

# A HOPF'S RATIO ERGODIC THEOREM FOR HEISENBERG GROUP

QINGYUAN CHEN

ABSTRACT. We present several classical theorems in ergodic theory and two examples of dynamical systems, geodesic flows on quotients of the hyperbolic plane and non-singular actions of the Heisenberg group. For the second example, we generalize recent works on ratio ergodic theorem for discrete Heisenberg group actions to continuous actions of real Heisenberg group. Besides, we will also summarize Hochman's result on the equivalence of a ratio maximal inequality and Besicovitch covering property of spaces.

## CONTENTS

1. Ergodicity and Ergodic Theorems	3
2. Amendable Groups and Ergodicity of Actions	13
3. Example: Geodesic Flows on quotients of Hyperbolic Plane	19
4. Maximal Inequality and Besicovitch Covering Property	25
5. Example: Non-singular Actions of Heisenberg Group	30
References	35

### Acknowledgement

I would like to express my special thanks of gratitude to my program advisor Amir Mohammadi, who offered me this opportunity to do this honor project, guided and supported me in completing this project.

1. ERGODICITY AND ERGODIC THEOREMS

In this section, we define ergodicity for dynamics systems on measure spaces and discuss a few classical ergodic theorems. In a vague sense, we say a system is ergodic if it does not have nontrivial independent sub-systems. A rigorous definition is as follows.

**Definition 1.1.** Let  $(X, \mathcal{B}, \mu)$  and  $(Y, \mathcal{C}, \nu)$  be probability spaces. A measurable function  $T : X \rightarrow Y$  is measure-preserving if  $\mu(T^{-1}B) = \nu(B)$  for all  $B \in \mathcal{C}$ .

We use the pre-images instead of images to define measure-preserving maps so that the examples such as the circle doubling map are included. The following theorem, Poincaré recurrence theorem, shows that almost all points in such dynamical systems return to a neighbourhood of themselves infinitely often.

**Theorem 1.2. (Poincaré Recurrence).** *Let  $(X, \mathcal{B}, \mu)$  be a probability space,  $T : X \rightarrow X$  be a measure-preserving map on  $X$ , and  $E \in \mathcal{B}$  be a measurable set. Then, there exists a measurable set  $F \in \mathcal{B}$  with  $\mu(F) = \mu(E)$ , so that for all  $x \in F$ , there exist an increasing sequence  $\{n_i\}_{i=1}^\infty$  of positive integers such that  $T^{n_i}x \in E$  for all  $i$ .*

*Proof.* Let  $B = \{x \in E | \forall n, T^n x \notin E\}$ . Then  $B$  is measurable since

$$B = E \cap \bigcap_{n \in \mathbb{N}} T^{-n}(X \setminus E).$$

Then for  $m \in \mathbb{N}$ ,

$$T^{-m}B = T^{-m}E \cap \bigcap_{n \in \mathbb{N}} T^{-n-m}(X \setminus E),$$

so  $\{T^{-n}B\}_{n \in \mathbb{N}}$  and  $B$  are mutually disjoint sets with same measure since  $T$  is measure-preserving. Since  $X$  is a probability space,  $\mu(B) = 0$ . Let  $F_1 = E \setminus B$ , then  $\mu(F_1) = \mu(E)$ , and every point in  $F_1$  returns to  $E$  at least once. Apply the same argument to all  $T^n$ , and for each  $n \in \mathbb{N}$ , there exists  $F_n \in \mathcal{B}$  so that  $\mu(F_n) = \mu(E)$  and every point in  $F_n$  returns to  $E$  at the  $k$ -th step where  $k$  is a multiple of  $n$ . Let

$F = \bigcap_{n \in \mathbb{N}} F^n$ . Then  $\mu(F) = \mu(E)$ , and every point in  $F$  returns to  $R$  infinitely often.  $\square$

A system is said to be ergodic if the invariant measurable subsets are of measure zero or have measure zero complement. An ergodic system can be considered as an indecomposable system. The following proposition presents several equivalent conditions for ergodicity.

**Definition 1.3.** Let  $(X, \mathcal{B}, \mu)$  be a probability space,  $T : X \rightarrow X$  be a measure-preserving map on  $X$ . Then  $T$  is ergodic if for any  $B \in \mathcal{B}$ ,

$$T^{-1}B = B \implies \mu(B) = 0 \text{ or } \mu(B) = 1.$$

**Proposition 1.4.** Let  $(X, \mathcal{B}, \mu)$  be a measure space and  $T$  be a measure preserving transformation on  $X$ . Then the following are equivalent.

- (1)  $T$  is ergodic.
- (2) For any  $B \in \mathcal{B}$ ,  $\mu(T^{-1}B \Delta B) = 0$  implies that  $\mu(B) = 0$  or  $\mu(B) = 1$ .
- (3) For  $A \in \mathcal{B}$ ,  $\mu(A) > 0$  implies that  $\mu(\bigcup_{n=1}^{\infty} T^{-n}A) = 1$ .
- (4) For  $A \in \mathcal{B}$ ,  $B \in \mathcal{B}$ ,  $\mu(A)\mu(B) > 0$  implies that there exists  $n \geq 1$  with  $\mu(T^{-n}A \cap B) > 0$ .
- (5) For  $f : X \rightarrow \mathbb{C}$  measurable function,  $f \circ T = f$  almost everywhere implies that  $f$  is constant almost everywhere.

*Proof.* (1)  $\implies$  (2): Let  $B \in \mathcal{B}$ , with  $\mu(T^{-1}B \Delta B) = 0$ . Let  $C = \bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} T^{-n}B$ . Then for any  $N \geq 0$ ,

$$B \Delta \bigcup_{n=N}^{\infty} T^{-n}B \subseteq \bigcup_{n=N}^{\infty} B \Delta T^{-n}B,$$

and for all  $n \geq 1$ ,

$$B \Delta T^{-n}B \subseteq \bigcup_{i=0}^{n-1} T^{-i}B \Delta T^{-i-1}B.$$

Thus,  $\mu(B \Delta T^{-n}B) = 0$  for all  $n \geq 1$ . Let  $C_N = \bigcup_{n=N}^{\infty} T^{-n}B$ , then  $C_N$  are nested and  $\mu(C_N \Delta B) = 0$  for all  $N$ . Thus,  $\mu(C \Delta B) = 0$ , so

$\mu(B) = \mu(C)$ . Then,

$$T^{-1}C = \bigcap_{N=0}^{\infty} \bigcup_{n=N}^{\infty} T^{-(n+1)}B = \bigcap_{N=0}^{\infty} \bigcup_{n=N+1}^{\infty} T^{-n}B = C.$$

By ergodicity,  $\mu(C) = 0$  or  $1$ , so  $\mu(B) = 0$  or  $1$ .

(2)  $\implies$  (3): Let  $A \in \mathcal{B}$  with  $\mu(A) > 0$ , and set  $B = \bigcup_{n=1}^{\infty} T^{-n}A$ . Then,  $\mu(T^{-1}B \Delta B) = 0$ , since  $\mu(T^{-1}B) = \mu(B)$ . Since  $T^{-1}A \subseteq B$ ,  $\mu(B) \neq 0$ . Then by (2),  $\mu(B) = 1$ .

(3)  $\implies$  (4): Let  $A, B \in \mathcal{B}$  with  $\mu(A)\mu(B) > 0$ . By (3),

$$\mu\left(\bigcup_{n=1}^{\infty} T^{-n}A\right) = 1.$$

Then,

$$\sum_{n=1}^{\infty} \mu(B \cap T^{-n}A) \geq \mu\left(\bigcup_{n=1}^{\infty} B \cap T^{-n}A\right) = \mu(B) > 0$$

Thus, there exists  $n \geq 1$  with  $\mu(B \cap T^{-n}A) > 0$ .

(4)  $\implies$  (1): Let  $A \in \mathcal{B}$  such that  $T^{-1}A = A$ . Then for all  $n \geq 1$ ,

$$\mu(T^{-n}A \cap X \setminus A) = \mu(A \cap X \setminus A) = 0.$$

By (4),  $\mu(A) = 0$  or  $\mu(X \setminus A) = 0$ .

(5)  $\implies$  (2): Let  $B \in \mathcal{B}$  so that  $\mu(T^{-1}B \Delta B) = 0$ . Then  $f = \chi_B$  is a  $\mathbb{T}$ -invariant function, so  $f$  is constant almost everywhere. Thus,  $f = 1$  almost everywhere or  $f = 0$  almost everywhere, which implies that  $\mu(B) = 0$  or  $1$ .

