Basic Ergodic Theory

C. S. Aravinda and Vishesh S. Bhat

1 Introduction

These notes are based on the course of six lectures given by the first named author at the well-run workshop organised at IIT-Delhi in the month of December, 2017. The lectures were intended to be self-contained covering some basic facts in ergodic theory including a discussion of the Birkhoff ergodic theorem which, in a sense, heralded the beginning of ergodic theory. Since the audience mainly consisted of graduate students with different mathematical backgrounds, the lectures began with a quick recap of the construction of the Lebesgue measure in $\mathbb{R}$ and progressed gradually to a discussion of more general measures. After setting up the groundwork on measure preserving transformations and flows on measure spaces, the notion of ergodicity was introduced.

Following a brief look at a couple of illustrative examples of dynamical systems, the focus shifted to a discussion of one of the early interesting examples of an ergodic system, namely the geodesic flow on closed surfaces of constant negative curvature. This necessitated a working recapitulation of the geometry of the upper-half plane with respect to the hyperbolic metric, the lectures culminated with a sketch of a proof, due to Eberhard Hopf, of the ergodicity of the geodesic flow in this setting.

This author acknowledges the support of the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.01-0520.

C. S. Aravinda (✉)
TIFR - Centre for Applicable Mathematics (TIFR-CAM), Bengaluru, India
e-mail: aravinda@math.tifrbng.res.in

V. S. Bhat
Mechanics and Materials Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Japan
e-mail: vishesh.bhat@oist.jp

A. Nagar et al. (eds.), *Elements of Dynamical Systems*, Texts and Readings
in Mathematics 79, https://doi.org/10.1007/978-981-16-7962-9_3

The notes, naturally, reflect the dynamics that the lectures carried and also include some historical titbits in an attempt to capture the significance of the exciting developments, that have shaped this field of study.

The first named author would like to record his deep gratitude to the organisers of this extremely well-run workshop, and to Nikita Agarwal who cheerfully conducted the afternoon tutorials at the workshop with great energy and lot of prior planning. Both the authors thank the efficient editors of this volume for their invitation to script the sketchy lecture notes into a coherent narrative, and the anonymous referees whose careful comments as well as suggestions to add a few explanatory lines at a couple of places helped weed out the several inadvertent typos and in improving the readability. The authors take full responsibility for any errors that may still remain despite their sincere efforts to make these notes error free.

2 Measure Theoretic Preliminaries

This section seeks to develop some rudimentary aspects of measure, starting with the illuminating case of the Lebesgue measure on the real line. Finding the measure of a set means to get a certain estimation of its size. A finite set could be measured by its cardinality, whereas what distinguishes an infinite set from a finite set is its intriguing property of being in bijective correspondence with a proper subset of itself. This begs the question as to how one would determine the size of an infinite subset of the real line $\mathbb{R}$?

For a subset which is an interval $I = (a, b) \subset \mathbb{R}$, its length $|I|$, namely $b - a$ seems a natural and a reasonable estimation of its size. In fact, the seminal investigation that Henri Lebesgue undertook culminating with the description of the so called Lebesgue measure, by exploiting the notion of length, appears in his fundamental paper of 1904 [10].

The basic idea of the Lebesgue measure on $\mathbb{R}$ stems from an effort to adapt the notion of length for an arbitrary subset of $\mathbb{R}$. This turns out to be a very profitable enterprise, as building on finer and subtle variants of this notion, allows one to describe a whole family of s-dimensional Hausdorff measures for each $s \in (0, 1]$; in turn giving rise to the notion of Hausdorff dimension of a given subset. We shall quickly uncover the main facets in this section, particularly mentioning the succinct and elegant work of Caratheodory [4].

We begin by first recalling the notion of outer measure.

Definition 2.1 If $A \subseteq \mathbb{R}$, the *(Lebesgue) outer measure* of A is

$$\mu^*(A) = \inf \left\{ \sum_{k=1}^{\infty} |I_k| \;:\; A \subseteq \bigcup_{k=1}^{\infty} I_k, \text{ where } (I_k)_{k=1}^{\infty} \text{ is} \right.$$
$$\left. \text{a collection of open intervals} \right\}.$$

The completeness property of the reals ensures that if at least one of the members of the above set is finite, then μ^* will be a finite non-negative real number. If no such finite number exists, then the outer measure of A is said to be infinite.

Definition 2.2 If $A \subseteq \mathbb{R}$ and $h \in \mathbb{R}$, the *translate* of A by h is

$$A + h = \{x + h \; : \; x \in A\}.$$

The outer measure on $\mathbb{R}$ exhibits the following properties which can easily be derived from first principles.

Theorem 2.3 *This theorem features the basic properties of outer measure on* $\mathbb{R}$.

1. *(Non-negativity)* $0 \le \mu^*(A) \le +\infty$.
2. *(Monotonicity)* $A \subseteq B \Longrightarrow \mu^*(A) \le \mu^*(B)$.
3. *(Countable subadditivity)* $A \subseteq \bigcup_{n=1}^{\infty} A_n \Longrightarrow \mu^*(A) \le \sum_{n=1}^{\infty} \mu^*(A_n)$.
4. *(Translation invariance)* $\mu^*(A + h) = \mu^*(A)$.
5. $\mu^*(A) = |A|$, *the length of A, if A is an interval.*

While the above mentioned properties inherently follow from the definition; one other natural and desirable property is to expect that the outer measures of two disjoint sets A and B add up to the outer measure of their disjoint union $A \cup B$. This expectation lies at the heart of our discussion and, in a sense, the real essence of the theory lies in understanding this rather innocuous requirement.

A moment's reflection on what the finite additivity property ensures, can be gathered from the following. If $\{A_i\}$, $i = 1, \dots, \infty$ is a countable collection of pairwise disjoint subsets of $\mathbb{R}$, then

$$\sum_{i=1}^{n} \mu^*(A_i) = \mu^*\left(\bigcup_{i=1}^{n} A_i\right) \le \mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} \mu^*(A_i).$$

Taking limits as $n \to \infty$ on both sides results in the countable additivity of the outer measure.

But, the outer measure μ^* defined above has a singular shortcoming in that it is not finitely additive! One way to see this fact, a posteriori, is to glean from Vitali's construction in 1905 [12], of a non-measurable subset of $\mathbb{R}$. Recall that Vitali exhibited a proper non-empty subset C of $\mathbb{R}$, taking rational translates of which, one obtains a countable collection of pairwise disjoint subsets of $\mathbb{R}$. It is on this collection, that the outer measure μ^* cannot be countably additive. In particular, there are disjoint subsets A and B of $\mathbb{R}$ such that $\mu^*(A \cup B) \neq \mu^*(A) + \mu^*(B)$.

In other words, there are subsets X and O of $\mathbb{R}$ such that for the partition by O of X, into disjoint subsets $X \cap O$ and $X \cap O^c$, one has

$$\mu^*(X) \neq \mu^*(X \cap O) + \mu^*(X \cap O^c).$$

To see this, consider disjoint sets A and B and take $X = A \cup B$ and $O = A$. Therefore, $\mu^*(X) = \mu^*(A \cup B) \neq \mu^*(A) + \mu^*(B) = \mu^*(X \cap O) + \mu^*(X \cap O^c)$.

Consequently, one looks at the collection, $\mathfrak{M}$, of all those sets $E \subseteq \mathbb{R}$ such that

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c), \ \forall A \subseteq \mathbb{R}. \tag{1}$$

On this collection, $\mathfrak{M}$, the outer measure μ^* is countably additive. The collection $\mathfrak{M}$, which includes open intervals, constitutes a σ-algebra, and the outer measure restricted to $\mathfrak{M}$ is called the Lebesgue measure on $\mathfrak{M}$. The expression (1) is termed as the Caratheodory criterion and naturally leads to the definition of a (Lebesgue) measurable set. The next two definitions make this observation precise.

Definition 2.4 A family of subsets, $\mathfrak{M}$ of a set X is said to be a *σ-algebra* if the following hold:

1. $X \in \mathfrak{M}$;
2. $A \in \mathfrak{M} \implies A^c \in \mathfrak{M}$;
3. $\{A_i\}_{i=1}^{\infty} \in \mathfrak{M} \implies \bigcup\limits_{i=1}^{\infty} A_i \in \mathfrak{M}$.

Definition 2.5 A set $E \subseteq \mathbb{R}$ is said to be *Lebesgue measurable* or *measurable* if the Caratheodory criterion (1) holds with respect to E.

In light of the preceding definitions, the conclusions of the next proposition can be deduced using properties of the outer measure given in Theorem 2.3.

Proposition 2.6

1. *If I is an interval, then $I \in \mathfrak{M}$ and $\mu^*(I) = |I|$.*
2. *If $A \in \mathfrak{M}$, then $A^c \in \mathfrak{M}$.*
3. *If $A, B \in \mathfrak{M}$, then $A \cup B,\ A \cap B \in \mathfrak{M}$.*
4. *If pairwise disjoint sets $A_1, A_2, \dots, A_N \in \mathfrak{M}$ and $E \subseteq \mathbb{R}$, then*

$$\mu^*\left(E \cap \left(\bigcup_{k=1}^{N} A_k\right)\right) = \sum_{k=1}^{N} \mu^*(E \cap A_k).$$

5. *(Countable additivity or σ-additivity) If $\{A_n\}_{n=1}^{\infty}$ is any sequence of measurable sets, then $\bigcap\limits_{n=1}^{\infty} A_n$ and $\bigcup\limits_{n=1}^{\infty} A_n$ are also measurable. Further, if $\{A_n\}_{n=1}^{\infty}$ is a sequence of pairwise disjoint measurable sets, then $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{M}$ and*

$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu^*(A_n).$$

Definition 2.7 Suppose $A \in \mathfrak{M}$. Then its *(Lebesgue) measure*, $\mu(A)$ is defined to be its outer measure: $\mu(A) = \mu^*(A)$.

Remark 2.8

- The reason for the need of two different concepts is that both have their disadvantages.
- μ is an additive measure, but is not defined for all subsets of $\mathbb{R}$.
- μ^* is defined for all subsets of $\mathbb{R}$, but is not additive, as demonstrated by Vitali's construction.

A more restricted class of Lebesgue measurable sets are the Borel measurable sets.

Definition 2.9 If X is any topological space (in this case $\mathbb{R}$), then the σ-algebra, $\mathfrak{B}$ generated by the class of open sets in X (resp. open intervals in $\mathbb{R}$) are called the *Borel sets* of X (resp. $\mathbb{R}$).

Remark 2.10 It can be easily shown that the Borel σ-algebra for $\mathbb{R}$ includes the half-open intervals such as $[a, b)$ as well as closed intervals and further that every Borel set is (Lebesgue) measurable.

The important properties of the outer measure μ^* continue to hold on replacing μ^* by μ whenever $A \in \mathfrak{M}$.

Theorem 2.11 *Here, we enlist some additional properties of measurable sets.*

1. *Continuity: Suppose $A_1 \supseteq A_2 \supseteq A_3 \cdots$ and $B_1 \subseteq B_2 \subseteq B_3 \cdots$ are sequences of measurable sets, and $\mu(A_1) < \infty$. Then,*

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n) \quad \textit{and} \quad \mu\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mu(B_n).$$

2. *Approximation: If $A \in \mathfrak{M}$, and $\mu(A) < \infty$, then for all $\epsilon > 0$ there exists a bounded closed set B and an open set C such that $B \subseteq A \subseteq C$, and $\mu(C \cap B^c) < \epsilon$.*

The previously sketched discussion of the construction of the Lebesgue measure on $\mathbb{R}$, starting from the notion of outer measure is, in a sense, a proto for the construction of measures more generally on complete metric spaces. In the setting of a metric space X together with the distance function d, one starts with the notion of a 'metric outer measure' which estimates the size of a subset A, by considering covers of A by a countable number of open balls; then, using radii of open balls, one considers an appropriate measure of their sizes to analogously replace lengths of intervals.

We shall elaborate more on this later when discussing Hausdorff measures, but will now proceed to a discussion of measures in general.

Definition 2.12 A *measure space* is a triple $(X, \mathfrak{M}, \mu)$, where X is any set, $\mathfrak{M}$ is a σ-algebra of measurable sets and μ is a σ-additive measure.

A measurable space is just the pair $(X, \mathfrak{M})$ with no specification about the measure. The concept of σ-finiteness is another desirable property for a measure to possess.

