is α_1-invariant by Lemma 8.7. It is sufficient to show that $h_\mu(T, \mathcal{P}) \geq 3\delta_1$.

For two integers $i < j$, abbreviate

$$\mathcal{P}_i^j := \alpha_1^{-i}\mathcal{P} \vee \alpha_1^{-(i+1)}\mathcal{P} \vee ... \vee \alpha_1^{-j}\mathcal{P}.$$

By Lemma 8.10, for each fixed k,

$$\frac{1}{k} H_\mu(\mathcal{P}_0^{k-1}) = \frac{1}{k} \lim_{n\to\infty} H_{\mu_n}(\mathcal{P}_0^{k-1}). \tag{23}$$

Lemma 8.12. *Let ν_1, ν_2 be two probability measures, $\lambda \in [0,1]$ and $\mathcal{Q} = (Q_i)$ be a finite measurable partition. Then*

$$H_{\lambda\nu_1+(1-\lambda)\nu_2}(\mathcal{Q}) \geq \lambda H_{\nu_1}(\mathcal{Q}) + (1-\lambda)H_{\nu_2}(\mathcal{Q}).$$

Proof. This follows from the convexity/concavity of $-x\log(x)$. $\square$

By applying this to $\mu_n = \frac{1}{d_n}\sum_{j=0}^{d_n-1}(\alpha_j)_*\mathrm{m}_{\mathcal{S}_n}$, we get

$$H_{\mu_n}(\mathcal{P}_0^{k-1}) \geq \frac{1}{d_n}\sum H_{(\alpha_j)_*\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_0^{k-1}) = \frac{1}{d_n}\sum_{j=0}^{d_n-1} H_{\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_j^{j+(k-1)}). \tag{24}$$

Let $l_n \in \mathbb{Z}_{\geq 0}$ be defined by

$$l_n k \leq d_n - 1 < (l_n+1)k.$$

By Lemma 8.4,

$$\sum_{j=0,k,...,l_n k} H_{\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_j^{j+(k-1)}) \geq H_{\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_0^{l_n k+k-1}).$$

In general, for every $r = 0, 1, ..., k-1$, let $l_n(r) \in \mathbb{Z}_{\geq 0}$ (so $l_n(0) = l_n$) be defined by

$$l_n(r)k + r \leq d_n - 1 < (l_n(r)+1)k + r.$$

By Lemma 8.4,

$$\begin{aligned} \sum_{j=r,r+,...,r+l_n(r)k} H_{\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_j^{j+(k-1)}) &\geq H_{\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_r^{r+l_n(r)k+k-1}) \\ &\geq H_{\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_0^{r+l_n(r)k+k-1}) - H_{\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_0^{r-1}). \end{aligned} \tag{25}$$

By Lemma 8.8, for every pair $\mathbf{s}_1 \neq \mathbf{s}_2$ in $\mathcal{S}_n$, there exists $0 \leq j \leq d_n - 1$ such that $d(\alpha_j.\Lambda_{\mathbf{s}_1}, \alpha_j.\Lambda_{\mathbf{s}_2}) > \delta_4$. Since $\mathrm{diam}(P) < \delta_4$ for every $P \in \mathcal{P}$, we conclude that $\Lambda_{\mathbf{s}_1}$ and $\Lambda_{\mathbf{s}_2}$ can not lie in the same element of the partition $\alpha_j^{-1}(\mathcal{P})$ and in particular $\mathcal{P}_0^{l_n(r)+r+k-1}$.

So we conclude that for every $r = 0, 1, ..., k-1$,

$$H_{\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_0^{l_n(r)k+r+k-1}) = \log(\#\mathcal{S}_n).$$

Combined with Equa.(24,25), we get

$$H_{\mu_n}(\mathcal{P}_0^{k-1}) \geq \frac{k}{d_n}\log(\#\mathcal{S}_n) - \frac{k}{d_n} H_{\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_0^{k-2}).$$

By the definition of $\mathcal{S}_n$ (as in Sect.8.6), $\log\#\mathcal{S}_n > \delta_1\log(\varepsilon_n^{-1}) - \log(l_0)$. Therefore,

$$H_{\mu_n}(\mathcal{P}_0^{k-1}) > \frac{k}{d_n}\delta_1\log(\varepsilon_n^{-1}) - \frac{k}{d_n}\left(H_{\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_0^{k-2}) + \log(l_0)\right).$$

By the choice of d_n (see Section 8.6), we have for any $\varepsilon > 0$ and n large enough,

$$\frac{\log(\varepsilon_n^{-1})}{d_n} \geq 3 - \frac{3+\log(\delta_3)}{d_n} \geq 3 - \varepsilon.$$

Hence,

$$H_{\mu_n}(\mathcal{P}_0^{k-1}) \geq (3-\varepsilon)k\delta_1 - \frac{k}{d_n}\left(H_{\mathrm{m}_{\mathcal{S}_n}}(\mathcal{P}_0^{k-2}) + \log(l_0)\right).$$

Combined with Equa.(23), we get

$$H_\mu(T,\mathcal{P}) = \lim_{k\to+\infty}\frac{1}{k}H_\mu(\mathcal{P}_0^{k-1}) = \lim_{k\to+\infty}\frac{1}{k}\lim_{n\to+\infty}H_{\mu_n}(\mathcal{P}_0^{k-1}) \geq (3-\varepsilon)\delta_1.$$

Letting $\varepsilon \to 0$ finishes the proof.