(2)  $\implies$  (5): Let  $f : X \rightarrow \mathbb{C}$  be a measurable function so that  $f \circ T = f$  almost everywhere. Then  $f = \operatorname{Re}(f) + i\operatorname{Im}(f)$ , and  $\operatorname{Re}(f)$  and  $\operatorname{Im}(f)$  are real valued functions. Let  $u = \operatorname{Re}(f)$  and  $v = \operatorname{Im}(f)$ . Then  $(u + iv) \circ T = u \circ T + iv \circ T = f = u + iv$ . Thus,  $u \circ T = u$ , and  $v \circ T = v$ . For all  $k \in \mathbb{Z}$  and  $n \in \mathbb{N}$ , define  $A_n^k = u^{-1}\left(\left[\frac{k}{n}, \frac{k+1}{n}\right)\right)$ . Then for all  $k$  and  $n$ ,

$$\mu(T^{-1}A_n^k \Delta A_n^k) \leq \mu(\{x \in X \mid u(T(x)) \neq u(x)\}) = 0.$$

Since (2) holds, either  $\mu(A_n^k) = 0$  or  $\mu(A_n^k) = 1$ . For each  $n$ ,  $\{A_n^k\}$  are all disjoint and  $X = \bigcup_k A_n^k$ . Therefore, there exists a unique  $k_n$  so that  $\mu(A_n^{k_n}) = 1$ . For  $x, y \in A_n^{k_n}$ ,

$$|u(x) - u(y)| \leq \left| \frac{k_n}{n} - \frac{k_n + 1}{n} \right| = \frac{1}{n}.$$

Let  $Y = \bigcap_n A_n^{k_n}$ , then  $\mu(Y) = 1$ , and for  $x, y \in Y$  and

$$|u(x) - u(y)| \leq \frac{1}{n} \quad \forall n \in \mathbb{N}.$$

Hence,  $u$  is constant on  $Y$ . Similarly,  $v$  is also constant almost everywhere, so  $f$  is constant almost everywhere.  $\square$

For a probability space  $(X, \mathcal{B}, \mu)$ , a measure-preserving map  $T$  on it induces linear operator  $U_T : L_\mu^2 \rightarrow L_\mu^2$ , which is defined by  $U_T(f) = f \circ T$  for all  $f \in L_\mu^2$ . Note that  $L_\mu^2$  is a Hilbert space and for all  $f, g \in L_\mu^2$ ,  $\langle U_T f, U_T g \rangle = \langle f, g \rangle$ . Thus, such  $U_T$  is an isometry, and if in addition  $T$  is invertible, then  $U_T$  is a unitary operator.

The ergodic theorems are generally presenting the relationship between the space average and the time average, where the time average represents the average taken along orbits of iterating the map  $T$ , and the space average represents the average taken over the whole space with respect to a  $T$ -invariant measure.

**Theorem 1.5. (Mean Ergodic Theorem).** *Let  $(X, \mathcal{B}, \mu, T)$  be a measure-preserving system and  $I = \{g \in L_\mu^2 | U_T g = g\}$ . Then  $I$  is a closed subspace of  $L_\mu^2$ . Let  $P_T$  be the orthogonal projection onto  $I$ . Then for any  $f \in L_\mu^2$ ,*

$$\frac{1}{N} \sum_{n=0}^{N-1} U_T^n f \rightarrow P_T f \text{ in } L_\mu^2.$$

We may define  $A_N = \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f$  as the  $N$ th ergodic average of  $f$ . If  $T$  is ergodic, then  $I$  is the subspace of functions that are constant almost everywhere. Then the limit function is an almost everywhere constant function that equals to the integral of  $f$  a.e., which can be realized as the space average.

*Proof.* Let  $\{g_i\}$  be a convergent sequence in  $I \subseteq L_\mu^2$ , then  $U_T g_i = g_i$  for all  $i$ , and let  $g = \lim_{i \rightarrow \infty} g_i$ . Then,  $\|U_T g - g\|_2 \leq \|U_T g - U_T g_i\|_2 + \|g - g_i\|_2 \leq 2\|g - g_i\|_2$  for all  $i$ , since  $U_T$  is an isometry. Thus,  $U_T g = g$ , so  $g \in I$ . Therefore,  $I$  is closed, and  $P_T$  is well defined. Let  $B = \{U_T g - g | g \in L_\mu^2\}$ . Then for  $f \in I$  and  $g \in B$ ,

$$\begin{aligned} \langle f, U_T g - g \rangle &= \langle f, U_T g \rangle - \langle f, g \rangle \\ &= \langle U_T f, U_T g \rangle - \langle f, g \rangle \\ &= \langle f, g \rangle - \langle f, g \rangle \\ &= 0. \end{aligned}$$

Thus,  $I \subseteq B^\perp$ . For  $f \in B^\perp$ ,

$$\langle U_T f, f \rangle = \langle U_T f - f, f \rangle + \|f\|_2^2 = \|f\|_2^2,$$

and

$$\langle f, U_T f \rangle = \langle f, U_T f - f \rangle + \|f\|_2^2 = \|f\|_2^2.$$

Then,

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

Hence,  $I = B^\perp$  and  $L_\mu^2 = I \oplus \bar{B}$ . Let  $f \in L_\mu^2$ , then  $f = P_T f + h$  for some  $h \in \bar{B}$ . We claim that  $\lim_{n \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \right\|_2 = 0$ . If  $h \in B$ , then  $h = U_T g - g$  for some  $g \in L_\mu^2$ . Then, as  $N \rightarrow \infty$

$$\left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \right\|_2 = \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n (U_T g - g) \right\|_2 = \frac{1}{N} \|U_T^N g - g\|_2 \rightarrow 0.$$

For  $h \in \bar{B}$ , there is a sequence  $\{h_i\}_{i=1}^\infty$  of functions in  $B$  converging to  $h$  in  $L_\mu^2$ . Then for all  $i$ , there is  $g_i \in L_\mu^2$  so that  $h_i = U_T g_i - g_i$ . Then for  $\epsilon > 0$ , there exist  $N \in \mathbb{N}$  so that  $\forall i > N, \|h - h_i\|_2 < \epsilon/2$ . Then for

each  $i > N$ , there exist  $M \in \mathbb{N}$  so that  $\forall m > M$ ,

$$\begin{aligned} \left\| \frac{1}{m} \sum_{n=0}^{m-1} U_T^n h \right\|_2 &\leq \left\| \frac{1}{m} \sum_{n=0}^{m-1} U_T^n (U_T h - h_i) \right\|_2 + \left\| \frac{1}{m} \sum_{n=0}^{m-1} U_T^n h_i \right\|_2 \\ &\leq \frac{1}{m} \frac{m\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

Hence,  $\lim_{n \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \right\|_2 = 0$ , and

$$\lim_{n \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n f - P_T f \right\|_2 = \lim_{n \rightarrow \infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} U_T^n h \right\|_2 = 0.$$

□

The following proposition is maximal inequality for positive operators. We will then use it to prove the maximal ergodic theorem, and we will also show in section 4 that a similar results holds for general metric spaces where the Besicovitch covering property holds.

**Proposition 1.6. (Maximal Inequality).** *Let  $(X, \mathcal{B}, \mu)$  be a probability space and  $U : L_\mu^1 \rightarrow L_\mu^1$  be a positive linear operator with  $\|U\| \leq 1$ , and  $f \in L_\mu^1$ . Define  $f_0 = 0$ , for  $n \geq 1$ ,  $f_n = \sum_{i=0}^n U^i f$ , and  $F_n = \max_{0 \leq i \leq n} \{f_i\}$ . Then for all  $N \geq 1$ ,*

$$\int \chi_{\{x | F_N(x) > 0\}} f d\mu \geq 0.$$

*Proof.* Since  $F_n \geq f_n$  for all  $0 \leq n \leq N$ , and  $U$  is a positive linear operator,  $UF_N \geq Uf_n$ . Then for all  $0 \leq n \leq N$ ,

$$UF_N + f \geq Uf_n + f = f_{n+1}.$$

Thus,

$$UF_N + f \geq \max_{1 \leq n \leq N} f_n.$$



Let  $S = \{x | F_N(x) > 0\}$ , then on  $S$ ,  $\max_{1 \leq n \leq N} f_n = \max_{0 \leq n \leq N} f_n = F_N$ , and therefore  $UF_N + f \geq F_N$  on  $S$ . Then,

$$\begin{aligned} \int_S f d\mu &\geq \int_S F_N - UF_N d\mu \\ &\geq \int_X F_N d\mu - \int_X UF_N d\mu \\ &= \|F_N\|_1 - \|UF_N\|_1 \\ &\geq 0. \end{aligned}$$

□

**Theorem 1.7. (Maximal Ergodic Theorem).** *Let  $(X, \mathcal{B}, \mu, T)$  be a measure-preserving system on a probability space, and  $f \in L^1_\mu$ . Define*

$$E_\alpha = \{x \in X | \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) > \alpha\}.$$

*Then,  $\alpha\mu(E_\alpha) \leq \int_{E_\alpha} f d\mu \leq \|f\|_1$ , and for all  $A \in \mathcal{B}$  so that  $T^{-1}A = A$ ,*

$$\alpha\mu(E_\alpha \cap A) \leq \int_{E_\alpha \cap A} f d\mu.$$

*Proof.* Let  $g = f - \alpha$  and  $U$  be the linear operator induced by  $T$ , i.e.  $Ug = g \circ T$  for all  $g \in L^1_\mu$ . Then  $U$  is positive with norm 1. With the notation in the last proposition,

$$E_\alpha = \{x \in X | \sup_{N \geq 1} F_N(x) > 0\} = \bigcup_{N \geq 1} \{x | F_N(x) > 0\}.$$

By the last proposition,  $\int_{E_\alpha} g d\mu \geq 0$ . Thus,  $\|f\|_1 \geq \int_{E_\alpha} f d\mu \geq \alpha\mu(E_\alpha)$ . For  $A \in \mathcal{B}$  so that  $T^{-1}A = A$ , consider the system restricted to  $A$ . Then by the same argument,

$$\alpha\mu(E_\alpha \cap A) \leq \int_{E_\alpha \cap A} f d\mu.$$

□

Beside of this proof, there is also an alternative proof for this theorem via the Vitali covering lemma, which presents a hint that the covering

lemmas and maximal inequality may be presenting the same property of the space.

The maximal ergodic theorem presents that the measure of the exceptional set where the ergodic sum goes beyond some value is controlled by the  $L^1$  norm of the function. On the other hand, the Birkhoff's pointwise ergodic theorem presents the behavior of generic points in a measure-preserving system.