Definition 2.13 A measure space $(X, \mathfrak{M}, \mu)$ is said to be *σ-finite* if X can be written as a countable union of measurable sets of finite measure i.e., $X = \bigcup\limits_{n=1}^{\infty} A_n$ with $\mu(A_n) < +\infty$, for all n. μ is then said to be a *σ-finite measure*.

Definition 2.14 Given a measure space $(X, \mathfrak{M}, \mu)$, a set $A \subset X$ is said to be a *null set* or a set of measure zero if there exists a set $A_1 \in \mathfrak{M}$ so that $A \subseteq A_1$ and $\mu(A_1) = 0$. Furthermore, two sets $A_1, A_2 \subset X$ are said to be *equivalent mod* 0 if their symmetric difference, $A_1 \Delta A_2$ i.e., $(A_1 \setminus A_2) \cup (A_2 \setminus A_1)$ has measure zero and this is denoted as $A_1 \equiv A_2 \pmod 0$.

Remark 2.15

1. It should be noted in this context that not every measurable set is a Borel set. In fact, it is possible to construct sets of measure zero which are Lebesgue measurable but not Borel measurable. Thus, the Lebesgue measure serves as a completion of the Borel measure.
2. Note that a more formal definition of a complete measure is as follows: Given a measure space $(X, \mathfrak{M}, \mu)$, μ is complete if and only if for any $N \in \mathfrak{M}$ where $\mu(N) = 0$, $E \subseteq N$ implies $E \in \mathfrak{M}$. The Lebesgue measure is complete precisely in the above sense.

Another example of a finite measure space is the probability space which is the space of choice for ergodic theory. For a measure space $(X, \mathfrak{M}, \mu)$, if $\mu(X) = 1$, then X is a said to be probability space and μ a probability measure.

Measure zero sets are very useful in characterising properties in measure theory.

Definition 2.16 A property P of points of a set $A \subseteq X$ is said to hold *almost everywhere* (a.e.) if the set of points of A which do not satisfy P form a set of measure zero.

2.1 Measurable Functions and Transformations

We now move on to the notion of a measurable function which closely mirrors the topological definition of a continuous function. The first definition is formulated in the setting of general measure spaces.

Definition 2.17 (*Measurable functions or transformations*) If $(X, \mathfrak{M})$ and $(Y, \mathfrak{N})$ are two measurable spaces, then a map $f : X \longrightarrow Y$ is *measurable* if $f^{-1}(A)$ is measurable i.e., $f^{-1}(A) \in \mathfrak{M}$ for every $A \in \mathfrak{N}$. Further, if X and Y are topological spaces, then $f : X \longrightarrow Y$ is said to be *(Borel-) measurable* if it is measurable with respect to the Borel σ-algebras of X and Y.

Remark 2.18 The above definition implies that every continuous function is (Borel-) measurable.

In the sequel, we use the extended real line $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ with the usual conventions. To keep things simple, in the remaining part of this section, we restrict ourselves to extended real-valued functions defined on $\mathbb{R}$ (equipped with the usual Lebesgue measure), unless otherwise explicitly stated, although the statements hold in the more general setting of complete measure spaces.

Remark 2.19 In particular, if $f : (\mathbb{R}, \mathfrak{L}) \longrightarrow (\bar{\mathbb{R}}, \mathfrak{B})$, where $\mathfrak{L}$ is the Lebesgue σ-algebra, and f is measurable as in the Definition 2.17, then f is said to be Lebesgue measurable.

For extended real-valued functions f, g, denote

$$(f \wedge g)(x) = \min\{f(x), g(x)\}, \quad (f \vee g)(x) = \max\{f(x), g(x)\}.$$

Proposition 2.20 *Measurable functions satisfy the following notable properties:*

1. *Suppose f, g are measurable functions and $c \in \mathbb{R}$, then $cf, f + g, fg, |f|, f \wedge g, f \vee g$ are measurable.*
2. *Suppose $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions and $\lim_{n\to\infty} f_n(x) = f(x)$, then f is measurable.*
3. *Suppose $\{f_n\}_{n=1}^{\infty}$ is a sequence of measurable functions. Let $g(x) = \inf\{f_n(x)\}$ and $h(x) = \sup\{f_n(x)\}$. Then g and h are measurable.*

Definition 2.21 The *indicator function* of a set $A \subseteq \mathbb{R}$ is the function

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

Definition 2.22 A *simple function* is a function of the form

$$f = a_1\chi_{A_1} + \cdots + a_n\chi_{A_n} \quad \text{where } a_i \in \mathbb{R},\ A_i \in \mathfrak{M} \text{ and } \mu(A_i) < \infty.$$

Definition 2.23 The *integral* of a simple function $f = \sum_{i=1}^{n} a_i\chi_{A_i}$ is

$$\int f\,d\mu = \int_{\mathbb{R}} f\,d\mu = \sum_{i=1}^{n} a_i\mu(A_i).$$

Definition 2.24 (*Integral of nonnegative measurable functions*) If $f : \mathbb{R} \longrightarrow \mathbb{R}$ is a nonnegative measurable function, then its integral is

$$\int f\,d\mu = \sup\left\{\int g\,d\mu \ :\ g \text{ is a simple function such that } 0 \leq g \leq f\right\}.$$

Proposition 2.25 *If f, g are nonnegative measurable functions and $a > 0$, then*

$$\int af\,d\mu = a\int f\,d\mu, \quad \int (f+g)\,d\mu = \int f\,d\mu + \int g\,d\mu.$$

Moreover, if $f \leq g$, then

$$\int f\,d\mu \leq \int g\,d\mu.$$

This additivity property will allow us to extend the definition of integration to functions that change sign.

Definition 2.26 For an extended real-valued function f, define functions

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) > 0, \\ 0 & \text{if } f(x) \leq 0; \end{cases} \qquad f^-(x) = \begin{cases} -f(x) & \text{if } f(x) < 0, \\ 0 & \text{if } f(x) \geq 0. \end{cases}$$

Note that f^+ and f^- are nonnegative. They are measurable if f is, and $f = f^+ - f^-$, $|f| = f^+ + f^-$.

Definition 2.27 A measurable function is *integrable* if $\int |f|\,d\mu < +\infty$.

Definition 2.28 If f is an integrable function, its *integral* is

$$\int f\,d\mu = \int f^+\,d\mu - \int f^-\,d\mu.$$

Definition 2.29 The *limit supremum* of a sequence is the least upper bound of the set of all subsequential limits of the sequence. That is,

$$\limsup_{n\to\infty} a_n := \lim_{n\to\infty} (\sup\{a_m : m \geq n\}) = \inf_{n\geq 0}\left(\sup_{m\geq n} a_m\right).$$

Similarly, we define

$$\liminf_{n\to\infty} a_n := \lim_{n\to\infty} (\inf\{a_m : m \geq n\}).$$

Theorem 2.30 (Fundamental convergence theorems) *Here, we record the fundamental convergence theorems in analysis, that we use in the sequel.*

1. *(Lebesgue's dominated convergence theorem) Suppose $(f_n)_{n=1}^{\infty}$ is a sequence of measurable functions and $\lim_{n\to\infty} f_n(x) = f(x)$, for all $x \in \mathbb{R}$, and $|f_n(x)| \leq g(x)$ for all $n \in \mathbb{N}$, $x \in \mathbb{R}$ where g is an integrable function. Then,*

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

2. *(Monotone Convergence Theorem) Suppose $(f_n)_{n=1}^{\infty}$ is a non-decreasing sequence of non-negative measurable functions $0 \leq f_1 \leq f_2 \leq \cdots$. Let $f(x) = \lim_{n\to\infty} f_n(x)$.*

Then,

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

3. *(Fatou's Lemma) If $(f_n)_{n=1}^{\infty}$ is a sequence of nonnegative measurable functions, then*

$$\int \liminf_{n\to\infty} f_n\, d\mu \leq \liminf_{n\to\infty}\int f_n\, d\mu.$$

Definition 2.31 Two functions f and g are said to be equal *almost everywhere*, written $f = g$ a.e., if $\{x \;:\; f(x) \neq g(x)\}$ is a set of measure zero.

Proposition 2.32 *If f is a function on a Lebesgue measurable set E and $g = f$ a.e., then g is Lebesgue measurable if and only if f is Lebesgue measurable.*

Definition 2.33 Consider the set of all integrable functions on $\mathbb{R}$. The function space L^1 is the set of all equivalence classes of integrable functions on $\mathbb{R}$, where we set $f \simeq g$ if $f = g$ a.e. The L^1 *norm* is given by

$$\|f\|_1 := \int |f|\, d\mu.$$

Theorem 2.34 *L^1 is complete, i.e., given a Cauchy sequence $\{f_n\}_{n=1}^{\infty}$ in L^1, there exists $f \in L^1$ such that $\lim\limits_{n\to\infty} \|f_n - f\|_1 = 0$.*

Generalising the L^1 notion to functions on arbitrary complete measure spaces, we have the following definition.

Definition 2.35 Let $(X, \mathfrak{M}, \mu)$ be a complete measure space and $f : X \longrightarrow \bar{\mathbb{R}}$ be a measurable function, then for each integer $p \geq 1$, we say that $f \in L^p(\mu)$ if

$$\int_X |f|^p\, d\mu < \infty.$$

For any such $f \in L^p(\mu)$, we may define the L^p*-norm* as

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

Identifying the functions whose values agree a.e. allows for defining a metric on the space $L^p(\mu)$ by means of the L^p-norm. We treat $L^p(\mu)$ as the set of equivalence classes of functions which coincide a.e.. Thus, $L^p(\mu)$ becomes a Banach space for $1 \leq p < \infty$. In particular, $L^2(\mu)$ is a Hilbert space with the inner product defined by

$$\langle f, g\rangle := \int\limits_X fg\, d\mu.$$

Definition 2.36 $f : X \longrightarrow \mathbb{R}$ is said to be *compactly supported* if the closure of the set of points in X where the value of f is non-zero, is a compact subset of X.

Notation 2.37 We denote the set of all compactly supported (real-valued) continuous functions on X as $C_c(X)$.

Theorem 2.38 (Lusin's Theorem) *If X is a locally compact Hausdorff topological space and if $f : X \longrightarrow \bar{\mathbb{R}}$ is a measurable function such that $f(x) = 0$, for all $x \notin A \subset X$, where $\mu(A) < \infty$, then given $\epsilon > 0$, there exists a $g \in C_c(X)$ so that*

$$\mu\left(\{x \ : \ f(x) \neq g(x)\}\right) < \epsilon.$$

Theorem 2.39 *For $1 \leq p < \infty$, $C_c(X)$ is dense in $L^p(\mu)$.*

Definition 2.40 Let $(X, \mathfrak{M})$ be a measurable space and $\mu, \ \nu : X \longrightarrow [0, \infty)$ be two measures on $\mathfrak{M}$. We say that μ is *absolutely continuous* with respect to ν if $A \in \mathfrak{M}$ and $\nu(A) = 0$ implies $\mu(A) = 0$. This is denoted as $\mu \ll \nu$.

Theorem 2.41 (Radon–Nikodym) *If $(X, \mathfrak{M}, \nu)$ is a σ-finite measure space, then $\mu \ll \nu$ if and only if there exists a function $f \in L^1(\nu)$ such that*

$$\mu(A) = \int\limits_A f\, d\nu \ \ \textit{for every } A \in \mathfrak{M}.$$

The function f is unique a.e. with respect to ν and is written as $\frac{d\mu}{d\nu}$, called the Radon-Nikodym derivative *of μ w.r.t ν.*

2.2 Hausdorff Measures

In this subsection, we outline the notion of more general measures called Hausdorff measures that subsume the Lebesgue measure. It is assumed that (X, d) is a non-empty metric space. The notion of Hausdorff dimension of a subset $A \subset X$ arises from the construction of Hausdorff measures [6].

Definition 2.42 A function μ defined on $\mathcal{P}(X)$ is called a *metric outer measure* if it satisfies the following:

1. $\mu^*(A) \geq 0$, for all $A \in \mathcal{P}(X)$;
2. $\mu^*(\varnothing) = 0$;
3. (Monotonicity) $A_1 \subseteq A_2 \implies \mu^*(A_1) \leq \mu^*(A_2)$;

4. (Countable subadditivity) if $\{A_n\}_{n=1}^{\infty}$ is a countable collection of members of $\mathcal{P}(X)$, then $\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} \mu^*(A_n)$;
5. if $A_1, A_2 \in \mathcal{P}(X)$ with $d(A_1, A_2) > 0$, then $\mu^*(A_1 \cup A_2) = \mu^*(A_1) + \mu^*(A_2)$.