8.10. **Conditional entropy.** We need the general notion of conditional entropy for step two. Let $(X, \mathscr{B}_X)$ be a compact metrizable space together with its Borel σ-algebra.

Definition 8.13. *Let $\mathcal{P}$ be a finite measurable partition and $\mathscr{A}$ be a σ-subalgebra. Let $(\mu_x^{\mathscr{A}})_{x\in X'}$ be a family of conditional measures where $X' \in \mathscr{A}$ is a co-null set in X. Note that the map $x \mapsto H_{\mu_x^{\mathscr{A}}}(\mathcal{P})$ is measurable and non-negative. Define the **conditional entropy of $\mathcal{P}$ given $\mathscr{A}$** by*

$$H_\mu(\mathcal{P}|\mathscr{A}) := \int_{X'} H_{\mu_x^{\mathscr{A}}}(\mathcal{P})\,\mu(x).$$

Note that when $\mathscr{A}$ is the σ-subalgebra generated by a finite measurable partition $\mathcal{Q}$, then $H_\mu(\mathcal{P}|\mathscr{A})$ coincides with the $H_\mu(\mathcal{P}|\mathcal{Q})$ defined previously.

Lemma 8.14. *Let $\mathscr{A}_1 \subset \mathscr{A}_2 \subset \dots$ be a sequence of σ-subalgebras and $\mathscr{A}_\infty$ be the smallest σ-subalgebra containing them. Let $\mathcal{P}$ be a finite measurable partition. Then*

$$\lim_{n\to\infty} H_\mu(\mathcal{P}|\mathscr{A}_n) = H_\mu(\mathcal{P}|\mathscr{A}).$$

Proof. By the theorem on conditional measures, for each $P \in \mathcal{P}$,

$$\mu_x^{\mathscr{A}_n}(P) \text{ converges to } \mu_x^{\mathscr{A}_\infty}(P) \text{ for almost every } x.$$

Thus

$$H_{\mu_x^{\mathscr{A}_n}}(\mathcal{P}) \text{ converges to } H_{\mu_x^{\mathscr{A}_\infty}}(\mathcal{P}) \text{ for almost every } x.$$

Also $H_{\mu_x^{\mathscr{A}_n}}(\mathcal{P})$ is bounded by $\log \#\mathcal{P}$. So the conclusion follows from the dominated/bounded convergence theorem. □

A useful observation is that

Lemma 8.15. *Let $\mathscr{A}$ be a countably generated σ-subalgebra and X' be a full measure subset, If for every $x \in \mathscr{C}$, there exists $P \in \mathcal{P}$ with*

$$[x]_{\mathscr{A}} \cap X' \subset P,$$

then $H_\mu(\mathcal{P}|\mathscr{A}) = 0$.

Proof. Indeed, for almost every x, $\mu_x^{\mathscr{A}}(P\cap X') = \mu_x^{\mathscr{A}}(P \cap X' \cap [x]_{\mathscr{A}})$ is equal to 0 or 1. Moreover, there exists a full measure subset such that for every x in this subset,

$$\mu_x^{\mathscr{A}}(X'\cap P) = \mu_x^{\mathscr{A}}(P), \quad \forall\, P \in \mathcal{P}.$$

Hence $H_{\mu_x^{\mathscr{A}}}(\mathcal{P})$ is equal to zero μ-a.e.,which implies the claim. □

If the condition as in last lemma is satisfied, we say that $\mathcal{P}$ is **μ-essentially contained** in $\mathscr{A}$.

8.11. **Step 2, construction of the measure.** The construction is just performing average along A^+. Let

$$\mathbf{a}_{s,t} := \begin{bmatrix} e^s & 0 & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^{-s-t} \end{bmatrix}$$

and

$$\nu_T := \frac{1}{T^2}\int_0^T\int_0^T (\mathbf{a}_{s,t})_*\mu\,\mathrm{dsdt}.$$

which is supported on $\mathscr{C}_1$ by assumption.

Lemma 8.16. *For every $T > 0$, $h_{\nu_T}(\alpha_1) \geq h_\mu(\alpha_1)$.*

Actually $h_{\nu_T}(\alpha_1) = h_\mu(\alpha_1)$, but it requires more work.

Proof. Take a finite partition $\mathcal{P}$ of $\mathscr{C}_1$, we show that $H_{\nu_T}(\mathcal{P}) \geq \frac{1}{T^2}\int_0^T\int_0^T H_{(\mathbf{a}_{s,t})_*\mu}(\mathcal{P})\mathrm{dsdt}$. Since $H_{(\mathbf{a}_{s,t})_*\mu}(\mathcal{P}) = H_\mu(\mathcal{P})$, the lemma follows.