**Theorem 1.8. (*Birkhoff's Pointwise Ergodic Theorem*).** *Let  $(X, \mathcal{B}, \mu, T)$  be a measure-preserving system and  $f \in L^1_\mu$ . Then, there exists a  $T$ -invariant function  $f^* \in L^1_\mu$  so that*

$$\frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) \rightarrow f^*(x) \text{ in } L^1_\mu$$

*almost everywhere, and  $\int f^* d\mu = \int f d\mu$ .*

*Moreover, if  $T$  is ergodic, then  $f^*(x) = \int f d\mu$  almost everywhere.*

*Proof.* Define  $A_n(f)(x) = \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x)$  for all  $x$ , and set

$$L(f) = \liminf_{n \rightarrow \infty} A_n(f)$$

and

$$M(f) = \limsup_{n \rightarrow \infty} A_n(f)$$

pointwisely. Then observe that

$$A_{n+1}(f)(x) = \frac{n}{n+1} A_n(f)(Tx) + \frac{1}{n+1} f(x).$$

Let  $\{A_{n_i}(f)(x)\}$  be a subsequence such that

$$\lim_{i \rightarrow \infty} A_{n_i}(f)(Tx) = M(f)(Tx).$$

Then

$$\limsup_{n \rightarrow \infty} A_n(f)(x) \geq \limsup_{i \rightarrow \infty} A_{n_{i+1}}(f)(x) \geq M(f)(Tx).$$

Let  $\{A_{n_j}(f)(x)\}$  be a subsequence such that

$$\lim_{j \rightarrow \infty} A_{n_j}(f)(x) = M(f)(x).$$

Then

$$\limsup_{n \rightarrow \infty} A_n(f)(Tx) \geq \limsup_{j \rightarrow \infty} A_{n_j-1}(f)(Tx) \geq M(f)(x).$$

Thus,  $M(f)(Tx) = M(f)(x)$  for all  $x$ . Similarly,  $L(f)(Tx) = L(f)(x)$  for all  $x$ . For  $\alpha \in \mathbb{R}$ , let  $E_\alpha = \{x \in X : \sup_{n \in \mathbb{N}} A_n(f)(x) > \alpha\}$ . For  $\alpha, \beta \in \mathbb{R}$ , let  $E_\alpha^\beta = \{x \in X : L(x) < \beta \text{ and } M(x) > \alpha\}$ . Then  $T^{-1}E_\alpha^\beta = E_\alpha^\beta$  and  $E_\alpha^\beta \subseteq E_\alpha$  for all  $\alpha > \beta$ . By theorem 1.7,

$$\int_{E_\alpha^\beta} f d\mu \leq \alpha \mu(E_\alpha^\beta).$$

Then apply the same argument with  $-f$ , we have

$$\int_{E_\alpha^\beta} f d\mu \geq \beta \mu(E_\alpha^\beta).$$

Then

$$\mu(\{x \in X : L(f)(x) < M(f)(x)\}) \leq \sum_{\alpha > \beta \in \mathbb{Q}} \mu(E_\alpha^\beta) = 0.$$

Thus, for a.e.  $x$ ,  $L(f)(x) = M(f)(x)$ , and  $A_n(f)(x)$  converges almost everywhere. Let  $f^*(x) = M(f)(x)$ , then  $f^*$  is  $T$ -invariant and  $A_n(f)(x) \rightarrow f^*(x)$  a.e. Then for all  $n$ ,

$$\int f d\mu = \int A_n(f) d\mu,$$

and by dominated convergence theorem,

$$\int f^* d\mu = \int f d\mu.$$

If  $T$  is ergodic, then  $f^*$  is constant almost everywhere, and it follows that

$$f^*(x) = \int f d\mu$$

almost everywhere. □

A natural generalization of the last theorem is the following theorem due to Hoph.

**Theorem 1.9. (*Hoph's Ratio Ergodic Theorem*).**

*Let  $(X, \mathcal{B}, \mu, T)$  be an ergodic conservative measure-preserving system*

with  $X$   $\sigma$ -finite and  $f, g \in L^1_\mu$  with  $g > 0$  almost everywhere. Then

$$\frac{\sum_{j=0}^{n-1} f(T^j x)}{\sum_{j=0}^{n-1} g(T^j x)} \rightarrow \frac{\int f}{\int g}$$

almost everywhere.

When  $X$  is a probability measure space, this theorem is equivalent to Birkhoff's ergodic theorem since in this case we may take  $g = 1$  to reduce this theorem to Birkhoff's and the other implication is clear. When  $X$  is not a probability space, however, the equivalence does not hold.

In the case where  $T$  is invertible, we can think of the transformation generating an  $\mathbb{Z}$  action on  $X$ . Then, the following theorem, proved by Hochman, further generalizes Hopf's ratio ergodic theorem: instead of  $\mathbb{Z}$  actions, we now consider  $\mathbb{Z}^d$  actions.

**Theorem 1.10.** *Let  $\{T^u\}_{u \in \mathbb{Z}^d}$  be a free, non-singular ergodic action on a  $\sigma$ -finite standard probability space  $(X, \mu)$ . Let  $\|\cdot\|$  be a norm on  $\mathbb{R}^d$ . Let  $B_n = \{u \in \mathbb{Z}^d : \|u\| \leq n\}$ . Then for  $f, g \in L^1(\mu)$  with  $\int g \neq 0$  and  $g > 0$ ,*

$$\frac{\int_{B_n} \hat{T}^u f d\nu}{\int_{B_n} \hat{T}^u g d\nu} \rightarrow \frac{\int f d\mu}{\int g d\mu}$$

almost everywhere.

In section 5, we will further generalize this theorem to the Heisenberg group. The Følner sequence and ratio maximal inequality are two main tools for proving our main theorem and they will be discussed in section 2 and 4, respectively. We also use the notion of ergodicity for group actions in this theorem, which will be defined in the next section.

## 2. AMENDABLE GROUPS AND ERGODICITY OF ACTIONS

In the last section, we discussed about the ergodic theory in the measure-preserving transformation setting. Most of the definitions and theorems can be generalized to the group action setting.

In order to consider the group actions generally, the following basic properties that a group action may possess will be used later in this thesis.

**Definition 2.1.** Let  $G \curvearrowright X$  be a group action, then we say the action is

- (1) free if for all  $x \in X$  and  $g \in G$ ,  $g \cdot x = x$  implies that  $g = e$ ;
- (2) conservative if it has no nontrivial wandering sets;
- (3) non-singular with respect to a measure  $\mu$  on  $X$  if  $\mu(A) = 0$  if and only if  $\mu(g \cdot A) = 0$  for all  $g$ ;
- (4) transitive if for all  $x, y \in X$ , there exists  $g \in G$  such that  $g \cdot x = y$ ;
- (5) isometric with respect to a metric  $d$  on  $X$  if for all  $x, y \in X$  and  $g \in G$ ,  $d(x, y) = d(g \cdot x, g \cdot y)$ .

**Definition 2.2.** A measure-preserving system is a quartet  $(X, \mathcal{B}, \mu, T)$  where  $(X, \mathcal{B}, \mu)$  is a probability space and  $T$  is a measure-preserving group action on  $X$ .

The measure-preserving systems we discussed in the last section can be regard as such systems where  $T$  is an  $\mathbb{N}$  or  $\mathbb{Z}$  action. Let  $MPT(X, \mathcal{B}, \mu)$  be the group of invertible measure-preserving transformation of  $X$ . Then  $T$  induces a homomorphism from the group  $G$  to  $MPT(X, \mathcal{B}, \mu)$ .

For our purpose, we may assume that the group action  $T : G \times X \rightarrow X$  is a continuous map,  $X$  is a  $\sigma$ -compact and locally compact metric space, and the group  $G$  is locally compact.

**Definition 2.3.** Let  $\mathcal{M}(X)$  be the space of all Borel probability measures on  $X$ . A measure  $\mu \in \mathcal{M}(X)$  is called  $T$ -invariant or invariant under  $G$  if for all  $g \in G$  and  $A \in \mathcal{B}_X$ ,  $\mu(A) = \mu(g^{-1}A)$ , i.e.  $g_*\mu = \mu$  for all  $g \in G$ . We denote the set of all the  $G$ -invariant measures on  $X$  as  $\mathcal{M}^G(X)$ .

There are examples of group actions where  $\mathcal{M}^G(X)$  is empty. To ensure the existence of invariant measures, we introduce the notion of amenability.

**Definition 2.4.** Let  $G$  be a  $\sigma$ -compact and locally compact group, and  $m_G$  be a left Haar measure on  $G$ .  $G$  is called amenable if for any compact subset  $K$  and  $\epsilon > 0$ , there is a measurable set  $F \subseteq G$  with  $\bar{F}$  compact such that  $KF$  is measurable and that

$$m_G(F \Delta KF) < \epsilon m_G(F).$$

Then, we call such set  $F$  a  $(K, \epsilon)$ -invariant set.

**Definition 2.5.** Let  $\{F_n\}_{n \in \mathbb{N}}$  be a sequence of compact sets in  $G$ .  $\{F_n\}_{n \in \mathbb{N}}$  is a Følner sequence if for any compact subset  $K$  and  $\epsilon > 0$ ,  $F_n$  is eventually  $(K, \epsilon)$ -invariant.

It follows from the definition that the existence of Følner sequences implies amenability. Moreover, along the Følner sequences, we can compute the ergodic averages of  $G$ -actions. The mean ergodic theorem and pointwise ergodic theorems can be generalized to amenable group actions via the Følner sequences.

The following theorem presents that amenability of a group implies existence of invariant measure. In fact, the existence of invariant probability measure of continuous actions and the existence of Følner sequences are both equivalent to the amenability. However, the other implications are not used in this thesis.