A familiar example of such an outer measure is the Lebesgue outer measure discussed in the earlier sections. Before defining the Hausdorff measure, we remark that as in the case of $\mathbb{R}$, a subset E of a space X is said to be measurable if

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c), \quad \forall A \in \mathcal{P}(X).$$

The class of measurable sets in X evidently form a σ-algebra $\mathfrak{P}$ so that μ when restricted to $\mathfrak{P}$, is countably additive and thus a measure in the usual sense. We henceforth use the usual μ notation for the measure.

Definition 2.43 Given a metric space (X, d) and $A \subset X$, the diameter of A is given as $\delta(A) := \sup\{d(x, y) \ : \ x, y \in A\}$.

Let (X, d) be a metric space and let $\alpha(> 0) \in \mathbb{R}$. Let $A \subset X$. Given $\epsilon > 0$, consider

$$H_\alpha^\epsilon(A) = \inf\left\{\sum_{k=1}^{\infty} \delta(A_k)^\alpha \ : \ A \subseteq \bigcup_{k=1}^{\infty} A_k \text{ where } \delta(A_k) < \epsilon \ \forall k\right\},$$

the infimum being taken over all countable covers of the set A whose members have diameter less than ϵ. Note that if $\epsilon_1 < \epsilon$, then $H_\alpha^{\epsilon_1}(A) \ge H_\alpha^\epsilon(A)$. Therefore, $\lim_{\epsilon \to 0} H_\alpha^\epsilon(A)$ exists, though it may be infinite, and we write $H_\alpha(A) = \lim_{\epsilon \to 0} H_\alpha^\epsilon(A)$.

Theorem 2.44 *For each $\alpha > 0$, H_α is a metric outer measure on X called the Hausdorff outer measure of dimension α and when restricted to the σ-algebra of measurable sets, is called the Hausdorff measure of dimension α on X.*

Note that if $\alpha = 0$, then H_α is merely, the counting measure.

Theorem 2.45 *(i) If $H_\alpha(A) < \infty$, then $H_\beta(A) = 0$ for $\beta > \alpha$.*
(ii) If $H_\alpha(A) > 0$, then $H_\beta(A) = \infty$ for $\beta < \alpha$.

Proof It is easy to see that *(i)* and *(ii)* are equivalent. Therefore, we prove *(i)*. Suppose $A = \bigcup_{k=1}^{\infty} A_k$, with $\delta(A_k) < \epsilon$. If $\beta > \alpha$, then

$$H_\beta^\epsilon(A) \le \sum_{k=1}^{\infty} \delta(A_k)^\beta \le \epsilon^{\beta-\alpha} \sum_{k=1}^{\infty} \delta(A_k)^\alpha.$$

That is, $H_\beta^\epsilon(A) \le \epsilon^{\beta-\alpha} H_\alpha^\epsilon(A)$. Letting $\epsilon \to 0$, we see that $H_\beta(A) = 0$ if $H_\alpha(A) < \infty$. □

As a consequence of the above theorem, for $A \subset X$, there exists $d \in \mathbb{R}$ such that

$$\begin{cases} H_m(A) = 0 & \text{if } m > d, \\ H_m(A) = \infty & \text{if } m < d. \end{cases}$$

The d, obtained as above, is called the *Hausdorff dimension* of the set A, denoted by $\mathcal{H}_{dim}(A)$.

Example 2.46

1. If A is any countable set then, $\mathcal{H}_{dim}(A) = 0$.
2. If $X = \mathbb{R}$ and $\alpha = 1$, then it is straightforward to check that H_1 is the Lebesgue measure.
3. The Cantor ternary set is an example of an uncountable set of zero Lebesgue measure, as opposed to countable sets which are also of Lebesgue measure zero. It can be shown that its Hausdorff dimension is $\dfrac{\ln 2}{\ln 3}$.

If $X = \mathbb{R}^n$, $n > 1$, then H_n is not the same as the Lebesgue measure, but is *comparable* to it, a fact elucidated in the next theorem.

Theorem 2.47 *Let $A \subset \mathbb{R}^n$.*

1. *Then, there exists positive constants C_1 and C_2 depending only on the dimension n such that*
$$C_1 H_n(A) \leq \lambda(A) \leq C_2 H_n(A),$$
for $A \subset \mathbb{R}^n$, λ being the Lebesgue measure on $\mathbb{R}^n$.
2. *If $\alpha > n$, then $H_\alpha(A) = 0$, for every $A \subset \mathbb{R}^n$.*

3 Recurrence and Ergodic Theorems

Let $(X, \mathfrak{M}, \mu)$ be a measure space. A transformation $T : X \longrightarrow X$ is said to be a *measurable transformation* (with respect to μ) if the inverse image of every μ-measurable set is μ-measurable. And a μ-measurable transformation T of X into itself is said to be *measure preserving* if $\mu(T^{-1}(E)) = \mu(E)$ for every μ-measurable subset E of X.

Example 3.1

1. Let $X = [0, 1)$ and λ be the Lebesgue measure on X. Let $c \in X$ be any point. Then the transformation $T : X \longrightarrow X$ defined by $T(x) = x + c$ (mod1) is measure preserving.

2. Let $X = [0, 1)$ and λ be the Lebesgue measure on X. Define $T : X \longrightarrow X$ as
$$T(x) = \begin{cases} 2x & \text{for } 0 \leq x < \frac{1}{2} \\ 2x - 1 & \text{for } \frac{1}{2} \leq x < 1. \end{cases}$$
It can be easily verified that T as defined above is a measure preserving transformation.
3. Given $a = (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ where $\mathbb{R}^n$ is equipped with the usual Lebesgue measure. The affine transformation $T : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ defined as $T(x) = x + a$ is invertible and measure preserving.

In the context of ergodic theory, a measurable space $(X, \mathfrak{M}, \mu)$ equipped with a measure preserving transformation T constitutes a *dynamical system* denoted by $(X, \mathfrak{M}, \mu, T)$.

3.1 Recurrence

In the sequel, we assume that $(X, \mathfrak{M}, \mu)$ is a probability space i.e. $\mu(X) = 1$. Given a measure preserving transformation T on a measure space $(X, \mathfrak{M}, \mu)$, T is said to be recurrent if for any given set of positive measure $A \subset X$, almost all points of A return to A after at most finitely many iterations of T.

Theorem 3.2 (Poincare recurrence theorem) *Let $(X, \mathfrak{M}, \mu)$ be a probability space and $T : X \longrightarrow X$ be a measure preserving transformation. Given $A \in \mathfrak{M}$, let A_0 be the set of points $x \in A$ such that $T^n(x) \in A$ for infinitely many $n \geq 0$. Then $A_0 \in \mathfrak{M}$ and $\mu(A_0) = \mu(A)$.*

Proof Let
$$C_n = \left\{x \in A \ : \ T^k(x) \notin A \ \forall k \geq n\right\}.$$

Therefore $A_0 = A \setminus \bigcup\limits_{n=1}^{\infty} C_n$. In order to prove the theorem, it is enough to show that

1. $C_n \in \mathfrak{M}$ and
2. $\mu(C_n) = 0$ for every $n \geq 1$.

1. Now, $C_n = A \setminus \bigcup\limits_{k \geq n} T^{-k}(A)$. Since $T^{-k}(A) \in \mathfrak{M}$ for every $k \geq 1$, we see that $C_n \in \mathfrak{M}$.
2. Also,

$$C_n \subset \bigcup_{k \geq 0} T^{-k}(A) \setminus \bigcup_{k \geq n} T^{-k}(A)$$
$$\implies \mu(C_n) \leq \mu\left(\bigcup_{k \geq 0} T^{-k}(A)\right) - \mu\left(\bigcup_{k \geq n} T^{-k}(A)\right).$$

Now, observe that $\bigcup\limits_{k \geq n} T^{-k}(A) = T^{-n}\left(\bigcup\limits_{k \geq 0} T^{-k}(A)\right)$. Since T is measure preserving, this implies

$$\mu\left(\bigcup_{k \geq 0} T^{-k}(A)\right) = \mu\left(\bigcup_{k \geq n} T^{-k}(A)\right).$$

Therefore $\mu(C_n) = 0$. □

3.2 Birkhoff Ergodic Theorem and the Notion of Ergodicity

Let $(X, \mathfrak{M}, \mu)$ be a probability space and $T : X \longrightarrow X$ be a measure preserving transformation. Let $E \in \mathfrak{M}$. Given $x \in X$, one would like to ask with what frequency do the elements of the set $\{x, Tx, T^2x, \ldots\}$ lie in the set E?

Clearly $T^i x \in E$ if and only if $\chi_E(T^i x) = 1$; therefore the number of elements of $\{x, Tx, T^2x, \ldots, T^{n-1}x\}$ in E is $\sum\limits_{k=0}^{n-1} \chi_E(T^k x)$ or the relative number of $\{x, Tx, \ldots, T^{n-1}x\}$ in E is $\dfrac{1}{n}\sum\limits_{k=0}^{n-1} \chi_E(T^k x)$.

Around the turn of the century, the work of Boltzmann and Gibbs on statistical mechanics raised a mathematical problem which can be stated as follows: Given a measure preserving transformation T of a probability space and an integrable function $f : X \longrightarrow \mathbb{R}$, find conditions under which

$$\lim_{n \to \infty} \frac{f(x) + f(Tx) + \cdots + f(T^{n-1}x)}{n}$$

exists and is constant almost everywhere.

In 1931 [3], Birkhoff proved that for any T and f, the above limit exists almost everywhere. From this, he concluded that a necessary and sufficient condition for its value to be constant almost everywhere, is that there exist no set $A \in \mathfrak{M}$ such that $0 < \mu(A) < 1$ and $T^{-1}A = A$. As we will see later, the fact that this limit is constant easily implies that it is equal to the integral of f over X. Transformations T which satisfy this condition are called *ergodic* and ergodic theory is essentially the study of such transformations. The Birkhoff Ergodic theorem is the first fundamental result that sets the tone for much of what follows.

Theorem 3.3 (Birkhoff Ergodic Theorem) *Let $(X, \mathfrak{M}, \mu)$ be a probability space and $T : X \longrightarrow X$ be a measure preserving transformation. If $f \in L^1(\mu)$ then the limit*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) = \widetilde{f}(x),$$

exists for almost every point $x \in X$, $\widetilde{f} \in L^1(\mu)$ *and* $\widetilde{f} \circ T = \widetilde{f}$ *almost everywhere. Furthermore,*

$$\int_X \widetilde{f}\, d\mu \;=\; \int_X f\, d\mu.$$

If f is any measurable function, let $g(x) = f(Tx)$. Since T is measurable, the function g is measurable so that, writing $g(x) = Uf(x)$, the transformation U assigns to each measurable function f, a measurable function g. Clearly, U is linear and g is non-negative if f is so. Moreover, we have:

Theorem 3.4 *If* $1 \le p \le \infty$ *and* $\|f\|_p$ *denotes the* L^p*-norm of* f, *then* $\|g\|_p = \|f\|_p$ *for* $g = Uf$.

Proof Let $E \in \mathfrak{M}$ and $f = \chi_E$. Then $g = Uf = f(Tx) = \chi_{T^{-1}(E)}$. Therefore,

$$\|g\|_p^p \;=\; \mu(T^{-1}(E)) \;=\; \mu(E) \;=\; \|f\|_p^p\,.$$

It follows that $\|g\|_p = \|f\|_p$ for every non-negative simple function. If f is any non-negative measurable function, there exists a sequence of simple, non-negative measurable functions $\{s_n\}_{n=1}^{\infty}$ such that $s_n \to f$, as $n \to \infty$, with $s_1 \le s_2 \le \cdots \le f$. Now, since $t_n = Us_n$ is also an increasing sequence of simple functions, converging to g, monotone convergence theorem implies that

$$\|g\|_p \;=\; \lim_{n\to\infty} \|t_n\|_p \;=\; \lim_{n\to\infty} \|s_n\|_p \;=\; \|f\|_p\,.$$

The general case of f now follows by writing $f = f^+ - f^-$ and applying the above conclusion to f^+ and f^- separately. □

In particular, if $f \in L^2(\mu)$ we have showed that $g(x) = Uf(x) = f(Tx)$ is also in $L^2(\mu)$ and that $\|g\|_2 = \|f\|_2$. In other words, U is an isometric transformation of the Hilbert space $L^2(\mu)$ into itself.