Let $Y := \mathscr{C}_1 \times [0,T]^2$, a compact metrizable space. Define a Borel probability measure ν on Y by

$$\nu(E\times F) := \frac{1}{T^2}\int_0^T\int_0^T (\mathbf{a}_{s,t})_*\mu(E)\mathrm{dsdt}.$$

Define a σ-subalgebra

$$\mathscr{A}_1 := \left\{\mathscr{C}_1 \times F \;\middle|\; F \text{ Borel measurable subset of } [0,T]^2\right\}$$

and a finite partition

$$\mathcal{P}_1 := \left\{P\times[0,T]^2 \;\middle|\; P\in\mathcal{P}\right\}.$$

One can immediately check that if $x\in Y$ is written as $(\Lambda,(s,t))$, then conditional measures $\mu_x^{\mathscr{A}_1}$ can be chosen to be $((\mathbf{a}_{s,t})_*\mu)\otimes\delta_{(s,t)}$. Hence

$$H_\mu(\mathcal{P}_1|\mathscr{A}_1) = \int H_{\nu_x^{\mathscr{A}_1}}(\mathcal{P}_1)\,\nu(x) = \frac{1}{T^2}\int_0^T\int_0^T H_{(\mathbf{a}_{s,t})_*\mu}(\mathcal{P})\mathrm{dsdt}.$$

On the other hand,

$$H_\nu(\mathcal{P}_1) = H_{\nu_T}(\mathcal{P}).$$

So it remains to prove that $H_\mu(\mathcal{P}_1|\mathscr{A}_1)\leq H_\mu(\mathcal{P}_1)$. Choose an increasing sequence of finite σ-subalgebras $\mathscr{B}_i$ converging to $\mathscr{A}_1$. By Lemma 8.14,

$$H_\mu(\mathcal{P}_1|\mathscr{B}_i)\to H_\mu(\mathcal{P}_1|\mathscr{A}_1).$$

But each $H_\mu(\mathcal{P}_1|\mathscr{B}_i)\leq H_\mu(\mathcal{P}_1)$. So we are done.

□

Find a convergent subsequence $\nu := \lim_n \nu_{T_n}$. As before, we can show that

Lemma 8.17. *ν is* A*-invariant.*

What is less trivial is

Lemma 8.18. $h_\nu(\alpha_1)\geq\limsup h_{\nu_{T_n}}(\alpha_1) = h_\mu(\alpha_1)$.

Thus the proof of Theorem 8.6 is complete modulo this lemma.

Note that for each finite measurable partition $\mathcal{P}$, we have

$$h_\nu(\alpha_1,\mathcal{P})\geq\limsup h_{\nu_n}(\alpha_1,\mathcal{P}).$$

whenever ν_n converges to ν (under weak$*$ topology) and boundary of each element in $\mathcal{P}$ has vanishing ν-measure. So the real task is to find a "generating partition" that works for all ν_n.

8.12. **Dynamical entropy as conditional entropy.**

Lemma 8.19. $h_\mu(T,\mathcal{P}) = H_\mu(\mathcal{P}|\mathcal{P}_1^\infty) = H_\mu(\mathcal{P}|\mathcal{P}_{-\infty}^{-1}) = h_\mu(T^{-1},\mathcal{P})$.

Remark 8.20. *If one does not assume knowledge of conditional measures, especially the "martingale convergence theorem", then the proof below shows that $h_\mu(T,\mathcal{P}) = \lim_{n\to\infty}H_\mu(\mathcal{P}|\mathcal{P}_1^n)$. Similar remarks apply to the the lemma below.*

Proof.

$$\begin{aligned}
&H_\mu(\mathcal{P}\vee T^{-1}\mathcal{P}\vee...\vee T^{-(n-1)}\mathcal{P})\\
=&H_\mu(T^{-(n-1)}\mathcal{P}) + H_\mu(T^{-(n-2)\mathcal{P}}|T^{-(n-1)}\mathcal{P}) + ... + H_\mu(\mathcal{P}|\mathcal{P}_1^{n-1})\\
=&H_\mu(\mathcal{P}) + H_\mu(\mathcal{P}|T^{-1}\mathcal{P}) + H_\mu(\mathcal{P}|T^{-1}\mathcal{P}\vee T^{-2}\mathcal{P}) + ... + H_\mu(\mathcal{P}|\mathcal{P}_1^{n-1}).
\end{aligned}$$

As the sequence $(H_\mu(\mathcal{P}|\mathcal{P}_1^{n-1}))$ converges to $H_\mu(\mathcal{P}|\mathcal{P}_1^\infty)$, it converges to the same limit on average. So we are done by invoking the definition of $h_\mu(T,\mathcal{P})$.

The rest of the equalities follow if $h_\mu(T,\mathcal{P}) = h_\mu(T^{-1},\mathcal{P})$, which is true since

$$H_\mu(\mathcal{P}\vee T^{-1}\mathcal{P}\vee...\vee T^{-(n-1)}\mathcal{P}) = H_\mu(T^{n-1}\mathcal{P}\vee T^{n-2}\mathcal{P}\vee...\vee\mathcal{P}).$$

□

More generally, we have

Lemma 8.21. *Let $\mathcal{P}$ and $\mathcal{Q}$ be two finite measurable partitions, then*

$$h_\mu(T,\mathcal{P}\vee\mathcal{Q}) = h_\mu(T,\mathcal{Q}) + H_\mu(\mathcal{P}|\mathcal{P}_1^\infty\vee\mathcal{Q}_{-\infty}^{+\infty}).$$

Proof. The proof is similar. We only present the first two steps since it gets quite complicated in general.