**Theorem 2.6.** *Let  $G$  be a locally compact amenable group,  $X$  be a compact metric space, and  $T : G \times X \rightarrow X$  be a continuous action. Then there exists a probability measure  $\mu$  on  $X$  so that  $\mu$  is  $T$ -invariant.*

*Proof.* As  $G$  is amenable, let  $\{F_n\}_{n \in \mathbb{N}}$  be a Følner sequence. Since  $T$  is a continuous group action, it induces a  $G$ -action on  $\mathcal{M}(X)$ . For each  $\nu \in \mathcal{M}(X)$ , we define the averaged measure  $\mu_n$  to be the measure so that for all  $f \in C(X)$ ,

$$\int f d\mu_n = \frac{1}{m_G(F_n)} \int_{F_n} \int f(g \cdot x) d\nu(x) dm_G(g).$$

Since  $X$  is compact and we equip  $\mathcal{M}(X)$  with the weak\* topology,  $\mu_n$  would have a weak\* converging subsequence converging to  $\mu \in \mathcal{M}(X)$ . Then,

$$\begin{aligned}
& \left| \int f(x) d\mu_n(x) - \int f(h \cdot x) d\mu_n(x) \right| \\
&= \frac{1}{m_G(F_n)} \int_{F_n} \int f(g \cdot x) d\nu(x) dm_G(g) \\
&\quad - \frac{1}{m_G(F_n)} \int_{hF_n} \int f(g \cdot x) d\nu(x) dm_G(g) \\
&\leq \frac{1}{m_G(F_n)} \int_{F_n \Delta hF_n} \int f(g \cdot x) d\nu(x) dm_G(g) \\
&\leq \frac{m_G(F_n \Delta hF_n)}{m_G(F_n)} \|f\|_\infty.
\end{aligned}$$

Thus,

$$\begin{aligned}
& \left| \int f(x) d\mu(x) - \int f(h \cdot x) d\mu(x) \right| \\
&= \lim_{n \rightarrow \infty} \left| \int f(x) d\mu_n(x) - \int f(h \cdot x) d\mu_n(x) \right| = 0.
\end{aligned}$$

Therefore,  $\mu$  is  $T$ -invariant.  $\square$

A simpler case is that when  $G$  compact, then  $G$  is amenable and unimodular, in which case we also have the existence of invariant probability measure.

**Proposition 2.7.** *Let  $G$  be a compact group,  $X$  be a compact metric space, and  $T : G \times X \rightarrow X$  be a continuous action. Then there exists a probability measure  $\mu$  on  $X$  so that  $\mu$  is  $T$ -invariant.*

*Proof.* Let  $m_G$  be the normalized Haar measure, and  $t \in X$  be an arbitrary point. Define  $\phi : G \rightarrow X$  by  $\phi(g) = g \cdot t$  for all  $g \in G$ . Then  $\phi$  is continuous and the pushforward measure  $\mu = \phi_*(m_G)$  is by

definition T-invariant, since for  $g \in G$  and  $A \in \mathcal{B}_X$ ,

$$\begin{aligned} \mu(g^{-1} \cdot A) &= m_G(\phi^{-1}(g^{-1} \cdot A)) \\ &= m_G(g^{-1} \cdot \phi^{-1}(A)) \\ &= m_G(\phi^{-1}(A)) \\ &= \mu(A). \end{aligned}$$

□

Similar to the general setting, we can define ergodicity of group actions.

**Definition 2.8.** Let  $G$  be a group acting continuously on a compact metric space  $X$ , and  $\mu \in \mathcal{M}^G(X)$ . The action is said to be ergodic if for all  $A \in \mathcal{B}_X$ ,  $\mu(g^{-1}A \Delta A) = 0$  for all  $g \in G$  implies that  $\mu(A) = 0$  or  $\mu(A) = 1$ .

There is a useful equivalent definition of ergodicity for group action, which could be realized as a generalization of proposition 1.4.

**Proposition 2.9.** *Let  $G$  be a group acting continuously on a compact metric space  $X$ , and  $\mu \in \mathcal{M}^G(X)$ . Then the following are equivalent:*

- (1) *The action is ergodic;*
- (2) *For all  $f : X \rightarrow \mathbb{C}$  measurable function,  $f(gx) = f(x)$  for  $\mu$ -a.e.  $x \in X$  and for all  $g \in G$  implies that  $f$  is constant a.e.;*

*Proof.* (2)  $\implies$  (1): Let  $A \in \mathcal{B}$  so that  $\mu(g^{-1}A \Delta A) = 0$  for all  $g \in G$  and set  $f = \chi_A$  to be the indicator function of  $A$ . Then for  $g \in G$ ,

$$\mu(\{x \in X : f(x) \neq f(gx)\}) = \mu(A \Delta gA) = 0.$$

Thus,  $f(x) = f(gx)$  almost everywhere, so  $f$  is constant almost everywhere. Thus,  $f = 1$  almost everywhere or  $f = 0$  almost everywhere, which implies that either  $\mu(A) = 0$  or 1.

(1)  $\implies$  (2): Let  $f : X \rightarrow \mathbb{C}$  be a measurable function such that  $f(gx) = f(x)$  for all  $g \in G$  and  $x \in X$ . Since  $\Re(f)(gx) + i\Im(f)(x) = \Re(f)(x) + i\Im(f)(x)$  implies that  $\Re(f)(gx) = \Re(f)(x)$  and  $\Im(f)(gx) = \Im(f)(x)$ , we may assume without loss of generality that  $f$  is a real



valued function. For all  $k \in \mathbb{Z}$  and  $n \in \mathbb{N}$ , define  $A_n^k = f^{-1}(\left[\frac{k}{n}, \frac{k+1}{n}\right))$ . Then for all  $k, g$  and  $n$ ,

$$\mu(g^{-1}A_n^k \Delta A_n^k) \leq \mu(\{x \in X | f(T(x)) \neq f(x)\}) = 0.$$

Since action is ergodic, either  $\mu(A_n^k) = 0$  or  $\mu(A_n^k) = 1$ . For each  $n$ ,  $\{A_n^k\}$  are all disjoint and  $X = \bigcup_k A_n^k$ . Therefore, there exists a unique  $k_n$  so that  $\mu(A_n^{k_n}) = 1$ . For  $x, y \in A_n^{k_n}$ ,

$$|f(x) - f(y)| \leq \left| \frac{k_n}{n} - \frac{k_n + 1}{n} \right| = \frac{1}{n}.$$

Let  $Y = \bigcap_n A_n^{k_n}$ , then  $\mu(Y) = 1$ , and for  $x, y \in Y$  and

$$|f(x) - f(y)| \leq \frac{1}{n} \quad \forall n \in \mathbb{N}.$$

Hence,  $f$  is constant almost everywhere.  $\square$

Another condition for group actions is mixing, which also measures how chaotic the system is. Mixing actions are necessarily ergodic, and there are examples of ergodic actions that are not mixing.

**Definition 2.10.** Let  $G$  be a group acting continuously on a compact metric space  $X$ , and  $\mu \in \mathcal{M}^G(X)$ . The action is said to be mixing if for all  $A, B \in \mathcal{B}$ , and sequence  $\{g_i\}_{i \in \mathbb{N}}$  in  $G$  so that  $|g_i \cap K| < \infty$  for all compact subset  $K \subseteq G$ ,

$$\lim_{n \rightarrow \infty} \mu(A \cap g_i^{-1}B) = \mu(A)\mu(B).$$

**Proposition 2.11.** *Let  $G$  be a group acting continuously on a compact metric space  $X$ , and  $\mu \in \mathcal{M}^G(X)$ . If the action is mixing, then it is ergodic.*

*Proof.* Let  $A \in \mathcal{B}_X$  such that  $\mu(g^{-1}A \Delta A) = 0$  and let  $\{g_i\}_{i \in \mathbb{N}}$  be a sequence in  $G$  so that  $|g_i \cap K| < \infty$  for all compact subset  $K \subseteq G$ . Then, for all  $i$ ,

$$\mu(g_i^{-1}A \Delta A) = \mu(A) + \mu(g_i^{-1}A) - 2\mu(A \cap g_i^{-1}A) = 0.$$

Thus,  $\mu(A) = \mu(A \cap g_i^{-1}A) = 0$  for all  $i$ , and

$$\mu(A) = \lim_{n \rightarrow \infty} \mu(A \cap g_i^{-1}A) = \mu(A)^2.$$

Hence,  $\mu(A) = 0$  or  $1$ .  $\square$

However, the other implication is not generally true. An example of ergodic action which is not mixing is circle rotation  $R_\alpha$  with  $\alpha$  irrational.

**Proposition 2.12.** *Consider  $\mathbb{S}^1 \cong \mathbb{R}/\mathbb{Z}$  as a circle and for  $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ , define  $R_\alpha : \mathbb{S}^1 \rightarrow \mathbb{S}^1$  as for all  $[s] \in \mathbb{S}^1$ ,  $R_\alpha([s]) = [s + \alpha]$ . Then the  $\mathbb{Z}$ -action generated by  $R_\alpha$  is ergodic but not mixing with respect to the Lebesgue measure.*

*Proof.* Let  $f \in L^2(m)$  be an invariant function. Then it can be written as Fourier series  $f(x) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x}$  almost everywhere. Then for all  $m$ ,

$$f(R_\alpha^m x) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n m \alpha} e^{2\pi i n x} = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x}.$$

Thus, for all  $n$  and  $m$ ,  $a_n = a_n e^{2\pi i n m \alpha}$ . For  $n \neq 0$ , since  $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ ,  $e^{2\pi i n m \alpha} \neq 1$ , so  $a_n = 0$ . Thus,  $f(x) = a_0$  almost everywhere, and the action is ergodic.