If, in addition, T is invertible (i.e., there exists a measure preserving transformation $S : X \longrightarrow X$ such that $ST = TS = Id_X$) and if V is the isometric transformation in $L^2(\mu)$ corresponding to its inverse, then $UV = VU$ is the identity transformation in $L^2(\mu)$. Therefore, the range of V is the whole of $L^2(\mu)$; in other words, U is a unitary transformation in $L^2(\mu)$ and V is its inverse. Thus, an invertible measure preserving transformation on a measure space (X, μ) induces an invertible unitary transformation in the Hilbert space $L^2(\mu)$.

Therefore, in so far as it concerns functions $f \in L^2(\mu)$, the existence of the limit of the averages is reduced to the problem of existence of the limit as $n \to \infty$ of $\dfrac{1}{n}\displaystyle\sum_{k=0}^{n-1} U^k f(x)$, where U is an isometric transformation in the Hilbert space $L^2(\mu)$. Precisely, this convergence, known as the mean ergodic theorem, was proven by J. von Neumann in 1932 [13].

Theorem 3.5 (Mean ergodic theorem) *If U is an isometric transformation in an arbitrary Hilbert space H and if P is the orthogonal projection on the closed linear subspace of all $f \in H$ satisfying $Uf = f$, then $\frac{1}{n}\sum_{k=0}^{n-1} U^k f$ converges in norm as $n \to \infty$ to Pf for all $f \in H$.*

We will skip a proof of this and prove the more general Birkhoff ergodic theorem (BET, for short). We prove the first part of the BET and prove the more general L^p version of the second part as a corollary. The key step in the proof of BET is itself a useful lemma known as the Maximal ergodic theorem.

Lemma 3.6 (Maximal ergodic theorem) *Given $f \in L^1(\mu)$, put*

$$E(f) \;=\; \left\{ x \;:\; \max_{n\geq 0}\left(\sum_{k=0}^{n-1} f(T^k x)\right) > 0 \right\}.$$

Then $\int_{E(f)} f \, d\mu \geq 0$.

Proof Define

$$\begin{aligned} f_0 &:= 0, \\ f_n &:= f + f\circ T + f \circ T^2 + \cdots + f \circ T^{n-1} \\ &= f + Uf + U^2 f + \cdots + U^{n-1} f. \end{aligned}$$

Let $F_n = \max_{0\leq k\leq n} f_k$. Therefore

$$E(f) = \bigcup_{n-1}^{\infty} \{x \;:\; F_n(x) > 0\} = \bigcup_{n-1}^{\infty} E_n.$$

Clearly, $F_n \in L^1(\mu)$ and, for $0 \leq k \leq n$, we have $F_n \geq f_k$. Therefore $UF_n \geq Uf_k$ because $U : L^1(\mu) \longrightarrow L^1(\mu)$ is a positive linear operator (i.e., $f \geq 0$ implies $Uf \geq 0$) and hence,

$$UF_n + f \geq Uf_k + f = f_{k+1}.$$

In other words,

$$UF_n + f \geq \max_{1\leq k\leq n} f_k(x) \;=\; \max_{0\leq k\leq n} f_k(x) \;=\; F_n(x) \text{ when } F_n(x) \;>\; 0.$$

That is, $f \geq F_n - UF_n$ on $\{x \;:\; F_n(x) > 0\} = E_n$. Therefore,

$$\begin{aligned}\int_{E_n} f\,d\mu &\geq \int_{E_n} F_n\,d\mu - \int_{E_n} UF_n\,d\mu\\ &= \int_X F_n\,d\mu - \int_{E_n} UF_n\,d\mu\\ &\geq \int_X F_n\,d\mu - \int_X UF_n\,d\mu\\ &= 0.\end{aligned}$$

The second equality above holds because $F_n = 0$ on $X \setminus E_n$, the third inequality holds because $F_n \geq 0$ implies $UF_n \geq 0$ and the last equality holds because $\|U\| = 1$. Finally, since $E_1 \subseteq E_2 \subseteq \cdots$, we have that $E_n \to E(f)$ and we are done. □

Corollary 3.7 *If $A \subset E(f)$, $A \in \mathfrak{M}$ and $T^{-1}A = A$, then,*

$$\int_A f\,d\mu \geq 0.$$

Proof Since $T^{-1}A = A$, we see that $E(f\chi_A) = A$. Therefore, the lemma above implies $0 \leq \int_{E(f\chi_A)} f\chi_A\,d\mu = \int_A f\chi_A\,d\mu = \int_A f\,d\mu$. □

Theorem 3.8 *Let $(X, \mathfrak{M}, \mu)$ be a probability space and $T : X \longrightarrow X$ be a measure preserving transformation. If $f \in L^1(\mu)$, then the limit*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$$

exists for almost every point $x \in X$.

Proof For each $\alpha, \ \beta \in \mathbb{R}$ with $\alpha < \beta$, let

$$E_{\alpha,\beta} \ = \ \left\{ x \in X \ : \ \liminf_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) < \alpha < \beta < \limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \right\}.$$

Clearly, $E_{\alpha,\beta} \in \mathfrak{M}$. We will show that $\mu(E_{\alpha,\beta}) = 0$ for each $\alpha, \ \beta$. This would imply that $\bigcup E_{\alpha,\beta}$, where $\alpha, \ \beta \in \mathbb{R}$ such that $\alpha < \beta$, has measure zero and hence the limit exists almost everywhere.

Put $f^*(x) = \sup_{n\geq 1} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$ and $f_*(x) = \inf_{n\geq 1} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$. Therefore,

$$E_{\alpha,\beta} \subset \{x \ : \ f^*(x) > \beta\} = \{x \ : \ (f^* - \beta)(x) > 0\} = E(f - \beta)$$

and $E_{\alpha,\beta} \subset \{x \ : \ f_*(x) < \alpha\}$.

We first show that $E_{\alpha,\beta}$ is T-invariant. That is, we show that $T^{-1}(E_{\alpha,\beta}) = E_{\alpha,\beta}$. Let $a_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$. Then, $\frac{n+1}{n}a_{n+1}(x) - a_n(Tx) = \frac{f(x)}{n}$. Therefore,

$$\limsup_{n\to\infty}(a_{n+1}(x) + \frac{1}{n}a_{n+1}(x) - a_n(Tx)) = \limsup_{n\to\infty} \frac{f(x)}{n}.$$

This implies that $\limsup_{n\to\infty}(a_{n+1}(x) - a_n(Tx)) = 0$. That is, $\limsup_{n\to\infty}(a_{n+1}(x)) = \limsup_{n\to\infty}(a_n(Tx))$. Similarly, $\liminf_{n\to\infty}(a_{n+1}(x)) = \liminf_{n\to\infty}(a_n(Tx))$.

Therefore, $T^{-1}(E_{\alpha,\beta}) = E_{\alpha,\beta}$.

By Corollary 3.7, we get $\int_{E_{\alpha,\beta}}(f - \beta)\, d\mu \geq 0$ or $\int_{E_{\alpha,\beta}} f\, d\mu \geq \beta\mu(E_{\alpha,\beta})$. Now $E_{\alpha,\beta} \subset \{x \ : \ f_*(x) < \alpha\} = \{x \ : \ -f_* > -\alpha\} = \{x \ : \ (-f)^* > -\alpha\}$.

Therefore, by the maximal ergodic theorem 3.6, $\int_{E_{\alpha,\beta}}(-f)\, d\mu \geq -\alpha\mu(E_{\alpha,\beta})$ or $\int_{E_{\alpha,\beta}} f\, d\mu \leq \alpha\mu(E_{\alpha,\beta})$. Thus, $\beta\mu(E_{\alpha,\beta}) \leq \int_{E_{\alpha,\beta}} f\, d\mu \leq \alpha\mu(E_{\alpha,\beta})$.

But $\alpha < \beta$. Therefore, the above inequality holds only if $\mu(E_{\alpha,\beta}) = 0$. □

Corollary 3.9 *(i) If $f \in L^p(\mu)$, $1 \leq p \leq \infty$, the function $\widetilde{f}$ defined by,*

$$\widetilde{f}(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$$

is in $L^p(\mu)$ and satisfies

$$\lim_{n\to\infty} \left\| \widetilde{f} - \frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \right\|_p = 0.$$

(ii) $\widetilde{f}(Tx) = \widetilde{f}(x)$.
(iii) For $f \in L^p(\mu)$, $\int_X \widetilde{f}\, d\mu = \int_X f\, d\mu$.

Proof *(i)* Since X is a probability space, $\mu(X) = 1$. Therefore, $f \in L^1(\mu)$ and $\widetilde{f}(x)$ makes sense. Moreover, $|f| \in L^1(\mu)$ and $\left|\widetilde{f}(x)\right| \leq \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \left|f(T^k x)\right|$ for a.e. x (this limit exists since $|f| \in L^1(\mu)$). That is, $\left|\widetilde{f}(x)\right|^p \leq \lim_{n\to\infty} \left(\frac{1}{n}\sum_{k=0}^{n-1} \left|f(T^k x)\right|\right)^p$. Since $\left|\widetilde{f}\right|^p \geq 0$,

$$\begin{aligned}\|\widetilde{f}\|_p^p = \int_X |\widetilde{f}|^p \, d\mu &= \int_X \lim_{n\to\infty} \left(\frac{1}{n}\sum_{k=0}^{n-1} \left|f(T^k x)\right|\right)^p d\mu \\ &= \int_X \liminf_{n\to\infty} \left(\frac{1}{n}\sum_{k=0}^{n-1} \left|f(T^k x)\right|\right)^p d\mu \\ &\le \liminf_{n\to\infty} \int_X \left(\frac{1}{n}\sum_{k=0}^{n-1} \left|f(T^k x)\right|\right)^p d\mu. \text{ (Fatou's Lemma)}\end{aligned}$$

Now

$$\begin{aligned}\int_X \left(\frac{1}{n}\sum_{k=0}^{n-1} \left|f(T^k x)\right|\right)^p d\mu &= \left\|\frac{1}{n}\sum_{k=0}^{n-1} \left|f(T^k x)\right|\right\|_p^p \\ &\le \left(\frac{1}{n}\sum_{k=0}^{n-1} \left\|f(T^k x)\right\|_p\right)^p \\ &= \left(\frac{1}{n}\sum_{k=0}^{n-1} \|f\|_p\right)^p \text{ (T^k is measure preserving)} \\ &= \|f\|_p^p .\end{aligned}$$

Therefore

$$\|\widetilde{f}\|_p^p \le \liminf_{n\to\infty} \int_X \left(\frac{1}{n}\sum_{k=0}^{n-1} \left|f(T^k x)\right|\right)^p d\mu \le \liminf_{n\to\infty} \|f\|_p^p = \|f\|_p^p < \infty,$$

since $f \in L^p(\mu)$. Therefore $\widetilde{f} \in L^p(\mu)$. □

Definition 3.10 (*Convergence in the L^p-norm*) Consider the case $f \in L^\infty(\mu)$, i.e., $\sup\limits_{x\in X} |f(x)| < \infty$ a.e. Clearly, $f \in L^1(\mu)$ and the sequence of functions

$$\left|\widetilde{f} - \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)\right|^p$$

converges to 0 a.e. Moreover,

$$\left|\widetilde{f}(x)\right| \le \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \left|f(T^k x)\right| \le \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \|f\|_\infty = \|f\|_\infty \text{ -a.e.}$$

Therefore,

$$\left|\widetilde{f}(x)-\frac{1}{n}\sum_{k=0}^{n-1}f(T^kx)\right|^p \leq \left|\|f\|_\infty+\frac{1}{n}\sum_{k=0}^{n-1}\left\|f\circ T^k\right\|_\infty\right|^p \leq \left(2\,\|f\|_\infty\right)^p = \text{constant}.$$

Hence by dominated Convergence theorem,

$$\int_X\left|\widetilde{f}-\frac{1}{n}\sum_{k=0}^{n-1}f(T^kx)\right|^p\ d\mu \to 0\text{ -a.e.}$$

That is, for $f\in L^p(\mu)$, $\lim\limits_{n\to\infty}\left\|\widetilde{f}-\frac{1}{n}\sum_{k=0}^{n-1}f\circ T^k\right\|_p = 0$. Now, let $f\in L^p(\mu)$ and let $\varepsilon>0$. There is an $f_0\in L^\infty(\mu)$ such that $\|f-f_0\|_p\leq\varepsilon/3$ and there exists an $N>0$ such that $\left\|\widetilde{f_0}-\frac{1}{n}\sum_{k=0}^{n-1}f_0\circ T^k\right\|_p\leq\varepsilon/3$ for $n\geq N$.