$$\begin{aligned}&H_\mu(\mathcal{P}\vee T^{-1}\mathcal{P}\vee\mathcal{Q}\vee T^{-1}\mathcal{Q})\\ =&H_\mu(T^{-1}\mathcal{Q}) + H_\mu(\mathcal{Q}|T^{-1}\mathcal{Q}) + H_\mu(T^{-1}\mathcal{P}|\mathcal{Q}\vee T^{-1}\mathcal{Q}) + H_\mu(\mathcal{P}|T^{-1}\mathcal{P}\vee\mathcal{Q}\vee T^{-1}\mathcal{Q})\\ =&H_\mu(\mathcal{Q}) + H_\mu(\mathcal{Q}|\mathcal{Q}_1^1) + H_\mu(\mathcal{P}|\mathcal{Q}_{-1}^0) + H_\mu(\mathcal{P}|\mathcal{P}_1^1\vee\mathcal{Q}_0^1)\end{aligned}$$

$$\begin{aligned}&H_\mu(\mathcal{P}\vee T^{-1}\mathcal{P}\vee T^{-2}\mathcal{P}\vee\mathcal{Q}\vee T^{-1}\mathcal{Q}\vee T^{-2}\mathcal{Q})\\ =&H_\mu(\mathcal{Q}) + H_\mu(\mathcal{Q}|\mathcal{Q}_1^1) + H_\mu(\mathcal{Q}|\mathcal{Q}_1^2) + H_\mu(\mathcal{P}|\mathcal{Q}_{-2}^0) + H_\mu(\mathcal{P}|\mathcal{P}_1^1\vee\mathcal{Q}_{-1}^1) + H_\mu(\mathcal{P}|\mathcal{P}_1^2\vee\mathcal{Q}_0^2)\end{aligned}$$

The general formula goes as

$$\begin{aligned}&H_\mu(\mathcal{P}_0^n\vee\mathcal{Q}_0^n)\\ =&\sum_{k=0}^n H_\mu(\mathcal{Q}|\mathcal{Q}_1^k) + \sum_{k=0}^n H_\mu\left(T^{-(n-k)}\mathcal{P}\middle|\bigvee_{i=0}^{k-1} T^{-(n-i)}\mathcal{P}\vee\bigvee_{i=1}^{n} T^{-i}\mathcal{Q}\right)\\ =&\sum_{k=0}^n H_\mu(\mathcal{P}|\mathcal{Q}_1^k) + \sum_{k=0}^n H_\mu\left(\mathcal{P}\middle|\bigvee_{i=0}^{k-1} T^{-(k-i)}\mathcal{P}\vee\bigvee_{i=-(n-k)+1}^{k} T^{-i}\mathcal{Q}\right)\\ =&\sum_{k=0}^n H_\mu(\mathcal{Q}|\mathcal{Q}_1^k) + \sum_{k=0}^n H_\mu(\mathcal{P}|\mathcal{P}_1^k\vee\mathcal{Q}_{-n+k+1}^k)\end{aligned}$$

The quotient by n of the left hand side converges to $H_\mu(T,\mathcal{P}\vee\mathcal{Q})$. And the limit $H_\mu(\mathcal{Q}|\mathcal{Q}_1^k)$ as $k\to+\infty$ converges to $H_\mu(\mathcal{Q}|\mathcal{Q}_1^\infty)$, and the limit of $H_\mu(\mathcal{P}|\mathcal{P}_1^a\vee\mathcal{Q}_{-b}^c)$ as $a,b,c\to+\infty$ converges to $H_\mu(\mathcal{P}|\mathcal{P}_1^\infty\vee\mathcal{Q}_{-\infty}^{+\infty})$. We are done by taking the limit of their averages. □

8.13. **Criterion of being a generating partition.** Combining efforts made in the last two subsections, we have shown:

Corollary 8.22. *Let $\mathcal{Q}$ be a finite measurable partition. If for every finite measurable partition $\mathcal{P}$ satisfying $\mu(\partial P)=0$ for all $P\in\mathcal{P}$, one has $\mathcal{P}$ is μ-essentially contained in $\mathcal{P}_1^\infty\vee\mathcal{Q}_{-\infty}^{+\infty}$, then*

$$h_\mu(T) = h_\mu(T,\mathcal{Q}).$$

Proof. By Lemma 8.9, there exists an increasing sequence of finite measurable partitions $(\mathcal{Q}_n)$ with diameter decreasing to 0 and μ-trivial boundary. Then we claim that $h_\mu(T)=\lim h_\mu(T,\mathcal{Q}_n)$.

Let $\mathcal{Q}_\infty$ be the smallest σ-subalgebra containing all $\mathcal{Q}_n$'s. Then $\mathcal{Q}_\infty$ is countably generated and every atom consists of one single point. For any other finite partition $\mathcal{P}$, one has

$$\inf H_\mu(\mathcal{P}|\mathcal{P}_1^\infty\vee(\mathcal{Q}_n)_{-\infty}^{+\infty}) \le \lim H_\mu(\mathcal{P}|\mathcal{Q}_n) = H_\mu(\mathcal{P}|\mathcal{Q}_\infty) = 0.$$

So the claim is true by Lemma 8.21.

On the other hand, by our assumption and Lemma 8.21, we have

$$h_\mu(T,\mathcal{Q}_n\vee\mathcal{Q}) = h_\mu(T,\mathcal{Q})$$

for every n. So we are done.