Let  $A = \{[x] : x \in [0, \frac{1}{10}]\}$ , and fix  $\{n_i\}_{i \in \mathbb{N}}$  a sequence in  $\mathbb{Z}$  so that  $|n_i \cap K| < \infty$  for all compact subset  $K \subseteq \mathbb{Z}$ . Then there are infinity many  $i \in \mathbb{N}$  such that  $n_i \notin \{[x] : x \in [-\frac{2}{10}, \frac{3}{10}]\}$ . Then for such  $i$ ,  $m(A \cap R_\alpha^{-n_i} A) = 0$ . Thus,

$$\liminf_{i \rightarrow \infty} m(A \cap R_\alpha^{-n_i} A) = 0 \neq m(A)^2,$$

so the action is not mixing.  $\square$

### 3. EXAMPLE: GEODESTIC FLOWS ON QUOTIENTS OF HYPERBOLIC PLANE

In this section, we will present a classical example of mixing group actions. We first introduce a classical model in hyperbolic geometry. Consider the upper half plane,

$$\mathbb{H} = \{x + iy : x, y \in \mathbb{R}, y > 0\}.$$

Then  $\mathbb{H}$  is a connected Hausdorff smooth manifold, and we can define tangent bundle on  $\mathbb{H}$ . For our purposes, we can identify the tangent bundle  $T\mathbb{H}$  as a disjoint union of spaces for derivatives at each point in  $\mathbb{H}$ . For each  $z \in \mathbb{H}$ , let  $T_z\mathbb{H} = \{z\} \times \mathbb{C}$ . Then,  $T\mathbb{H} \cong \sqcup_{z \in \mathbb{H}} T_z\mathbb{H} \cong \mathbb{H} \times \mathbb{C}$ .

For  $z = x + iy \in \mathbb{H}$ , we can define an inner product on  $T_z\mathbb{H}$  as, for  $v, w \in \mathbb{C} \cong T_z\mathbb{H}$ ,  $\langle v, w \rangle_z = \frac{1}{y^2} v\bar{w}$ . The collection of such inner product is a hyperbolic Riemannian metric on  $\mathbb{H}$ , and it is a smooth structure on  $\mathbb{H}$ .

Given a differentiable function  $f : [0, 1] \rightarrow \mathbb{H}$  and  $t \in [0, 1]$ , we define its derivative at  $t$  as  $Df(t) = (f(t), f'(t)) \in T_z\mathbb{H}$ . Then for  $z, w \in \mathbb{C}$ , a function  $\phi : [0, 1] \rightarrow \mathbb{H}$  is said to be a path from  $z$  to  $w$  if it is continuous and piecewise differentiable, with  $\phi(0) = z, \phi(1) = w$ . We can also define its length as

$$L(\phi) = \int_0^1 \sqrt{\langle D\phi(t), D\phi(t) \rangle_{\phi(t)}} dt.$$

Finally, we can define the distance between two points  $z, w \in \mathbb{H}$  as the length of shortest path between them, i.e.,

$$d(z, w) = \inf\{L(\phi) : \phi \text{ a path from } z \text{ to } w\}.$$

One can easily see from the construction that this is indeed a metric on  $\mathbb{H}$ , and it generates the same topology as the one induced by the inclusion map to  $\mathbb{C}$ . We can also compute the metric explicitly, assuming some of the following propositions, which is given by

$$d(z, w) = \log\left(\frac{|z - \bar{w}| + |z - w|}{|z - \bar{w}| - |z - w|}\right)$$

for all  $z, w \in \mathbb{H}$ .

Then, the projective special linear group

$$PSL_2(\mathbb{R}) = SL_2(\mathbb{R})/\{\pm I_2\}$$

acts naturally on  $\mathbb{H}$  by the conformal maps, for  $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PSL_2(\mathbb{R})$  and  $z \in \mathbb{H}$ ,

$$g \cdot z = \frac{az + b}{cz + d}.$$

Each  $g \in PSL_2(\mathbb{R})$  can be realized as a differentiable map from  $\mathbb{H}$  to  $\mathbb{H}$ , and we may define the derivative action  $D_g : T\mathbb{H} \rightarrow T\mathbb{H}$  by

$$D_g(z, v) = (g(z), g'(z)v) = \left( \frac{az + b}{cz + d}, \frac{v}{(cz + d)^2} \right)$$

for all  $(z, v) \in T\mathbb{H}$ . By the laws of differentiation, one can check that this is indeed an action on  $T\mathbb{H}$ . Also, for  $g, z$  and  $v \in T_z\mathbb{H}$ , let  $(D_g)_z(v) = \frac{v}{(cz+d)^2} \in T_{g \cdot z}\mathbb{H}$ .

**Proposition 3.1.** *The action defined above satisfies the following properties:*

- (1) *it is a well defined action;*
- (2) *it is transitive;*
- (3) *it is isometric;*
- (4)  $Stab_{PSL_2(\mathbb{R})}(i) = PSO_2(\mathbb{R}) = SO_2(\mathbb{R})/\{\pm I_2\}$ .

*Proof.* (1) Let  $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PSL_2(\mathbb{R})$  and  $z = x + iy \in \mathbb{H}$ . Then,

$$\Im(g \cdot z) = \Im\left(\frac{ax + b + iay}{cx + d + icy}\right) = \frac{\det(g)}{y} > 0.$$

The axioms for group action follow from the law of matrix multiplication.

(2) Let  $z = x + iy$  and  $w = a + ib \in \mathbb{H}$ . Set

$$M_z = \begin{pmatrix} \sqrt{y} & \frac{x}{\sqrt{y}} \\ 0 & \frac{1}{\sqrt{y}} \end{pmatrix}$$

and

$$M_w = \begin{pmatrix} \sqrt{b} & \frac{a}{\sqrt{b}} \\ 0 & \frac{1}{\sqrt{b}} \end{pmatrix}$$

Then  $M_z$  and  $M_w \in PSL_2(\mathbb{R})$   $M_z \cdot i = z$  and  $M_w \cdot i = w$ , so  $(M_w^{-1}M_z) \cdot z = w$ .

(3) Let  $z \in \mathbb{H}$ , then by calculation, for  $v, w \in T_z\mathbb{H}$ ,

$$\langle (D_g)_z(v), (D_g)_z(w) \rangle_{g \cdot z} = \langle v, w \rangle_z.$$

Thus, the length of vectors are preserved by the derivative map. It follows that for any piecewise differentiable path  $f$ ,  $L(f) = L(g \circ f)$ . Then by definition of the metric, for all  $p, q \in \mathbb{H}$  and  $g \in \mathbb{H}$ ,

$$d(p, q) = d(g \cdot p, g \cdot q),$$

and the action is therefore isometric.

(4) Let  $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PSL_2(\mathbb{R})$ . By calculation, we have for all  $z \in \mathbb{H}$  that

$$\Im(g \cdot z) = \frac{\Im(z)}{|cz + d|^2}$$

Suppose  $g \cdot i = i$ , Then  $|ci + d| = 1$ . Therefore, there exists  $\theta \in [0, 2\pi)$  such that  $c = \sin \theta$  and  $d = \cos \theta$ . Then  $g \cdot i = i$  if and only if

$$\frac{ai + b}{i \sin \theta + \cos \theta} = i,$$

which is equivalent to  $a = -\sin \theta$  and  $b = \cos \theta$ . Thus,  $g \cdot i = i$  if and only if  $g \in PSO_2(\mathbb{R})$ .  $\square$

By the last statement, we may make the identification that

$$\mathbb{H} \cong PSL_2(\mathbb{R})/PSO_2(\mathbb{R}).$$

In order to analyze the geodesic flows, we firstly find all geodesics of  $\mathbb{H}$ .

**Proposition 3.2.** *For  $z, w \in \mathbb{C}$ , there is a unique path  $\phi$  of constant unit speed with  $\phi(0) = z$  and  $\phi(d(z, w)) = w$  whose image is in*

- (1) a vertical line, if  $\Re(z) = \Re(w)$ ;
- (2) a semicircle with center on the real line, if  $\Re(z) \neq \Re(w)$ .

*Proof.* (1) Suppose  $\Re(z) = \Re(w) = 0$  and assume without loss of generality that  $\Im(z) = y_z$  and  $\Im(w) = y_w$  where  $y_z < y_w$ . Then the path

$$\phi(t) = iy_z \left( \frac{y_w}{y_z} \right)^{\frac{t}{\log(y_w) - \log(y_z)}}$$

is of constant unit speed with length  $L(\phi) = \log(y_w) - \log(y_z)$ . Then for any path  $\psi$  from  $z$  to  $w$ , since  $\psi$  is piecewise differentiable,  $\Im(\psi)'$  is defined almost everywhere. Thus,

$$L(\psi) = \int_0^1 \frac{\|\psi'(t)\|}{\Im(\psi(t))} dt \geq \int_0^1 \frac{\Im(\psi)'}{\Im(\psi)} dt = \log(y_w) - \log(y_z).$$

Since  $\phi$  achieves this lower bound, a reparameterization of  $\phi$  will have the same length and reach  $w$  at  $t = 1$ . Thus,  $d(z, w) = \log(y_w) - \log(y_z)$ . In the inequality above, the equality holds if and only if  $\Re(\psi)(t) = 0$  a.e. and the uniqueness follows.

Now for  $z, w \in \mathbb{H}$ , we claim that there is a  $g \in PSL_2(\mathbb{R})$  so that  $g \cdot z = i$  and  $g \cdot w = iy$  for some  $y > 1$ . If the claim holds, since  $PSL_2(\mathbb{R})$  acts transitively and isometrically, the unique path from  $i$  to  $iy$  of unit constant speed will be translated to a unique path from  $z$  to  $w$  with unit constant speed hitting  $w$  at time  $d(z, w)$ .

Then we prove the claim. By the previous proposition, there is a  $g_0 \in PSL_2(\mathbb{R})$  so that  $g_0 \cdot z = i$ . We pick  $g_1 \in PSO_2(\mathbb{R})$  such that  $\Im(g_1 g_0 \cdot w) = \sup\{\Im(g g_0 \cdot w) : g \in PSO_2(\mathbb{R})\}$ . Then let  $g = g_1 g_0$  and by calculation,  $\Re(g \cdot w) = 0$  and  $\Im(g \cdot w) > 1$ . Since  $PSO_2(\mathbb{R})$  fixes  $i$ ,  $g \cdot z = i$ .