Then,

$$\left\|\widetilde{f}-\frac{1}{n}\sum_{k=0}^{n-1}f(T^kx)\right\|_p$$
$$\leq \left\|\widetilde{f}-\widetilde{f_0}\right\|_p + \left\|\widetilde{f_0}-\frac{1}{n}\sum_{k=0}^{n-1}f_0(T^kx)\right\|_p + \left\|\frac{1}{n}\sum_{k=0}^{n-1}(f_0-f)(T^kx)\right\|_p.$$

Now, $\widetilde{f}-\widetilde{f_0}=\widetilde{f-f_0}$ and hence,

$$\left\|\widetilde{f}-\widetilde{f_0}\right\|_p = \left\|\widetilde{f-f_0}\right\|_p \leq \|f-f_0\|_p \leq \frac{\varepsilon}{3},$$

and

$$\left\|\frac{1}{n}\sum_{k=0}^{n-1}(f_0-f)(T^kx)\right\|_p \leq \frac{1}{n}\sum_{k=0}^{n-1}\|f_0-f\|_p = \|f_0-f\|_p \leq \frac{\varepsilon}{3}.$$

Therefore, for $n\geq N$,

$$\left\|\widetilde{f}-\frac{1}{n}\sum_{k=0}^{n-1}f(T^kx)\right\|_p < \varepsilon,$$

which implies that

$$\lim_{n\to\infty}\left\|\widetilde{f}-\frac{1}{n}\sum_{k=0}^{n-1}f(T^kx)\right\|_p = 0.$$

We now prove the remainder of the statements in Corollary 3.9.

(ii)

$$\begin{aligned}\widetilde{f}(Tx) &= \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} f(T^k(Tx))\\ &= \lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=0}^{n} f(T^k x) - \frac{1}{n}f(x)\right)\\ &= \lim_{n\to\infty}\frac{n+1}{n}\frac{1}{n+1}\sum_{k=0}^{n} f(T^k x) - \lim_{n\to\infty}\frac{1}{n}f(x)\\ &= \lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^{n} f(T^k x)\\ &= \widetilde{f}(x).\end{aligned}$$

(iii) If $f \in L^p(\mu)$, note that by *(ii)*, the sequence $\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x)$ converges to $\widetilde{f}$ in $L^1(\mu)$. Hence,

$$\int_X \widetilde{f}\,d\mu = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\int_X f(T^k x)\,d\mu = \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\int_X f\,d\mu = \int_X f\,d\mu.$$

□

In Birkhoff Ergodic Theorem, suppose the limit $\widetilde{f}(x) = c$, where c is a constant. Then,

$$\int_X f\,d\mu = \int_X \widetilde{f}\,d\mu = c\mu(X).$$

That is,

$$c = \widetilde{f}(x) = \frac{1}{\mu(X)}\int_X f\,d\mu.$$

In other words, we see that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) = \frac{1}{\mu(X)}\int_X f\,d\mu.$$

The left hand side is the *time average* of f and the right hand side is the *space average* of f. This is what the physicists call the *ergodic hypothesis*, (the equality of the time and space averages of f).

Proposition 3.11 *Let T be an invertible measure preserving transformation of X, $f \in L^1(\mu)$ and let*

$$f_n^+(x) \;=\; \frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) \quad f_n^-(x) \;=\; \frac{1}{n}\sum_{k=0}^{n-1} f(T^{-k}x).$$

Then, $\widetilde{f}^+ = \lim\limits_{n\to\infty} f_n^+$ and $\widetilde{f}^- = \lim\limits_{n\to\infty} f_n^-$ exist and are equal almost everywhere, i.e., $\widetilde{f}^+ = \widetilde{f}^-$ -a.e.

Proof We first observe that

$$f_N^+ \circ T^{-(N-1)}(x) = \frac{1}{N}\sum_{k=0}^{N-1} f(T^k(T^{-(N-1)}x)) \;=\; \frac{1}{N}\sum_{k=0}^{N-1} f(T^{-k}x) = f_N^-(x).$$

Also, since $\widetilde{f_N^+} \circ T = \widetilde{f_N^+}$ and $\widetilde{f_N^-} \circ T^{-1} = \widetilde{f_N^-}$, we get $\widetilde{f_N^+} \circ T^{-1} = \widetilde{f_N^+}$ and hence, $\widetilde{f_N^+} \circ T^{-k} = \widetilde{f_N^+}$ for all $k \in \mathbb{N}$. Therefore,

$$\begin{aligned}
\widetilde{f_N^+}(x) = \widetilde{f_N^+} \circ T^{-(N-1)}(x) &= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f_N^+(T^k(T^{-(N-1)}x)) \\
&= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f_N^+ \circ T^{-(N-1)}(T^k x) \\
&= \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} f_N^-(T^k x) \\
&= \widetilde{f_N^-}(x).
\end{aligned}$$

Hence $\widetilde{f}^+ = \lim_{n\to\infty} \widetilde{f_n^+} = \lim\limits_{n\to\infty} \widetilde{f_n^-} = \widetilde{f}^-$ (this holds because, $f_n \to f$ implies $\widetilde{f_n} \to \widetilde{f}$). □

Definition 3.12

1. A *measurable flow* in a measure space $(X, \mathfrak{M}, \mu)$ is a map $\tau : X \times \mathbb{R} \longrightarrow X$ that satisfies the following two conditions:

   (a) τ is measurable with respect to the product measure $\mu \times \lambda$ on $X \times \mathbb{R}$ and the measure μ on X. Here, λ is the Lebesgue measure on $\mathbb{R}$.

   (b) For $t \in \mathbb{R}$, the maps $\tau_t(x) := \tau(x, t)$ form a one-parameter group of transformations of X to itself with $\tau_0 =$ identity on X and $\tau_{t+s} = \tau_t \circ \tau_s$ for $t, s \in \mathbb{R}$.

2. A measurable flow τ_t is *measure preserving* or is μ-invariant if $\mu(\tau_t A) = \mu(A)$ for every $t \in \mathbb{R}$ and every $A \in \mathfrak{M}$.

Remark 3.13 If τ_t is a measure preserving flow on a finite measure space $(X, \mathfrak{M}, \mu)$ and if $f \in L^1(\mu)$, then the limits

$$f^{+} = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(\tau_t x)\, dt \quad \text{and} \quad f^{-} = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(\tau_{-t} x)\, dt$$

exist and are equal for μ - a.e x.

Proof Let $F(x) = \int_0^1 f(\tau_t x)\, dt$. Since f and τ are measurable, $f \circ \tau(x,t) = f(\tau_t x)$ is measurable and by Fubini theorem $F(x) = \int_0^1 f(\tau_t x)\, dt$ is μ-measurable and is in $L^1(\mu)$ since $f \in L^1(\mu)$.

Now

$$\lim_{n\to\infty} \frac{1}{n} \int_0^n f(\tau_t x)\, dt = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\tau_1^k(x))$$

(where $\tau_1(x) = \tau(x, 1) : X \times \mathbb{R} \to X$) exists for μ a.e. x by Birkhoff ergodic theorem.

Let

$$\widetilde{f}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} F(\tau_1^k(x)) = \lim_{n\to\infty} \frac{1}{n} \int_0^n f(\tau_t x)\, dt.$$

If $t \in \mathbb{R},\ t > 0$ is such that $n < t < n+1$ for $n \in \mathbb{N} \cup \{0\}$, then

$$\begin{aligned}
\left| \int_0^t f(\tau_t x)\, dt - \int_0^n f(\tau_t x)\, dt \right| &= \left| \int_n^t f(\tau_t x)\, dt \right| \\
&\leq \left| \int_n^{n+1} f(\tau_t x)\, dt \right| \\
&\leq \int_n^{n+1} |f(\tau_t x)|\, dt \\
&= \int_0^1 \left| f(\tau_1^n \circ \tau_t(x)) \right|\, dt \\
&= \int_0^1 |f(\tau_t x)|\, dt,
\end{aligned}$$

where the last equality follows from Theorem 3.4.

Since

$$\frac{1}{n} \int_0^1 |f(\tau_t x)|\, dt \to 0 \ \text{ as } \ n \to \infty,$$

we have

$$\frac{1}{t} \left| \int_n^t f(\tau_t x)\, dt \right| \ \leq \ \frac{1}{n} \left| \int_n^t f(\tau_t x)\, dt \right| \to 0 \ \text{ as } \ n \to \infty.$$

Since $t \to \infty$ as $n \to \infty$, we have

$$\frac{1}{t} \int_n^t f(\tau_t x)\, dt \to 0 \ \text{ as } \ t \to \infty,$$

and hence

$$\frac{1}{t}\int_0^t f(\tau_t x)\,dt \to \widetilde{f}(x) \text{ as } t \to \infty.$$

Now the remark follows by virtue of the preceding Proposition 3.11. □

Definition 3.14

1. Let $(X, \mathfrak{M}, \mu)$ be a probability space. If $A \in \mathfrak{M}$ and T is a measure preserving transformation of X, then A is said to be T-*invariant* if $\mu(T^{-1}A\Delta A) = 0$. A is said to be strictly T-invariant if $T^{-1}A = A$.
2. A measurable function $f : X \longrightarrow \mathbb{R}$ is T-*invariant* if $\mu\left(\{x : f(Tx) \neq f(x)\}\right) = 0$. f is *strictly* T-*invariant* if $f(Tx) = f(x)$ for all x.

The next two observations seek to bridge the divide between T-invariant and strictly T-invariant sets (or functions).

Lemma 3.15

1. *If* $A \in \mathfrak{M}$ *is a* T-*invariant set, then there is a strictly* T-*invariant set* A_∞ *such that* $\mu\,(A_\infty) = \mu(A)$.
2. *If* f *is a* T-*invariant function, then there is a strictly* T-*invariant function* $\bar{f}$ *such that* $\bar{f}(x) = f(x)$ *-a.e.*

Proof

1. Let

$$A_\infty = \bigcap_{n=0}^{\infty}\bigcup_{i=n}^{\infty} T^{-i}A.$$

It is easy to check that $A_\infty \in \mathfrak{M}$, $T^{-1}A_\infty = A_\infty$ and $\mu\,(A_\infty) = \mu(A)$.

2. Let

$$A_f = \left\{x : f(T^k x) = f(x) \text{ for some } k \in \mathbb{N}\right\}.$$

Clearly, A_f has measure 1, since the set $\{x : f(Tx) = f(x)\}$ is contained in A_f. Let

$$\bar{f}(x) = \begin{cases} f(y) & \text{if } y = T^k(x) \in A_f \text{ for some } k \in \mathbb{N} \\ 0 & \text{otherwise.} \end{cases}$$

It is easy to see that $\bar{f}$ is well-defined, strictly T-invariant and $\bar{f} = f$-a.e. □

Let us find out the conditions under which the limit $\widetilde{f}(x)$ in the ergodic theorem is constant a.e. for every $f \in L^1(\mu)$.

Suppose $\widetilde{f}(x)$ =constant -a.e. for every $f \in L^1(\mu)$. Let $A \in \mathfrak{M}$ be a strictly T-invariant set and let χ_A be the characteristic function of A.

The ergodic theorem for χ_A implies $\int_X \widetilde{\chi}_A\,d\mu = \int_X \chi_A\,d\mu = \mu(A)$. Now

$$\widetilde{\chi}_A(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_A(T^k x).$$

Since $A = T^{-1}A$, $Tx \in A$ if and only if $x \in T^{-1}A = A$ or $T^k x \in A$ if and only if $x \in T^{-k}A = A$ for $k \in \mathbb{N}$. Therefore,

$$\widetilde{\chi}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

By assumption, $\widetilde{\chi}_A(x)$ =constant -a.e. Therefore, $\widetilde{\chi}_A = 0$ or 1 -a.e. This implies $\mu(A) = 0$ or 1. That is, every T-invariant set has measure either 0 or 1.