□

8.14. **Expansive modulo centralizer.** Let μ be an α_1-invariant probability measure supported on $\mathscr{C}_1$. Fix a finite measurable partition $\mathcal{Q}$ (of $\mathscr{C}_1$) with $\operatorname{diam}(\mathcal{Q})<\delta_?$. Let $\mathcal{P}$ be any other finite measurable partition $\mathcal{P}$ with μ-trivial boundary. The goal is to show that for μ-almost every x, the atom $[x]_{\mathcal{P}_1^\infty\vee\mathcal{Q}_{-\infty}^{+\infty}}$ is contained in one element of $\mathcal{P}$.

Lemma 8.23. *If $y\in[x]_{\mathcal{Q}_{-\infty}^{+\infty}}$, then $y=z_y^x.x$ for some $z_y^x\in\mathrm{Z}(\alpha_1)$ with $\left\|z_y^x\right\|<\delta_?$.*

8.15. **Poincare recurrence.**

Slogan: If something happens once, then it should happen infinitely many times.

Lemma 8.24 (Poincare recurrence)**.** *For every measurable set E, there exists a subset $E' \subset E$ with $\mu(E \setminus E') = 0$ such that for every $x \in E'$, there exists infinitely many $n \in \mathbb{Z}^+$ with $\alpha_1^n.x$ belongs to E.*

Proof. For every $N \in \mathbb{Z}^+$, let

$$E_N := \{x \mid \alpha_1^n.x \in E, \, \exists n > N\}$$

It suffices to show that $\mu(E \setminus E_N) = 0$ for every N, since then we can take $E' := \cap E_N$.

Write

$$F_N := \bigcup_{i=N+1}^{\infty} \alpha_1^{-i}(E)$$

then it suffices to show that $\mu(E \setminus F_N) = 0$.

Note that

$$F_N \subset \alpha_1(F_N) \subset \alpha_1^2(F_N) \subset ...$$

On the other hand, μ is preserved by α_1. Thus, for each k, $\alpha_1^k(F_N) \setminus F_N$ has zero μ-measure. Applied when $k = N+1$, we see that $\mu(E \setminus F_N) = 0$. $\square$

8.16. **Idea of the proof.** For every $z \in Z(\alpha_1)$ with $\|z\| < \delta_?$, let

$$S_z := \{x \in X \mid x, z.x \text{ lie in the same } P \in \mathcal{P}\} = \bigcup_{P \in \mathcal{P}} (P \cap z^{-1}P)$$

$$D_z := \{x \in X \mid x, z.x \text{ lie in different } P \in \mathcal{P}\} = \bigcup_{P \in \mathcal{P}} (P \setminus z^{-1}P)$$

and the associated recurrence sets

$$\mathcal{R}S_z := \{x \in S_z, \ \alpha_1^n.x \in S_z \text{ for infinitely many } n \in \mathbb{Z}^+\},$$
$$\mathcal{R}D_z := \{x \in D_z, \ \alpha_1^n.x \in D_z \text{ for infinitely many } n \in \mathbb{Z}^+\},$$
$$\mathcal{R}_z := \mathcal{R}S_z \sqcup \mathcal{R}D_z$$

By Poincare recurrence, $\mu(\mathcal{R}_z) = 1$.

Now for $x \in \bigcap_z \mathcal{R}_z$, we can show that $y \in [x]_{\mathcal{P}_1^\infty \vee \mathcal{Q}_{-\infty}^{+\infty}}$ implies that $y \in [x]_{\sigma(\mathcal{P})}$.

Indeed, by Lemma 8.23, $y \in [x]_{\mathcal{Q}_{-\infty}^{+\infty}}$ implies that $y = z.x$ for some z in the centralizer with $\|z\| < \delta_?$. On the other hand, $y \in [x]_{\mathcal{P}_1^\infty}$ together with the fact that z commutes with α_1 imply that $\alpha_1^n.x \in S_z$ for all $n \in \mathbb{Z}^+$. Hence x lies in $\mathcal{R}S_z$, in particular, $x \in S_z$ or $y \in [x]_{\sigma(\mathcal{P})}$.

Unfortunately, $\bigcap_z \mathcal{R}_z$, being an uncountable intersection, may not have full measure (even not clear if it is measurable). So a refined argument is needed and will be given in the next subsection.

8.17. **The proof.** For $n \in \mathbb{Z}^+$ and $P \in \mathcal{P}$, let

$$P(\frac{1}{n}) := \left\{ x \;\middle|\; d(x, P^c) > \frac{1}{n} \right\}$$

$$S_{z,\frac{1}{n}} := \left\{ x \in X \;\middle|\; x, z.x \text{ lie in the same } P(\frac{1}{n}), \exists P \in \mathcal{P} \right\}$$

$$D_{z,\frac{1}{n}} := \left\{ x \in X \;\middle|\; x, z.x \text{ lie in different } P(\frac{1}{n}), \exists P \in \mathcal{P} \right\}$$

Let $\mathcal{R}S_{z,\frac{1}{n}}$ and $\mathcal{R}D_{z,\frac{1}{n}}$ denote the corresponding recurrence sets and

$$\mathcal{R}_{z,\frac{1}{n}} := \mathcal{R}S_{z,\frac{1}{n}} \sqcup \mathcal{R}D_{z,\frac{1}{n}}.$$

Then by Poincare recurrence, $\mathcal{R}_{z,\frac{1}{n}}$ is of full measure in $X(\frac{1}{n}) := \bigsqcup_{P \in \mathcal{P}} P(\frac{1}{n})$.