Note that  $PSL_2(\mathbb{R})$  acts on  $\mathbb{H}$  by fractional linear transformations, so the image of the positive imaginary line will either be a subset of straight line or a subset of circle in the complex plane. Then if  $\Re(z) = \Re(w) = b$ , take  $g = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$  and  $g$  maps the positive imaginary line to a vertical line with real part  $b$ . Otherwise, we consider the intersection points  $t, s$  of the real line with the semicircle containing  $z$  and  $w$ . Take  $g$  such that  $g^{-1}$  is the fractional linear transformation that maps  $t$  to  $0$ ,  $z$  to  $i$ , and  $s$  to  $\infty$ , then  $g$  maps the positive imaginary line to the circle as desired.  $\square$

A restatement of this proposition is that the geodesic curves in  $\mathbb{H}$  is either a vertical line or a semicircle with center on the real line. It follows from the previous two proposition that given  $z, w \in \mathbb{H}$  the geodesic curve is determined by the initial point  $z$  and the initial direction  $v$  of the geodesic path. Moreover, there is a unique  $g \in PSL_2(\mathbb{R})$  so that  $D_g(i, i) = (z, v) \in T_z^1\mathbb{H}$ .

Then we can define the geodesic flow on  $\mathbb{H}$  as  $g_t : T^1\mathbb{H} \rightarrow T^1\mathbb{H}$ , for  $(z, v) \in T^1\mathbb{H}$ , its image under  $g_t$  is the image and direction of the geodesic path determined by  $(z, v)$  at time  $t$ .

Note that we may identify  $T^1\mathbb{H}$  with  $PSL_2(\mathbb{R})$ , since we may identify  $\mathbb{H}$  with  $PSL_2(\mathbb{R})/PSO_2(\mathbb{R})$ , and for each  $z \in \mathbb{H}$ ,  $T_z^1\mathbb{H} \cong PSO_2(\mathbb{R})$ . We can also easily check that the action of  $PSL_2(\mathbb{R})$  on  $T^1\mathbb{H}$  is transitive, free, and of no non-trivial isotropy groups.

Via this identification, the geodesic flow  $\mathbb{R} \curvearrowright T^1\mathbb{H}$  corresponds to the action  $\mathcal{T} \curvearrowright PSL_2(\mathbb{R})$  by left translation, where

$$\mathcal{T} = \{g_t \in PSL_2(\mathbb{R}) : g_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}, t \in \mathbb{R}\}.$$

**Proposition 3.3.** *Let  $\Gamma$  be a lattice in  $PSL_2(\mathbb{R})$ , i.e.,  $\Gamma$  is discrete and  $PSL_2(\mathbb{R})/\Gamma$  has finite measure. Then the geodesic flow of  $\Gamma \backslash PSL_2(\mathbb{R})$  is mixing.*

*Proof.* Let  $X = \Gamma \backslash PSL_2(\mathbb{R})$  and  $\mu$  be the normalized Haar measure on  $X$ . Then  $\mathcal{T} \curvearrowright X$  by left translation, which also induces an action  $\pi : \mathcal{T} \curvearrowright L^2(X)$  defined for  $g_t \in \mathcal{T}$  and  $f \in L^2(X)$ ,

$$\pi(g_t)f(x) = f(g_t \cdot x).$$

Then for each  $g_t \in \mathcal{T}$ ,  $\pi(g)$  is a unitary transformation on  $L^2(X)$ . We observe in this setting that the geodesic flow of  $\Gamma \backslash PSL_2(\mathbb{R})$  is mixing if and only if for all  $f, g \in L^2(X)$ ,

$$\langle f, \pi(g_t)g \rangle \rightarrow \int f d\mu \int \bar{g} d\mu.$$

Suppose for contradiction that there exist  $f, g \in L^2(X)$  and  $t_{n_i} \rightarrow \infty$  so that the equation above does not hold. By linearity, we may assume

without loss of generality that  $\int f d\mu = \int g d\mu = 0$ . Then,

$$\langle f, \pi(g_{t_{n_i}})g \rangle \rightarrow 0.$$

Since the ball of radius  $\|g\|$  is weakly compact in  $L^2(X)$ , there is a subsequence  $\{\pi(g_{t_{n_{i_j}}})g\} \rightarrow \hat{g} \in L^2(X)$  weakly. Let  $h_j = g_{t_{n_{i_j}}}$  for all  $j$ , and define

$$U = \{u_s = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} : s \in \mathbb{R}\}.$$

Then by calculation,

$$h_j^{-1}u_s h_j = u_{se^{-t_{n_{i_j}}}} \rightarrow I_2.$$

Then for all  $\phi \in L^2(X)$ ,

$$\begin{aligned} \langle \pi(u_s)\hat{g} - \hat{g}, \phi \rangle &= \lim_{j \rightarrow \infty} \langle \pi(u_s)\pi(h_j)g - \pi(h_j)g, \phi \rangle \\ &= \lim_{j \rightarrow \infty} \langle \pi(h_j)\pi(u_{se^{-t_{n_{i_j}}}})g, \phi \rangle - \lim_{j \rightarrow \infty} \langle \pi(h_j)g, \phi \rangle \\ &= 0 \end{aligned}$$

Thus,  $\hat{g}$  is invariant under  $U$ . We can also check by a similar calculation that  $\langle \pi(g_t)\hat{g}, \hat{g} \rangle = \|\hat{g}\|^2$ . Thus,  $\hat{g}$  is invariant under  $SL_2(\mathbb{R})$ . For all  $A \in SL_2(\mathbb{R})$ ,  $\hat{g}(x) = \hat{g}(Ax)$  a.e.. Thus,  $\hat{g} = c$  for some  $c \in \mathbb{C}$  almost everywhere since the action is transitive. Then,

$$c = \langle c, 1 \rangle = \lim_{j \rightarrow \infty} \langle \pi(h_j)g, 1 \rangle = 0,$$

but that implies

$$\langle f, \pi(g_{t_{n_i}})g \rangle \rightarrow 0,$$

which is a contradiction.  $\square$



#### 4. MAXIMAL INEQUALITY AND BESICOVITCH COVERING PROPERTY

In the proof for maximal ergodic theorem in section 1, we used the maximal inequality on probability spaces. That maximal inequality gives a bound of the size of the set where the partial sum of functions exceeds a given bound. In fact, inequalities with similar idea are helpful tools for proving ergodic theorems. In order to prove a ratio ergodic theorem, we will make use of the ratio maximal inequality. On the other hand, Besicovitch covering property is describing the behaviour of open covers in a metric space.

**Definition 4.1.** Let  $(M, d)$  be a metric space. We say  $M$  admits the Besicovitch Covering Property if there exist  $N \in \mathbb{N}$  so that for all bounded subset  $A \subseteq M$ , and family  $\{B_\alpha\}_{\alpha \in I}$  of balls such that  $B = B_\alpha$  is centered at  $\alpha$  for all  $\alpha \in I$ , there exists a subfamily  $F \subseteq B$  such that  $F$  covers  $A$  and each point in  $M$  is contained in at most  $N$  balls in  $F$ .

Similarly, in our setting, we may define Besicovitch Covering Property for subsets of a group as follows:

**Definition 4.2.** Let  $\{B_n\}_{n \in \mathbb{N}}$  be a sequence of subsets of group  $G$ . We say it has the Besicovitch covering property with constant  $C$  if for all  $E \subseteq G$  finite, and a collection of subsets  $D = \{D_g\}_{g \in E}$  such that  $D_g = B_n g$  for some  $n \in \mathbb{N}$ , there exists a subfamily  $F \subseteq D$  so that  $F$  covers  $E$  and each point in  $G$  is contained in at most  $N$  subsets in  $F$ .

Then if  $G$  can be equipped with a metric  $d$  so that  $(G, d)$  has the Besicovitch Covering Property with multiplicity  $N$ , then any sequence of subsets of  $G$  will satisfy the Besicovitch Covering Property with constant  $N$ .

In order to introduce the ratio maximal inequality, we first define the ratio ergodic mean for measure-preserving systems of group actions.

**Definition 4.3.** Let  $G$  be a amenable group acting on measure space  $(X, \mu)$ , and  $\{B_n\}_{n \in \mathbb{N}}$  be a sequence of subsets containing the identity.

For  $f, g \in L^1(\mu)$ , let

$$R_n(f, g) = \frac{\int_{B_n} f \circ u d\nu(u)}{\int_{B_n} g \circ u d\nu(u)}$$

where  $\nu$  is the Haar measure on  $G$ .

**Definition 4.4.** Let  $G \curvearrowright X$  be an ergodic action, and  $\{B_n\}_{n \in \mathbb{N}}$  be a sequence of subsets containing the identity. We say the action admits the ratio maximal inequality with respect to  $\{B_n\}$  if for every  $g \in L^1$  and  $g \geq 0$ , there exists a constant  $M$  so that for all  $f \in L^1$ , and  $\epsilon > 0$ ,

$$\mu_g\{x \in X : \sup_n R_n(f, g) > \epsilon\} \leq \frac{M}{\epsilon} \int f d\mu$$

where  $d\mu_g = g d\mu$ . We say  $G$  admits the ratio maximal inequality with respect to  $\{B_n\}$  if every action of it has the ratio maximal inequality.

We notice that for probability space, the ratio maximal inequality can be reduced to the ordinary maximal inequality since we can take  $g$  to be constant 1.

The Besicovitch covering property only describes the property of the metric, while the ratio maximal inequality shows the well behavior of the whole analytic structure on the space. Thus, it is not apparently that they are actually equivalent. It is known to us before that the Besicovitch covering property implies the ratio maximal inequality, and Hochman showed the converse also holds for countable groups and proved the following theorem.