Now, suppose on the contrary that if $A \in \mathfrak{M}$ is T-invariant then $\mu(A) = 0$ or 1. Let $f \in L^1(\mu)$ and let $\widetilde{f}(x)$ be the limit as in the ergodic theorem. By ergodic theorem, $\widetilde{f} \circ T = \widetilde{f}$ -a.e. on X.

Let

$$A(k,n) \;=\; \left\{ x \;:\; \frac{k}{2^n} \leq \widetilde{f}(x) < \frac{k+1}{2^n} \right\} \text{ for } k \in \mathbb{Z},\ n \in \mathbb{N}.$$

Now

$$T^{-1}(A(k,n))\Delta A(k,n) \subset \left\{ x \;:\; \widetilde{f} \circ T(x) \neq \widetilde{f}(x) \right\}.$$

Therefore,

$$\mu(T^{-1}(A(k,n))\Delta A(k,n)) = 0$$

and hence, $A(k,n)$ is a T-invariant set and therefore $\mu(A(k,n)) = 0$ or 1.

Now, for a fixed $n \in \mathbb{N}$, $\bigcup\limits_{k\in\mathbb{Z}} A(k,n) = X$ is a disjoint union. Therefore, for each $n \in \mathbb{N}$, there exists a unique $k_n \in \mathbb{Z}$ such that $\mu(A(k_n,n)) = 1$.

Let $Y = \bigcap\limits_{n=1}^{\infty} A(k_n,n)$. Then $\mu(Y) = 1$ (because $\mu(Y^c) = 0$). Since $\widetilde{f}$ is constant on Y and $\mu(Y) = 1$, $\widetilde{f}$ is constant a.e on X.

Definition 3.16 A measure preserving transformation $T : X \longrightarrow X$, where $(X, \mathfrak{M}, \mu)$ is a probability measure space, is said to be *ergodic* if for every set $A \in \mathfrak{M}$ which is T-invariant, one has $\mu(A) = 0$ or 1.

Indeed we have shown that a measure preserving transformation T is ergodic if and only if every T-invariant function f is constant a.e. on X.

Proposition 3.17 *Let $(X, \mathfrak{M}, \mu)$ be a second countable probability measure space such that every non-empty open subset of X has positive measure. If $T : X \longrightarrow X$ is an ergodic transformation then*

$$\mu\left(\left\{x \;:\; \left\{T^n x \;:\; n \geq 0 \textit{ is dense in } X\right\}\right\}\right) = 1.$$

That is, almost all points in X have dense orbits.

Proof Let $\{U_n\}_{n=1}^{\infty}$ be a basis for X. Let

$$Y \;=\; \left\{x \;:\; \left\{T^n x \;:\; n \geq 0 \text{ is dense in } X\right\}\right\}.$$

Clearly $x \notin Y$ if and only if there is a basic open set U_k such that $x \in \bigcap\limits_{n=0}^{\infty}(X \setminus T^{-n}(U_k)) = P$, say. It is easy to see that $P \subset T^{-1}(P)$. Since T is measure preserving and $P \in \mathfrak{M}$, $\mu(T^{-1}P) = \mu(P)$. Therefore, $T^{-1}P \equiv P \pmod 0$ and hence, P is T-invariant. Also $U_n \cap P = \varnothing$ and since $\mu(U_k) > 0$, we must have $\mu(P) = 0$, which implies $\mu(P^c) = 1$. Ergo, P^c consists of points x whose T-orbits are dense in X. □

Example 3.18 Let $X = [0, 1)$ be equipped with the Lebesgue measure. If $c \in \mathbb{R}$, the map $T_c : X \longrightarrow X$ defined by

$$T_c(x) \;=\; x + c \pmod 1 \;=\; \{x + c\} \;\text{ i.e., fractional part of } x + c.$$

It is clear that T_c preserves the Lebesgue measure, and it is easy to see that if $c \in \mathbb{Q}$, then T_c is periodic and all orbits are finite having same cardinality. Therefore, T_c is not ergodic when c is rational.

Example 3.19 If X is the circle $\mathbb{S} = \{z \in \mathbb{C} \;:\; |z| = 1\}$ with the normalised Lebesgue measure, then $T : \mathbb{S} \longrightarrow \mathbb{S}$ defined as $T(z) = az$ is measure preserving, as can be easily verified. Then T is ergodic iff a is not a root of unity. For, suppose a is a root of unity, i.e., $a^p = 1$ for some $p \neq 0$. Then $f(z) = z^p$. Clearly $f \circ T = f$, but f is not constant a.e. Therefore, T is not ergodic.

Conversely, suppose a is not a root of unity and let $f(z) = \sum\limits_{n=-\infty}^{\infty} b_n z^n$ be its Fourier expansion. Now, $f \circ T = f$ implies $\sum\limits_{n=-\infty}^{\infty} b_n a^n z^n = \sum\limits_{n=-\infty}^{\infty} b_n z^n$. Hence, $b_n(a^n - 1) = 0$. As $a^n \neq 1$, for any $n \neq 0$, we must have $b_n = 0$ for all $n \neq 0$. Consequently, it follows that f is constant a.e. and that T is ergodic. Alternatively, if $a = e^{2\pi icn}$, then T is ergodic whenever c is irrational.

4 Geodesic Flows on Closed Surfaces

Let M be a compact or, more generally, a complete, smooth manifold endowed with a Riemannian metric g, and let SM denote the associated unit tangent bundle. That is,

$$SM \;=\; \{(x, v) \;:\; x \in M,\; v \text{ is a unit tangent vector to } M \text{ at } x\}.$$

For each $t \in \mathbb{R}$, consider the transformation $\phi^t : SM \longrightarrow SM$ defined as follows: Given $(x, v) \in SM$, let γ_v be the unique geodesic in M passing through the point $x \in M$ and with v as its tangent vector at x. Since M is a manifold which is complete,

γ_v is defined on all of $\mathbb{R}$. Moreover, given any two points $p, q \in M$ there exists a geodesic joining p and q that realises the distance between them. Now set

$$\phi^t(x, v) \;=\; (\gamma_v(t), \gamma_v'(t)). \tag{2}$$

It is easy to verify that ϕ^t as defined above for all $t \in \mathbb{R}$ constitutes a 1-parameter group of transformations, called the geodesic flow, and satisfies the following properties:

1. $\phi^t \circ \phi^s = \phi^{t+s} = \phi^{s+t} = \phi^s \circ \phi^t$ and $\phi^0 = \mathrm{Id}|_{SM}$.
2. ϕ^t is measure preserving where the measure under consideration is the Liouville measure given locally by the product of the Riemannian volume [form] on M, (i.e., $\sqrt{\det(g_{ij})}\, dx_1 \wedge \cdots \wedge dx_n$) - also called the Riemannian measure and the usual Lebesgue measure on the unit sphere.

It would be illuminating to look at a simple example of the geodesic flow.

Example 4.1 Suppose $M = \mathbb{S}^2$, the unit 2-sphere, then M admits a metric of constant positive curvature. Since all of its geodesics are great circles, it means that every orbit of the geodesic flow is periodic, and is therefore not ergodic.

Following up on the previous example, the question of ergodicity of the geodesic flow on closed surfaces of constant negative curvature is treated in the sequel.

The Gauss-Bonnet theorem suggests that a compact Riemann surface with genus ≥ 2 admits a Riemannian metric of constant negative curvature.

We shall initially see how to define such a metric on these surfaces. The universal cover of the surface is, in fact, the upper half plane $\mathbb{H}^2$, where $\mathbb{H}^2 = \{z \in \mathbb{C} \;:\; \mathrm{Im}(z) > 0\}$, equipped with the metric $ds = \dfrac{\sqrt{dx^2 + dy^2}}{y}$, which is a metric of constant negative curvature, called the *hyperbolic metric*. Therefore, we first discuss the geometry of the upper half plane.

4.1 Isometries and Geodesics of $\mathbb{H}^2$

Let $\gamma : I \longrightarrow \mathbb{H}^2$ be a piecewise differentiable path parametrised as

$$\gamma(t) \;=\; \left\{z(t) \;=\; x(t) + iy(t) \in \mathbb{H}^2 \;:\; t \in I\right\}, \quad \text{where } I = [0, 1].$$

Then, the hyperbolic length $l(\gamma)$ of the path is given by

$$l(\gamma) \;=\; \int_0^1 \frac{\sqrt{(\frac{dx}{dt})^2 + (\frac{dy}{dt})^2}}{y(t)}\, dt \;=\; \int_0^1 \frac{|\frac{dz}{dt}|}{y(t)}\, dt. \tag{3}$$

The hyperbolic distance $\rho_h(z, w)$ between any two points $z, w \in \mathbb{H}^2$ is given as $\rho_h(z, w) = \inf l(\gamma)$, where the infimum is taken over all piecewise differentiable paths γ joining z and w in $\mathbb{H}^2$.

A natural question is to look at the isometries of $\mathbb{H}^2$; i.e., transformations on $\mathbb{H}^2$ preserving the hyperbolic distance ρ_h defined above. This leads us to a particular group of matrices denoted as $\mathrm{PSL}(2, \mathbb{R})$.

In order to place the elements in $\mathrm{PSL}(2, \mathbb{R})$, we first look at the group of matrices $\mathrm{SL}(2, \mathbb{R})$ consisting of all 2×2 real matrices of the form

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ where } \det(g) = 1. \tag{4}$$

Quite clearly, the above group of matrices assumes a correspondence with the group of all *fractional linear transformations* of $\mathbb{C}$ onto itself of the form

$$\left\{ z \longmapsto \frac{az+b}{cz+d} \; : \; ad - bc = 1; \; a, b, c, d \in \mathbb{R} \right\}$$

with the product of two such transformations being equivalent to the product of two corresponding matrices in $\mathrm{SL}(2, \mathbb{R})$ and the inverse of a given transformation corresponding to the inverse matrix.

However the correspondence is not 1-1, rather any such fractional linear transformation is represented by a pair of matrices $\pm g$. Ergo, the group of all fractional linear transformations, henceforth identified with $\mathrm{PSL}(2, \mathbb{R})$, is isomorphic to $\mathrm{SL}(2, \mathbb{R})/\pm I$, where I is the 2×2 identity matrix. The corresponding identity transformation in $\mathrm{PSL}(2, \mathbb{R})$ will be denoted by Id.

Remark 4.2 Note that $\mathrm{PSL}(2, \mathbb{R})$ contains all fractional linear transformations of the form $z \longmapsto \frac{az+b}{cz+d}$, where $ad - bc = \Delta > 0$, as dividing the numerator and denominator by $\sqrt{\Delta}$ gives a new matrix of determinant 1, but resulting in the same transformation on $\mathbb{H}^2$. In particular, $\mathrm{PSL}(2, \mathbb{R})$ contains transformations of the form $z \longmapsto az + b, \; a, b \in \mathbb{R}, \; a > 0$ and those of the form $z \longmapsto \dfrac{-1}{z}$.

Remark 4.3 $\mathrm{PSL}(2, \mathbb{R})$ acts on $\mathbb{H}^2$ by homeomorphisms. In fact, $\mathrm{PSL}(2, \mathbb{R}) \subset \mathrm{Isom}(\mathbb{H}^2)$, the group of all isometries of $\mathbb{H}^2$ (i.e., transformations of $\mathbb{H}^2$ onto itself preserving the hyperbolic distance on $\mathbb{H}^2$).

Proof Firstly, any transformation of the form $z \longmapsto \dfrac{az+b}{cz+d}$ on $\mathbb{C}$ maps $\mathbb{H}^2$ onto itself. Given any $T \in \mathrm{PSL}(2, \mathbb{R})$, let $w = T(z) = \dfrac{az+b}{cz+d}$. Then,

$$w = \frac{(az+b)(c\bar{z}+d)}{|cz+d|^2} = \frac{ac\,|z|^2 + adz + bc\bar{z} + bd}{|cz+d|^2}.$$

Hence, the imaginary part $\mathrm{Im}(w)$ of w is,

$$\mathrm{Im}(w) \;=\; \frac{w-\overline{w}}{2i} \;=\; \frac{z-\overline{z}}{2i\,|cz+d|^2} \;=\; \frac{\mathrm{Im}(z)}{|cz+d|^2}.$$

Therefore, $\mathrm{Im}(z) > 0 \iff \mathrm{Im}(w) > 0$. As T is continuous and its inverse exists, we conclude that T is a homeomorphism of $\mathbb{H}^2$ onto itself.