Fix a countable dense subset

$$\mathrm{CZ} \subset \{z \in Z \mid \|z\| < \delta_?\}.$$

Define

$$X' := \bigcup_n \bigcap_{z \in CZ} \mathcal{R}_{z,\frac{1}{n}} \subset \bigsqcup_{P \in \mathcal{P}} \mathrm{Int}(P)$$

Then X' is of full measure in X.

Now let $x, y \in X'$ and $y \in [x]_{\mathcal{P}_1^\infty \vee \mathcal{Q}_{-\infty}^{+\infty}}$, we claim $y \in [x]_{\sigma(\mathcal{P})}$. Assume otherwise and we shall derive a contradiction.

Find $n = n_{x,y} \in \mathbb{Z}^+$ such that

$$y \in X(\frac{1}{n}), \quad x \in \bigcap_{z \in CZ} \mathcal{R}_{z, \frac{1}{n}}.$$

So we find $z_y \in Z$ with $\|z_y\| < \delta_?$ such that $y = z_y.x$. Choose $z \in CZ$ sufficiently close to z_y such that

$$zz_y^{-1}.P(\frac{1}{n}) \subset P(\frac{1}{n+1}), \quad \forall P \in \mathcal{P};$$
$$z_y z^{-1}.P(\frac{1}{n+1}) \subset P(\frac{1}{n+2}), \quad \forall P \in \mathcal{P}.$$

Thus x and $z.x$ lies in different $P(\frac{1}{n+1})$'s. In particular, $x \in \mathcal{R}D_{z, \frac{1}{n+1}}$. So there exists $l \in \mathbb{Z}^+$ such that

$$z\alpha_1^l.x \in P(\frac{1}{n+1}), \quad \alpha_1^l.x \in P'(\frac{1}{n+1}), \quad \exists\, P \neq P' \in \mathcal{P}$$

By assumption on z,

$$\alpha_1^l.y = z_y \alpha_1^l.x \in P(\frac{1}{n+2}).$$

This contradicts against the fact that $\alpha_1^l.x$ and $\alpha_1^l.y$ are supposed to be contained in the same $P \in \mathcal{P}$.

8.18. **Exercises.** Recall $X := \{0,1\}^{\mathbb{Z}_{\geq 0}}$ from Exercise A Define a continuous map $\sigma : X \to X$ by $\sigma(x)_i := x_{i+1}$.

Exercise A. *Let X be equipped with the metric defined above,* $\dim_H(X) = 1$.

Exercise B. *Let X be equipped with the metric defined above,* $\dim_\square(X) = 1$.

For $\lambda \in [0,1]$, there is a Borel probability measure m_λ on X such that

$$\mathrm{m}_\lambda(C_{\mathbf{a}}) := \lambda^{l_0}(1-\lambda)^{l-l_0}.$$

for every word $\mathbf{a}$ of length l. Here $l_0 := \#\{i \in \{0,1,...,l-1\} \mid \mathbf{a}_i = 0\}$. Accept the fact that m_λ is σ-invariant for every $\lambda \in [0,1]$. Let $\mathcal{P}_0$ be the partition $\{C_{(0)}, C_{(1)}\}$.

Exercise C. *For every σ-invariant Borel probability measure ν on X, show $h_\nu(\sigma) = h_\nu(\sigma, \mathcal{P}_0)$.*

Exercise D. *Show that $h_{\mathrm{m}_\lambda}(\sigma) \leq \log 2$. And equality holds iff $\lambda = 1/2$.*

For $l \geq 0$, let $\mathcal{P}_0^{l-1} := \mathcal{P}_0 \vee \sigma^{-1}\mathcal{P}_0 \vee ... \vee \sigma^{-(l-1)}\mathcal{P}_0$. For $x_\star \in X$, let $\mathcal{P}_0^{l-1}(x_\star)$ be the unique element of $\mathcal{P}_0^{l-1}$ containing $x_\star$.

Exercise E. *Fix $\lambda \in [0,1]$. Show that for m_λ-almost every $x_\star \in X$, we have*

$$\lim_{l \to \infty} \frac{-\log \mu(\mathcal{P}_0^{l-1}(x_\star))}{n} = h_{\mathrm{m}_\lambda}(\sigma).$$

Exercise F. *Take $E \subset X$ such that $\dim_\square(E) > 0$. Let E' be the closure of the union of $\sigma^n(E)$ as n varies in $\mathbb{Z}_{\geq 0}$. Show that there exists a σ-invariant Borel probability measure ν supported on E' such that $h_\nu(\sigma) > 0$.*

In the reverse direction, we have,

Exercise G. *Let ν be a σ-invariant probability measure on X with $h_\nu(\sigma) > 0$, show that* $\dim_{\mathrm{H}}(\mathrm{supp}(\nu)) > 0$.

Recall that $x \in \mathrm{supp}(\nu)$ iff every open neighborhood of x has positive ν-measure.