**Theorem 4.5.** *Let  $G$  be a countable group and  $\{B_r\} \subseteq G$  an increasing sequence of symmetric sets with  $\cap B_r = \{e\}$ . Then  $G$  has the ratio maximal inequality if and only if  $\{B_r\}$  has the Besicovitch property.*

For our purpose, we want the same equivalence to hold for locally compact Hausdorff groups generally, and make the following claim.

**Theorem 4.6.** *Let  $G$  be a locally compact and  $\sigma$ -compact Hausdorff amenable group and  $\{B_r\} \subseteq G$  an Følner sequence with  $\cap B_r = \{e\}$ . Then  $G$  has the ratio maximal inequality along  $\{B_r\}$  if and only if  $\{B_r\}$  has the Besicovitch covering property.*

*Proof.* Suppose the Besicovitch covering property does not hold for  $\{B_r\}$ . We consider the group acting on itself by left translation. For  $\epsilon > 0$  and  $f, g \in L^1$  with  $\int f d\mu \neq 0$ , let

$$C(f, g) = \frac{\mu_g\{x \in G : \sup_n R_n(f, g)(x) > \epsilon\}}{\int f d\mu}$$

where  $d\mu_g = g \cdot d\mu$  and  $\mu$  is the left Haar measure on  $G$ . As the Besicovitch covering property does not hold for  $\{B_r\}$ , for every  $C > 0$  there exists a  $s > 0$  and  $A, B$  finite subsets of  $G$  such that

$$\frac{|A|}{|B|} < \frac{s}{C},$$

and for any  $g \in A \cup B$ ,

$$\frac{|A \cap B_n g|}{|B \cap B_n g|} > s$$

for some  $n \in \mathbb{N}$ . Then for all  $x \in B$ ,

$$\sup_n \frac{|\{z \in B_n : z \cdot x \in A\}|}{|\{z \in B_n : z \cdot x \in B\}|} > s.$$

Let  $f = \chi_A$  and  $g = \chi_B$ , then

$$\mu_g(\{z \in G : \sup_n \frac{\int_{B_n} f(w \cdot z) d\mu(w)}{\int_{B_n} g(w \cdot z) d\mu(w)} > s\}) \geq \mu_g(B),$$

and

$$\mu_g(B) = |B| > \frac{C}{s}|A| = \frac{C}{s} \int f d\mu.$$

Since  $C$  is arbitrary, this implies that the maximal inequality does not hold for this action and thus  $G$  does not admit the ratio maximal inequality.

Suppose the Besicovitch covering property holds with multiplicity  $C$ . Let  $G \curvearrowright X$  be an ergodic action, and  $f, g \in L^1$  with  $\int g > 0$ . Define

$$T_n(f, g)(h, x) = R_n(\hat{T}^h f, \hat{T}^h g)(x)$$

for all  $x \in X$  and  $h \in G$ . Set

$$S_k(f, g)(h, x) = \sup_{0 \leq n \leq k} T_n(|f|, g)(h, x),$$

and

$$S(f, g)(h, x) = \sup_{n \geq 0} T_n(|f|, g)(h, x).$$

Then  $S_k(f, g) \leq S_l(f, g)$  for all  $k < l$ , and  $\lim_{k \rightarrow \infty} S_k(f, g) = S(f, g)$  pointwisely. Let  $\epsilon > 0$ , and consider  $A_{\epsilon, k} = \{(h, x) : S_k(f, g)(h, x) > \epsilon\}$ . We can observe that since  $\cap B_n = \{e\}$ ,

$$\int_{A_{\epsilon, k}} \hat{T}^s g(x) dm_G(s) d\mu_X(x) \geq m_G(B_k) \int_E g d\mu_X,$$

where  $E = \{x : (e, x) \in A_{\epsilon, k}\} = \{x : \sup_n R_n(|f|, g)(x) > \epsilon\}$ .

On the other hand, we define

$$T^*(f, g)(h, x) = \sup_{n \geq 0} T_n(f, g)(h, x)$$

and

$$T_k^*(f, g)(h, x) = \sup_{0 \leq n \leq k} T_n(f, g)(h, x).$$

Then  $T_k^*(f, g) \leq T_l^*(f, g)$  for all  $k < l$ , and  $\lim_{k \rightarrow \infty} T_k^*(f, g) = T^*(f, g)$  pointwisely. Fix  $x$ , let  $E^x = \{h \in G : T_k^*(f, g)(h, x) > \epsilon\}$ . Then for each  $h \in E^x$ , there exists  $n_h \leq k$  so that  $T_{n_h}^*(\hat{T}^{n_h} f, \hat{T}^{n_h} g)(x) > \epsilon$ . Consider the family  $\{hB_{n_h}\}_{h \in E^x}$ , since Besicovitch covering property holds, there is a subfamily  $\{h_i B_{n_{h_i}}\}_{i \in \mathcal{I}}$  such that

$$\chi_{E^x} \leq \sum_{i \in \mathcal{I}} \chi_{h_i B_{n_{h_i}}} \leq C.$$

Then,

$$\begin{aligned} \int_{A_{\epsilon, k}} \hat{T}^s g(x) dm_G(s) d\mu_X(x) &\leq \int \int_{E^x} \hat{T}^h g(x) dm_G(h) d\mu_X(x) \\ &\leq \frac{C}{\epsilon} \int \int |f| dm_G d\mu_X \\ &= \frac{C}{\epsilon} \|f\|_1. \end{aligned}$$

Hence, let  $k \rightarrow \infty$ , we get

$$\int_E g d\mu \leq \frac{C}{\epsilon} \|f\|_1.$$

□

In fact, by a similar method, one can show that if the Besicovitch covering property does not hold, then the ratio maximal inequality also fails for any free action of  $G$ .

The Besicovitch covering property is a rare condition for metric spaces and is sensitive to the local property of the metric. The following lemma gives a characterization of the Besicovitch covering property when the metric is doubling.

**Proposition 4.7.** *Let  $(X, d)$  be a doubling metric space, i.e., there exists a constant  $M > 0$  so that for any  $x \in X$  and  $r > 0$ , there exists at most  $M$  points  $x_1, x_2, \dots, x_{M'}$  so that  $B(x, r) \subseteq \bigcup_{i=1}^{M'} B(x_i, \frac{r}{2})$ . Then  $(M, d)$  admits the Besicovitch covering property if and only if there exists a constant  $N$  so that for any collection of balls  $B = \{B_i = B(c_i, r_i)\}_{i \in I}$  such that  $\bigcap_{i \in I} B_i \neq \emptyset$  and  $x_i \notin B_j$  for  $i \neq j$ , the cordiality of  $B$ ,  $|I| \leq N$ .*

The later condition may also be called the weak Besicovitch covering property. As the name suggested, spaces with Besicovitch covering property will share this property and there are examples of non-doubling metric spaces where the weak Besicovitch covering property is satisfied but the Besicovitch covering property is not. With this proposition, we can think of the Besicovitch covering property as describing the roundness of the balls in a space. A good example for this intuition is that in  $\mathbb{R}^2$ , we may not have infinite many balls covering the origin with each center not in other balls. In fact, we can prove that  $\mathbb{R}^n$  with the usual metric admits the Besicovitch covering property in this way.

## 5. EXAMPLE: NON-SINGULAR ACTIONS OF HEISENBERG GROUP

In the last section, we discussed the ratio maximal inequality. One of the application of it is proving the ratio ergodic theorems. A general method for proving such theorem is proving the convergence of ratio ergodic mean for a dense set of  $L^1$ , and then use the ratio maximal inequality to control the measure of the exceptional set. To illustrate this method, we use the example of the Heisenberg group.

$$\mathcal{H} = \left\{ \begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} : x, y, z \in \mathbb{R} \right\}$$

Our goal is to prove the ratio ergodic theorem for  $\mathcal{H}$ . To shorten our writing, we first do the following identification. As sets,  $\mathcal{H} = \mathbb{R}^3$  via the bijection

$$\begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \mapsto (x, y, z).$$

Then the multiplication on  $\mathcal{H}$

$$\begin{bmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & x' & z' \\ 0 & 1 & y' \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & x+x' & z+z'+xy' \\ 0 & 1 & y+y' \\ 0 & 0 & 1 \end{bmatrix}$$

can be translated to the following binary operation on  $\mathbb{R}^3$

$$(x, y, z)(x', y', z') = (x+x', y+y', z+z'+xy').$$

A natural metric for  $\mathcal{H}$  is the Carnot-Carathéodory metric, which is defined as the minimal time for connecting two points with curve of derivative in the eigenspace with eigenvalue 1. Its explicit formula is

$$d_{C-C}((x, y, z), (x', y', z')) = \max\{\sqrt{(x-x')^2 + (y-y')^2}, \sqrt{|z-z'|}\}.$$

The unfortunate fact proved by Rigot is that  $\mathcal{H}$  with such metric does not satisfy the Besicovitch covering property. Thus, by the previous theorem, the ratio maximal inequality would not hold for the balls

centered at  $I_3$  with radius  $n \in \mathbb{N}$ , and the method we introduced does not apply.

However,  $\mathcal{H}$  does admit a homogeneous metric with which it satisfies the Besicovitch covering property, as defined below.

**Definition 5.1.** For each  $\lambda > 0$ , let  $\delta_\lambda : \mathcal{H} \rightarrow \mathcal{H}$  be defined as

$$\delta_\lambda(x, y, z) = (\lambda x, \lambda y, \lambda^2 z).$$

For  $p, q \in \mathcal{H}$ , let

$$d(p, q) = \inf\{r > 0 : \delta_{r^{-1}}(pq^{-1}) \in \mathbf{B}\},$$

where  $\mathbf{B} = \{(x, y, z) \in \mathbb{R}^3 : \sqrt{x^2 + y^2 + z^2} \leq 1\}$ .