To show that $T \in \mathrm{PSL}(2, \mathbb{R})$ is an isometry of $\mathbb{H}^2$ onto itself, we show that if $\gamma : I \longrightarrow \mathbb{H}^2$ is a piecewise differentiable path in $\mathbb{H}^2$, then $l\,(T(\gamma)) = l(\gamma)$. Therefore, suppose $\gamma := z(t) = x(t) + iy(t)$, and $T(\gamma)$ is given by $w(t) = T(z(t)) = u(t) + iv(t)$. Now

$$\frac{dw}{dz} \;=\; \frac{a(cz+d) - c(az+b)}{(cz+d)^2} \;=\; \frac{1}{(cz+d)^2}.$$

Since $v = \dfrac{y}{|cz+d|^2}$, we have $\left|\dfrac{dw}{dz}\right| = \dfrac{v}{y}$. Therefore,

$$\begin{aligned} l(T(\gamma)) &= \int_0^1 \frac{\left|\frac{dw}{dt}\right|}{v(t)}\,dt \;=\; \int_0^1 \frac{\left|\frac{dw}{dz}\frac{dz}{dt}\right|}{v(t)}\,dt \\ &= \int_0^1 \frac{\left|\frac{dw}{dz}\right|\left|\frac{dz}{dt}\right|}{v(t)}\,dt \;=\; \int_0^1 \frac{\left|\frac{dz}{dt}\right|}{y(t)}\,dt \;=\; l(\gamma). \end{aligned}$$

□

It is a fact that isometries take geodesics to geodesics and hence any transformation in PSL(2, $\mathbb{R}$) maps geodesics to geodesics. We now determine the geodesics on the hyperbolic plane.

Theorem 4.4 *The geodesics in $\mathbb{H}^2$ are semicircles and straight lines orthogonal to the real axis.*

Proof Let $z_1, z_2 \in \mathbb{H}^2$. First suppose $z_1 = ia$ and $z_2 = ib$ with $b > a$ which are two points on the imaginary axis. If $\gamma : [0, 1] \longrightarrow \mathbb{H}^2$ is any path joining ia to ib, with $\gamma(t) = x(t) + iy(t)$, then

$$l(\gamma) = \int_0^1 \frac{\sqrt{(\frac{dx}{dt})^2 + (\frac{dy}{dt})^2}}{y(t)}\,dt \ge \int_0^1 \frac{\left|\frac{dy}{dt}\right|}{y(t)}\,dt \ge \int_a^b \frac{dy}{y} \ge \ln\frac{b}{a}.$$

It is easy to verify that the equality in the above expression is realised by the hyperbolic length of the segment of the y-axis joining ia to ib which is of length $\ln\dfrac{b}{a}$ and hence the geodesic joining the points ia and ib is the segment of the imaginary axis between them.

If $z_1, z_2 \in \mathbb{H}^2$ are arbitrary, let L be the unique Euclidean semi-circle or straight line orthogonal to the real axis passing through z_1 and z_2, then there exists a transformation in $\mathrm{PSL}(2,\mathbb{R})$ which maps L into the imaginary axis. The transformation $T(z)=\dfrac{-1}{z-a}$ takes a to ∞ and b to $\dfrac{1}{b-a}$ (>0), and the transformation $S(z)=z-\dfrac{1}{b-a}=z-c$ takes ∞ to ∞ and c to 0. Thus,

$$S\circ T=\begin{pmatrix}1 & -c\\ 0 & 1\end{pmatrix}\begin{pmatrix}0 & -1\\ 1 & -a\end{pmatrix}=\begin{pmatrix}-c & -1-ac\\ 1 & -a\end{pmatrix}$$

is the transformation in $\mathrm{PSL}(2,\mathbb{R})$ that takes (a,b) to $(\infty,0)$. Since each element of $\mathrm{PSL}(2,\mathbb{R})$ is an isometry of $\mathbb{H}^2$ and segments of the imaginary axis are geodesics, we conclude that the geodesic joining z_1 and z_2 is the segment of L joining them. □

Since $\mathrm{PSL}(2,\mathbb{R})$ acts by isometries on $\mathbb{H}^2$, it acts on the unit tangent bundle $S\mathbb{H}^2$ as

$$g(z,\zeta)=(g(z),D_zg(\zeta))=\left(g(z),\frac{1}{(cz+d)^2}\right),$$

where $z\in\mathbb{H}^2$, $\zeta\in T_z\mathbb{H}^2$ such that $\|\zeta\|=1$ and $g=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in\mathrm{PSL}(2,\mathbb{R})$.

Lemma 4.5 *The action of* $\mathrm{PSL}(2,\mathbb{R})$ *on* $S\mathbb{H}^2$ *is transitive and free, i.e., all isotropy groups are trivial.*

Proof Let $z_0=i$ and ζ_0 be the unit tangent vector at z_0 pointing in the positive direction of the imaginary axis. Let $(z,\zeta)\in S\mathbb{H}^2$ and σ be the positive imaginary half axis starting from z_0. Let L be the unique geodesic determined by (z,ζ). Let $g\in\mathrm{PSL}(2,\mathbb{R})$ be the transformation taking σ to L, i.e., $g(\sigma)=L$, with $g(z_0)=z$. Since transformations of $\mathrm{PSL}(2,\mathbb{R})$ have positive determinant, they preserve orientation and hence the condition that $D_{z_0}g(\zeta_0)=\zeta$ forces g to be unique; we will, therefore, denote it by $g_{z\zeta}$. □

Remark 4.6 In the above lemma, taking $(z,\zeta)\in S\mathbb{H}^2$ to $g_{z\zeta}\in\mathrm{PSL}(2,\mathbb{R})$, sets up a bijection F between $S\mathbb{H}^2$ and $\mathrm{PSL}(2,\mathbb{R})$, and is easily seen to be a diffeomorphism.

Let $z_0=i$ and ζ_0 be as in the proof of Lemma 4.5. Given an arbitrary $(z,\zeta)\in S\mathbb{H}^2$, let $g_{z\zeta}$ be the unique element of $\mathrm{PSL}(2,\mathbb{R})$ (which exists by virtue of the lemma) that takes (z_0,ζ_0) to (z,ζ) in $S\mathbb{H}^2$. The uniqueness of the element $g_{z\zeta}$ shows that the diffeomorphism F intertwines the action of $\mathrm{PSL}(2,\mathbb{R})$ on $S\mathbb{H}^2$ with the left multiplication in the group. That is,

$$g((z,\zeta))=g\cdot g_{z\zeta}\ \ \forall g\in\mathrm{PSL}(2,\mathbb{R}).$$

Proposition 4.7 *The geodesic flow on* $S\mathbb{H}^2$ *corresponds to the flow on the group* $\mathrm{PSL}(2,\mathbb{R})$ *given by the right translation*

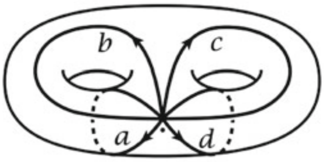
$$g\longmapsto g\cdot g_t,\ \ \textit{where}\ \ g_t=\begin{pmatrix}e^{\frac{t}{2}} & 0\\ 0 & e^{\frac{-t}{2}},\end{pmatrix}\ \ \forall t\in\mathbb{R}.$$

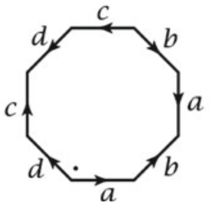
Proof It is clear that $\phi^t(z_0, \zeta_0) = g_t(z_0, \zeta_0)$, where ϕ^t is the geodesic flow. Therefore, for $(z, \zeta) \in S\mathbb{H}^2$,

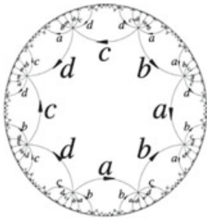
$$\phi^t(z, \zeta) = \phi^t\left(g_{z\zeta}(z_0, \zeta_0)\right) = g_{z\zeta}\left(\phi^t(z_0, \zeta_0)\right) = g_{z\zeta}\left(g_t(z_0, \zeta_0)\right) = g_{z\zeta}g_t.$$

The second equality is a result of the fact that the action of $\mathrm{PSL}(2, \mathbb{R})$ on $\mathbb{H}^2$ is by isometries, and hence takes geodesics to geodesics as described in the proof of Lemma 4.5. □

Let Σ be a compact Riemann surface of genus $g \geq 2$. Then Σ has $\mathbb{H}^2$ as its universal cover, i.e., if $\Gamma = \pi_1(\Sigma)$, the fundamental group of Σ, then Γ acts freely and discontinuously on $\mathbb{H}^2$ by deck transformations. Consequently, Γ can be identified with a discrete subgroup of $\mathrm{PSL}(2, \mathbb{R})$ such that the quotient space $\Sigma = \mathbb{H}^2/\Gamma$ is compact. Further Σ is a Riemannian manifold with constant negative curvature -1 with respect to the metric induced from $\mathbb{H}^2$ via the quotient map. The pictures in this page roughly serve to illustrate this procedure.







Proposition 4.8 *The identification of* $S\mathbb{H}^2$ *with* $\mathrm{PSL}(2, \mathbb{R})$ *induces an identification* $S\left(\mathbb{H}^2/\Gamma\right) \cong \Gamma\backslash\mathrm{PSL}(2, \mathbb{R})$. *The geodesic flow on* $S\Sigma$ *corresponds to the flow*

$$\Gamma \setminus \mathrm{PSL}(2, \mathbb{R}) \longrightarrow \Gamma\backslash\mathrm{PSL}(2, \mathbb{R}), \quad \Gamma g \longmapsto \Gamma g g_t,$$

where $g_t = \begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{\frac{-t}{2}} \end{pmatrix}$.

Proof Since $(z, \zeta) \longmapsto g_{z\zeta}$ intertwines the action of $\mathrm{PSL}(2, \mathbb{R})$, the proof follows from the previous proposition and is left as an exercise to the reader. □

4.2 Hopf's Proof of Ergodicity

In this section, we sketch a proof of the ergodicity of the geodesic flow g_t on $\Gamma\backslash$ $\mathrm{PSL}(2, \mathbb{R})$ that was originally presented by E. Hopf [9]. In this context, we introduce the notion of horocycles, some of whose illustrative examples are the lines parallel to the x-axis in $\mathbb{H}^2$. As we shall soon discover, horocycles have a very special role in the study of the dynamics of the geodesic flow.

Lines parallel to the x-axis can also be viewed as orbits of points in $\mathbb{H}^2$ under the action of the 1-parameter subgroup of $\mathrm{PSL}(2, \mathbb{R})$ consisting of matrices of the form

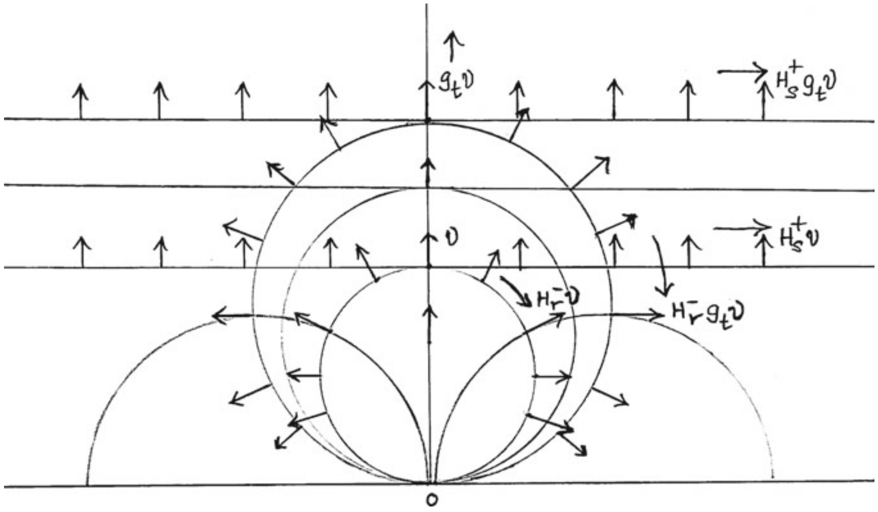


Fig. 1 Geodesic and horocycle flows

$H_s^+ = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$; that is, transformations of the form $z \longmapsto z + s$. Being orthogonal to the lines parallel to the y-axis in $\mathbb{H}^2$, it turns out that their images, under a typical element of PSL$(2, \mathbb{R})$ taking ∞ to a point x_0 on the x-axis, are the Euclidean circles in $\mathbb{H}^2$ tangent to the x-axis at the point x_0.