For $x_\star \in X$ and $n \in \mathbb{Z}^+$, define

$$\mathbf{w}_n(x_\star) := \#\{f : \{0,1,...,n-1\} \to \{0,1\} \mid f(i) = x_{n+i},\ \exists\, n \in \mathbb{Z}_{\geq 0}\}$$

and its "word entropy" as

$$\mathrm{WH}(x_\star) := \limsup_{n \to \infty} \frac{\log \mathbf{w}_n(x_\star)}{n}.$$

Exercise H. *Let $x_\star \in X$ and E' be the closure of $\{\sigma^n(x_\star),\ n \in \mathbb{Z}_{\geq 0}\}$. Prove that if* $\mathrm{WH}(x_\star) > 0$, *then there exists a σ-invariant Borel probability measure ν supported on E' such that $h_\nu(\sigma) > 0$.*

8.19. **Definition of topological entropy.** Let (X, d) be a compact metric space and $f : X \to X$ be a continuous map. We let

$$\mathrm{Prob}(X, f) := \{f\text{-invariant Borel probability measures on } X\}.$$

Recall for every $\mu \in \mathrm{Prob}(X, f)$, we have defined the measure entropy:

$$h_\mu(f) := \sup_{\mathcal{P}} h_\mu(f, \mathcal{P}) = \lim_{n\to+\infty} \frac{1}{n} H_\mu(\mathcal{P} \vee f^{-1}\mathcal{P} \vee ... \vee f^{-(n-1)}\mathcal{P}).$$

Topological entropy is nothing but

$$h_{\mathrm{top}}(f) := \sup_{\mu\in\mathrm{Prob}(X,f)} h_\mu(f).$$

If $\mu \in \mathrm{Prob}(X, f)$ is such that $h_{\mathrm{top}}(f) = h_\mu(f)$, then we say that μ is a **measure of maximal entropy**.

Below we give examples on "what entropy is" without proof.

8.20. **Genus zero: Rational functions on the Riemann sphere.** Let $P(z), Q(z)$ be two relatively prime polynomials in $\mathbb{C}[X]$ of degree p, q respectively. Let $f(z) = \frac{P(z)}{Q(z)}$. We regard $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ as an endomorphism of the Riemann sphere $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$.

Theorem 8.25. *Let f be as above. Then $h_{\mathrm{top}}(f) = \log(\deg(f)) = \log(\max\{p, q\})$. If nonzero, there exists a unique f-invariant probability measure μ (supported on the Julia set) with $h_{\mathrm{top}}(f) = h_\mu(f)$.*

8.21. **Genus one: Endomorphism of tori.** Let M be a 2-by-2 matrix with entries in $\mathbb{Z}$ acting naturally on $\mathbb{R}^2$. Since M preserves the lattice $\mathbb{Z}^2$, it induces a continuous map $f_M : \mathbb{R}^2/\mathbb{Z}^2 \to \mathbb{R}^2/\mathbb{Z}^2$ on the 2-dimensional torus (one can replace 2 by n to get the same conclusion).

Theorem 8.26. *Let $(\lambda_i)_{i=1}^2$ be the set of eigenvalues of M, counted with multiplicities. Then*

$$h_{\mathrm{top}}(f_M) = \sum_{|\lambda_i|\geq 1} \log(|\lambda_i|).$$

Theorem 8.27. *Assume further that M is a hyperbolic matrix (namely, every eigenvalue has absolute value different from 1) in $\mathbf{GL}_2(\mathbb{Z})$, then the standard Lebesgue measure m is the unique measure such that*

$$h_{\mathrm{top}}(f_M) = h_{\mathrm{m}}(f_M).$$

8.22. **Higher genus: Geodesic flow.** For a Riemannian manifold $(\mathcal{M}, d)$, let $T^1\mathcal{M}$ be its unit-tangent bundle (i.e., sub-bundle of the tangent bundle consisting of unit vectors). Let $f_t : T^1\mathcal{M} \to T^1\mathcal{M}$ be the geodesic flow. Let $\widetilde{\mathcal{M}}$ be the universal cover of $\mathcal{M}$ equipped with the induced Riemannian metric. Let Vol be the induced volume. For a point $x \in \widetilde{\mathcal{M}}$, let $B_x(R) := \{y \in \widetilde{\mathcal{M}},\ d(y, x) < R\}$. Let $\mathcal{L}(R)$ (resp., $L(R)$) denote the set (resp., number) of closed prime geodesics of length smaller than R.

Theorem 8.28. *Assume the Riemannian manifold $(\mathcal{M}, d)$ is a closed surface (of genus at least two) and has negative sectional curvatures, then for every $x \in \widetilde{\mathcal{M}}$,*

$$h_{\mathrm{top}}(f_1) = \lim_{R\to+\infty} \frac{1}{R} \log \mathrm{Vol}(B_x(R)) = \lim_{R\to+\infty} \frac{1}{R} \log L(R).$$

Theorem 8.29. *Keep assumptions as above. There exists a unique probability measure μ invariant under the geodesic flow such that $h_\mu(f_1) = h_{\mathrm{top}}(f_1)$. Also, if for a closed geodesic γ, let m_γ denote the natural probability measure on γ. Then*

$$\mu = \lim_{n\to\infty} \frac{1}{L(R)} \sum_{\gamma\in\mathcal{L}(R)} \mathrm{m}_\gamma.$$

Theorem 8.30. *Keep assumptions as above. Normalize the Riemannian metric such that the volume of $T^1\mathcal{M}$ is one. Then*

$$h_{\mathrm{top}}(f_1) \geq \sqrt{2g-2} = \sqrt{-\chi(\mathcal{M})}.$$

The equality holds iff the curvature is constant. Also, the (normalized probability) Liouville measure is the measure of maximal entropy iff the curvature is constant.