Donne and Rigot proved in 2004 that the distance function defined above is indeed a homogeneous metric on  $\mathcal{H}$ , and the Besicovitch covering property holds on  $\mathcal{H}$  with this metric. In fact, their result is much stronger: they proved that the definition above could be generalized to the  $n$ -dimensional Heisenberg group  $\mathcal{H}^n$  and we can also replace  $\mathbf{B}$  by any closed ball centered at the origin of  $\mathbb{R}^{2n+1}$  with any radius  $\alpha > 0$ . The corresponding metric would be a homogeneous metric and  $\mathcal{H}^n$  equipped with such metric will satisfy the Besicovitch covering property. Thus, we could expect the results in this section can be generalized to  $\mathcal{H}^n$ .

Let  $B_n = \{p \in \mathcal{H} : d(p, I_3) \leq n\}$  for all  $n \in \mathbb{N}$ . Then,  $\{B_n\}$  satisfies the Besicovitch covering property. In fact, it is shown by Jarrett that it is also a Følner sequence. Thus, we have the ratio maximal inequality for  $\mathcal{H}$  along  $\{B_n\}$ .

With this Følner sequence, the ratio ergodic sum make sense and we expect the following theorem holds.

**Theorem 5.2.** *Let  $\{T^u\}_{u \in \mathcal{H}}$  be a free, non-singular ergodic action on a standard probability space  $(X, \mu)$ . Let  $\nu$  be a left Haar measure on  $\mathcal{H}$ . Let  $B_n = \{u \in \mathcal{H} : \|u\|_{\mathcal{H}} \leq n\}$ . Then for  $f, g \in L^1(\mu)$  with  $\int g \neq 0$  and  $g > 0$ ,*

$$\frac{\int_{B_n} \hat{T}^u f d\nu}{\int_{B_n} \hat{T}^u g d\nu} \rightarrow \frac{\int f d\mu}{\int g d\mu}$$

almost everywhere.

Following the generic method, in order to prove this theorem, we need to firstly construct a dense subset in  $L^1(\mu)$ . The candidate for the dense set is

$$S = \{c + h - \hat{g}h : c \in \mathbb{R}, g \in \mathcal{H}, h \in L^\infty\}.$$

**Lemma 5.3.** *S is dense in  $L^1(\mu)$ .*

*Proof.* Suppose for contradiction that S is not dense in  $L^1$ , then by Hahn Banach theorem, there exists  $f \in L^\infty \setminus \{0\}$  such that for all  $s \in S$ ,  $\int s f d\mu = 0$ . Then for all  $h \in L^\infty$  and  $g \in \mathcal{H}$ ,

$$\int (h - \hat{g}h) f d\mu = 0.$$

Then,

$$\int h(x) f(x) d\mu(x) = \int h(x) f(g^{-1}x) d\mu(x)$$

. Since  $L^\infty$  dense in  $L^1$  and  $h \in L^\infty$  arbitrary,  $f(x) = f(\hat{g}x)$  for all  $g \in \mathcal{H}$ . By ergodicity, f is constant almost everywhere. Then  $\int c f d\mu = 0$  for  $c \in \mathbb{R}$ . Hence,  $f = 0$  a.e., which is a contradiction.  $\square$

**Lemma 5.4.** *Let  $\{T^u\}_{u \in \mathcal{H}}$  be a free, non-singular ergodic action on a standard probability space  $(X, \mu)$ . Let  $\nu$  be a left Haar measure on  $\mathcal{H}$ . Let  $B_n = \{u \in \mathcal{H} : \|u\|_{\mathcal{H}} \leq n\}$ . Then for  $f \in S$ ,*

$$\frac{\int_{B_n} \hat{T}^u f d\nu}{\int_{B_n} \hat{T}^u 1 d\nu} \rightarrow \int f d\mu$$

almost everywhere.

*Proof.* Let  $f = c + h - \hat{g}h \in S$ , then  $\int f d\mu = c$ , and

$$\frac{\int_{B_n} \hat{T}^u f d\nu}{\int_{B_n} \hat{T}^u 1 d\nu} = \frac{\int_{B_n} \hat{T}^u (c + h - \hat{g}h) d\nu}{\int_{B_n} \hat{T}^u 1 d\nu} = c + \frac{\int_{B_n \Delta g B_n} \hat{T}^u h d\nu}{\int_{B_n} \hat{T}^u 1 d\nu}.$$

Since  $h \in L^\infty$ , it suffices to show that for all  $g \in \mathcal{H}$ ,

$$\frac{\int_{B_n \Delta g B_n} \omega_a d\nu(a)}{\int_{B_n} \omega_a d\nu(a)} \rightarrow 0 \text{ a.s.,}$$



where  $\omega_a = \frac{a_* d\mu}{d\mu}$  is the Lebesgue derivative. This follows from the fact that  $\{B_n\}_{n \in \mathbb{N}}$  is a følner sequence, which implies that for all  $g \in \mathcal{H}$ ,

$$\frac{\nu(B_n \Delta g B_n)}{\nu(B_n)} \rightarrow 0,$$

and the doubling property and Besicovitch covering property of the metric.  $\square$

With these two lemmas, we can start to prove theorem 5.2.

*Proof.* First observe that

$$R_n(f, g) = \frac{R_n(f, 1)}{R_n(g, 1)}$$

for all  $n, f$  and  $g$ , and that the theorem follows if it holds in the case  $g \equiv 1$ . Thus, without loss of generality, we may assume that  $g \equiv 1$ . For  $f \in L^1$ , there is a sequence  $\{f_i\}$  in  $S$  converging to  $f$  in  $L^1$ . Suppose for all  $i$ ,

$$f_i = c_i + h_i - \hat{g}_i h_i$$

for some  $c_i \in \mathbb{R}, g_i \in \mathcal{H}, h_i \in L^\infty$ . Then for all  $i$ ,  $\int f_i d\mu = \int c_i d\mu = c_i$ . Thus,  $\lim_{i \rightarrow \infty} c_i = \int f d\mu$ . Then we apply the ratio maximal inequality to  $f - f_i$  and constant function 1. Let  $\epsilon > 0$ , since  $\mu_1 = \mu$ , we have that there exists  $M$  such that for all  $i$ ,

$$\mu(\{x \in X : \sup_n \frac{\int_{B_n} (f - f_i)(ux) d\nu(u)}{\int_{B_n} 1 d\nu(u)} > \epsilon\}) \leq \frac{M}{\epsilon} (\int f d\mu - c_i).$$

Thus,

$$\mu(\{x \in X : \limsup_{n \rightarrow \infty} |R_n(f, 1) - c_i| > 2\epsilon\}) \leq \frac{2M}{\epsilon} (\int f d\mu - c_i).$$

Then there exists  $N \in \mathbb{N}$  such that for all  $i > N$ ,  $|\int f d\mu - c_i| < \epsilon^2$ . For such  $i$ ,

$$\mu(\{x \in X : \limsup_{n \rightarrow \infty} |R_n(f, 1) - \int f d\mu| > 2\epsilon + \epsilon^2\}) \leq 2M\epsilon.$$

Hence,  $R_n(f, 1) \rightarrow \int f d\mu$  almost everywhere, which completes the proof.  $\square$

Balls centered at origin with respect to the Carnot-Carathéodory metric does not satisfy the Besicovitch covering property. Thus, this standard method for proving ratio maximal inequality does not generalize to that case. However, there are examples where the maximal fails but the ergodic theorems hold. For this specific example, we have neither a counter example of functions  $f, g \in L^1$  where ratio ergodic theorem fails, nor new methods for proving it.

The ergodic theorem for  $\mathbb{R}^n$  has helped us understanding the action

$$\left\{ \begin{bmatrix} 1 & z \\ 0 & 1 \end{bmatrix} : z \in \mathbb{R} \right\} \curvearrowright SL_2(\mathbb{R})/\Gamma$$

where  $\Gamma$  is a discrete, finite generated subgroup with infinite co-volume. In the same way, this theorem would help us understanding the orbits of the Heisenberg group acting on a homogeneous space with infinite volume.

## REFERENCES

- [1] Mariá E. Becker. *A ratio ergodic theorem for groups of measure-preserving transformations*. Illinois J. Math., 27(4):562–570, 1983.
- [2] M. Bachir Bekka and Matthias Mayer. *Ergodic Theory and Topological Dynamics of Group Actions on Homogeneous Spaces*. Cambridge University Press, 2000.
- [3] R. V. Chacon and D. S. Ornstein. *A general ergodic theorem*. Illinois J. Math. 4 (1960), no. 2, 153–160, 2009.
- [4] Enrico Le Donne and Séverine Rigot. *Besicovitch Covering Property for homogeneous distances in the Heisenberg groups*. J. Eur. Math. Soc. (JEMS) 19 (2017), no. 5, 1589-1617, 2017.
- [5] Manfred Einsiedler and Thomas Ward. *Ergodic Theory, with a view towards Number Theory*. Springer-Verlag London, 2011.
- [6] Michael Hochman. *A ratio ergodic theorem for multiparameter non-singular actions*. J. Eur. Math. Soc. (JEMS), 12(2):365–383, 2010.
- [7] Kieran Jarrett. *An ergodic theorem for non-singular actions of the Heisenberg groups*. arXiv:1702.04157v1 [math.DS] ,2017.
- [8] David Kerr and Hanfeng Li. *Ergodic Theory, independence and Dichotomies*. Springer International Publishing, 2006.
- [9] Elon Lindenstrauss. *Invariant measures and arithmetic quantum unique ergodicity*. Ann. of Math. (2), 163(1):165–219, 2006.
- [10] Elon Lindenstrauss. *Pointwise theorems for amenable groups*. Invent. Math., 146(2):259–295, 2001.
- [11] Séverine Rigot. *Counter example to the Besicovitch covering property for some Carnot groups equipped with their Carnot-Carétheodory metric*. Math. Z. 248 (2004), no. 4, 827-848.