Moving a step further, and using the identification of PSL$(2, \mathbb{R})$ with $S\mathbb{H}^2$, we see that the 1-parameter subgroup H_s^+, of PSL$(2, \mathbb{R})$, defines a measure preserving flow on $S\mathbb{H}^2$. In a similar fashion, we observe that the 1-parameter subgroup $H_r^- = \begin{pmatrix} 1 & 0 \\ r & 1 \end{pmatrix}$ of PSL$(2, \mathbb{R})$ also defines a measure preserving flow on $S\mathbb{H}^2$. The flow H_s^+ is termed *the stable horocycle flow* while H_r^- is termed *the unstable horocycle flow*.

The next figure serves to illustrate the orbits of a vector $v \in S\mathbb{H}^2$ under the dynamics of the two horocycle flows, in relation to the geodesic flow.

The two horocycle flows determine vector fields on $S\mathbb{H}^2$ which are linearly independent, i.e., at any given point of $S\mathbb{H}^2$, the tangent vectors of the corresponding vector fields are linearly independent and hence, together with the tangent vector given by geodesic flow vector field, span the tangent space to $S\mathbb{H}^2$ at that point.

4.2.1 A Historical Interlude

Eberhard Hopf exploited the interrelation between the stable and unstable horocycle flows and the geodesic flow in his proof. Historically it was G.A. Hedlund [7] who, in 1934, first proved that the geodesic flow on closed surfaces of constant negative curvature is ergodic (which was called metric transitivity at that time). In 1936, E.

Hopf gave another proof of ergodicity in the case considered by Hedlund. Hedlund was also the first to recognize the importance of the close relationship between horocycle and geodesic flows. Later, in 1939, Hedlund proved [8] stronger properties (like mixing) for geodesic flow on surfaces of finite area and constant negative curvature. Erogdicity was extended to arbitrary dimensions for manifolds of constant negative curvature by Hopf in 1939. In the same paper [9], Hopf also proved that the geodesic flow is ergodic for a surface of finite area and of variable negative curvature under the restriction that the curvature and its first derivatives are bounded in absolute value (Fig. 1).

Gelfand and Fomin, in 1952 [5], provided the next impetus by proving the stronger property of mixing for the case of manifolds of higher dimension and constant negative curvature. Their approach and method was generalised by Mautner in 1957 [11] to prove ergodicity of the geodesic flow on locally symmetric spaces of negative curvature and arbitrary dimensions.

However the question remained open in the case of variable curvature in arbitrary dimension until 1960s when the work of Anosov and Sinai [2] led Anosov to prove ergodicity for closed manifolds of negative curvature and arbitrary dimension [1]. The approach adopted in the work of Anosov and Sinai enabled Anosov to overcome the difficulty faced by Hopf, and Anosov proved ergodicity for manifolds of finite volume and variable negative curvature under exactly the same hypothesis considered by Hopf in 1939 [9], namely when the covariant derivative of the curvature tensor is bounded in absolute value.

Remark 4.9 For manifolds of finite volume and variable negative curvature without the boundedness assumption on the first derivatives of curvature, to the best of our knowledge, the question of ergodicity is still an outstanding open problem (even for surfaces!).

Resuming the sketch of Hopf's proof, let $f : S\Sigma \longrightarrow \mathbb{R}$ be a continuous function with compact support where Σ is a surface of genus $g \geq 2$ with the hyperbolic metric. Note that as a consequence of Theorem 2.39, it suffices to consider continuous functions with compact support. We will show that f is constant a.e. when f is g_t-invariant.

For the three smooth flows g_t, H_s^+ and H_r^- on $\mathrm{PSL}(2, \mathbb{R})$, a routine computation shows that

$$H_s^+ g_t = g_t H_{e^{-t}s}^+ \text{ and } H_r^- g_t = g_t H_{e^{-t}r}^-.$$

From this, it follows that

$$f(xH_s^+ g_t) = f(xg_t H_{e^{-t}s}^+) \text{ and } f(xH_r^- g_t) = f(xg_t H_{e^{-t}r}^-).$$

Uniform continuity of f then implies that

$$\lim_{t\to\infty} \big(f(xH_s^+ g_t) - f(xg_t)\big) = \lim_{t\to\infty} \big(f(xg_t H_{e^{-t}s}^+) - f(xg_t)\big) = 0$$

and

$$\lim_{t\to\infty}\big(f(xH_r^- g_t) - f(xg_t)\big) \;=\; \lim_{t\to\infty}\big(f(xg_t H_{e^{-t}r}^-) - f(xg_t)\big) \;=\; 0.$$

Therefore,

$$\lim_{\tau\to\infty}\frac{1}{\tau}\int_0^\tau \big(f(xg_t) - f(xH_s^+ g_t)\big)\,dt \;=\; 0.$$

Similarly,

$$\lim_{\tau\to\infty}\frac{1}{\tau}\int_0^\tau \big(f(xg_{-t}) - f(xH_r^- g_{-t})\big)\,dt \;=\; 0.$$

With the notation introduced in an earlier remark in this chapter, we note that $\widetilde{f}^+(xH_s^+)$ and $\widetilde{f}^-(xH_r^-)$ exist whenever $\widetilde{f}^+(x)$ and $\widetilde{f}^-(x)$ exist. Further, we conclude from the above that $\widetilde{f}^+(x) = \widetilde{f}^+(xH_s^+)$ and $\widetilde{f}^-(x) = \widetilde{f}^-(xH_r^-)$, and are equal a.e.

Let $x_0 \in S\Sigma$. We will construct an open neighbourhood of x_0 as follows. Let $\delta_1, \delta_2, \delta_3 > 0$ be sufficiently small. Construct a smooth curve $\gamma_{\delta_1}(x_0)$ through x_0 by defining

$$\gamma_{\delta_1}(x_0) \;=\; \big\{x_0 H_r^- \;:\; |r| < \delta_1\big\}$$

and then construct an open smooth surface $\sigma_{\delta_1,\delta_2}(x_0)$ by defining

$$\begin{aligned}\sigma_{\delta_1,\delta_2}(x_0) &= \big\{x_0 H_r^- g_t \;:\; |r| < \delta_1, |t| < \delta_2\big\}\\ &= \bigcup_{|t|<\delta_2}\big(\gamma_{\delta_1}(x_0)\big)\,g_t.\end{aligned}$$

Finally, construct an open neighbourhood $U_{\delta_1,\delta_2,\delta_3}(x_0)$ by

$$U_{\delta_1,\delta_2,\delta_3}(x_0) \;=\; \bigcup_{|s|<\delta_3}\big(\sigma_{\delta_1,\delta_2}(x_0)\big)\,H_s^+.$$

It follows from the smoothness of the corresponding vector fields that for sufficiently small $\delta_1, \delta_2, \delta_3$, the surfaces $\big(\sigma_{\delta_1,\delta_2}(x_0)\big)\,H_s^+$ are disjoint for distinct s with $|s| < \delta_3$ and for the point

$$x \;=\; x_0 H_r^- g_t H_s^+ \;\in\; U_{\delta_1,\delta_2,\delta_3}(x_0),$$

the numbers r, t, s are smooth coordinates in $U_{\delta_1,\delta_2,\delta_3}(x_0)$. In fact, as x_0 varies over a compact set on $S\Sigma$, all of $\delta_1, \delta_2, \delta_3$ can be chosen to be independent of x_0. Now, the Liouville measure on $S\Sigma$ induces conditional measures on each of the surfaces $\big(\sigma_{\delta_1,\delta_2}(x_0)\big)\,H_s^+$, for all s and invoking Fubini's theorem shows that for a.e. $y \in \sigma_{\delta_1,\delta_2}(x_0)$ (with respect to the induced conditional measure), one has $\widetilde{f}^+(y) = \widetilde{f}^-(y)$; and this holds for x_0 a.e. in $S\Sigma$(with respect to μ).

We will now show that $\widetilde{f}(x)$ is constant for $x (= x_0 H_r^- g_t H_s^+)$ a.e. in $U_{\delta_1,\delta_2,\delta_3}(x_0)$. To this end, let

$$\widetilde{U} = \Bigg\{ x \in U_{\delta_1,\delta_2,\delta_3}(x_0) \ : \ \widetilde{f}^+(x) \text{ exists and } \\ \text{for } y = x_0 H_r^- g_t \in \sigma_{\delta_1,\delta_2}(x_0), \ \widetilde{f}^+(y) = \widetilde{f}^-(y) \Bigg\}.$$

Since the vector fields are smooth, it follows from Fubini's theorem that $\widetilde{U}$ has full measure in $U_{\delta_1,\delta_2,\delta_3}(x_0)$. Further, if $x_1, x_2 \in \widetilde{U}$, with $x_1 = x_0 N_{r_1}^- g_{t_1} N_{s_1}^+$ and $x_2 = x_0 N_{r_2}^- g_{t_2} N_{s_2}^+$, and if y_1, y_2, z_1, z_2 denote $x_0 N_{r_1}^- g_{t_1}$, $x_0 N_{r_2}^- g_{t_2}$, $x_0 N_{r_1}^-$ and $x_0 N_{r_2}^-$ respectively, then we have,

$$\begin{aligned} \widetilde{f}^+(x_1) = \widetilde{f}^+(y_1) &= \widetilde{f}^-(y_1) = \widetilde{f}^-(z_1) \\ = \widetilde{f}^-(z_2) &= \widetilde{f}^-(y_2) = \widetilde{f}^+(y_2) = \widetilde{f}^+(x_2). \end{aligned}$$

Thus $\widetilde{f}^+$ is constant in $\widetilde{U}$, i.e., $\widetilde{f}^+$ is constant a.e. in $U_{\delta_1,\delta_2,\delta_3}(x_0)$, which proves the ergodicity of g_t.

References

1. Anosov, D. V. (1967). Geodesic flows on closed Riemannian manifolds of negative curvature. *Proceedings of the Steklov Institute of Mathematics, 90*, 235.
2. Anosov, D. V., & Sinai, Y. G. (1967). Some smooth ergodic systems. *Russian Mathematical Surveys, 22*(5), 107–172.
3. Birkhoff, G. D. (1931). Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences of the United States of America, 17*, 656–660.
4. Caratheodory, C. (1914). Über das lineare Mass von Punktmengen -eine Verallgemeinerung des Lïgenbegriffs, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse (pp. 404–426).
5. Gel'fand, I. M., & Fomin, S. V. E. (1952). Geodesic flows on manifolds of constant negative curvature. *Uspekhi Matematicheskikh Nauk, 7*, 118–137.
6. Hausdorff, F. (1918). Dimension und äusseres mass. *Mathematische Annalen, 79*, 157–179.
7. Hedlund, G. A. (1934). On the metrical transitivity of the geodesics on closed surfaces of constant negative curvature. *Annals of Mathematics, 35*, 787–808.
8. Hedlund, G. A. (1939). Fuchsian groups and mixtures. *Annals of Mathematics, 40*, 370–383.
9. Hopf, E. (1939). Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung. *Ber. Verh. Sachs. Akad. Wiss. Leipzig, 91*, 261–304.
10. Lebesgue, H. (1904). *Leçons sur l'intégration et la recherche des fonctions primitives*. Paris: Gauthier-Villars.
11. Mautner, F. I. (1957). Geodesic flows on symmetric Riemann spaces. *Annals of Mathematics, 65*(2), 416–431.
12. Vitali, G. (1905). Sul problema della misura dei gruppi di punti di una retta, Bologna, Tip. Gamberini e Parmeggiani.
13. von Neumann, J. (1932). Proof of the quasi-ergodic hypothesis. *Proceedings of the National Academy of Sciences of the United States of America, 18*, 70–82.