References

[Bad13] Dzmitry A. Badziahin, *On multiplicatively badly approximable numbers*, Mathematika **59** (2013), no. 1, 31–55. MR 3028170 5

[BRV16] Victor Beresnevich, Felipe Ramírez, and Sanju Velani, *Metric Diophantine approximation: aspects of recent work*, Dynamics and analytic number theory, London Math. Soc. Lecture Note Ser., vol. 437, Cambridge Univ. Press, Cambridge, 2016, pp. 1–95. MR 3618787 3

[Cas50] J. W. S. Cassels, *Some metrical theorems in Diophantine approximation. I*, Proc. Cambridge Philos. Soc. **46** (1950), 209–218. MR 36787 3

[Cas59] ———, *An introduction to the geometry of numbers*, Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete, vol. Bd. 99, Springer-Verlag, Berlin-Göttingen-Heidelberg, 1959. MR 157947 12

[CSD55] J. W. S. Cassels and H. P. F. Swinnerton-Dyer, *On the product of three homogeneous linear forms and the indefinite ternary quadratic forms*, Philos. Trans. Roy. Soc. London Ser. A **248** (1955), 73–96. MR 70653 33

[Ein10] Manfred Einsiedler, *Applications of measure rigidity of diagonal actions*, Proceedings of the International Congress of Mathematicians. Volume III, Hindustan Book Agency, New Delhi, 2010, pp. 1740–1759. MR 2827864 35

[EKL06] Manfred Einsiedler, Anatole Katok, and Elon Lindenstrauss, *Invariant measures and the set of exceptions to Littlewood's conjecture*, Ann. of Math. (2) **164** (2006), no. 2, 513–560. MR 2247967 4, 56

[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 12, 52

[Fur67] Harry Furstenberg, *Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation*, Math. Systems Theory **1** (1967), 1–49. MR 213508 35

[Kle23] Dmitry Kleinbock, *Ergodic theory on homogeneous spaces and metric number theory*, Ergodic theory, Encycl. Complex. Syst. Sci., Springer, New York, [2023] ©2023, pp. 597–613. MR 4647091 12

[KM96] D. Y. Kleinbock and G. A. Margulis, *Bounded orbits of nonquasiunipotent flows on homogeneous spaces*, Sinaĭ's Moscow Seminar on Dynamical Systems, Amer. Math. Soc. Transl. Ser. 2, vol. 171, Amer. Math. Soc., Providence, RI, 1996, pp. 141–172. MR 1359098 26

[KM99] ———, *Logarithm laws for flows on homogeneous spaces*, Invent. Math. **138** (1999), no. 3, 451–494. MR 1719827 12

[Lin22] Elon Lindenstrauss, *Recent progress on rigidity properties of higher rank diagonalizable actions and applications*, Dynamics, geometry, number theory—the impact of Margulis on modern mathematics, Univ. Chicago Press, Chicago, IL, [2022] ©2022, pp. 362–425. MR 4422060 35

[LW01] Elon Lindenstrauss and Barak Weiss, *On sets invariant under the action of the diagonal group*, Ergodic Theory Dynam. Systems **21** (2001), no. 5, 1481–1500. MR 1855843 33

[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 33

[Mat95] Pertti Mattila, *Geometry of sets and measures in Euclidean spaces*, Cambridge Studies in Advanced Mathematics, vol. 44, Cambridge University Press, Cambridge, 1995, Fractals and rectifiability. MR 1333890 26, 56

[McM87] Curt McMullen, *Area and Hausdorff dimension of Julia sets of entire functions*, Trans. Amer. Math. Soc. **300** (1987), no. 1, 329–342. MR 871679 26

[Mor15] Dave Witte Morris, *Introduction to arithmetic groups*, Deductive Press, [place of publication not identified], 2015. MR 3307755 12

[Phe01] Robert R. Phelps, *Lectures on Choquet's theorem*, second ed., Lecture Notes in Mathematics, vol. 1757, Springer-Verlag, Berlin, 2001. MR 1835574 38

[Rud87] Walter Rudin, *Real and complex analysis*, third ed., McGraw-Hill Book Co., New York, 1987. MR 924157 19, 54

[Rud90] Daniel J. Rudolph, $\times 2$ *and* $\times 3$ *invariant measures and entropy*, Ergodic Theory Dynam. Systems **10** (1990), no. 2, 395–406. MR 1062766 35

[SS05] Elias M. Stein and Rami Shakarchi, *Real analysis*, Princeton Lectures in Analysis, vol. 3, Princeton University Press, Princeton, NJ, 2005, Measure theory, integration, and Hilbert spaces. MR 2129625 6

[Zaf17] Agamemnon Zafeiropoulos, *Inhomogeneous diophantine approximation, m_0 sets and projections of fractals.*, Ph.D. thesis, University of York, 2017. 3