Master's thesis

A dynamical system with oscillating time average

Joel Nilsson

7 June, 2020

Supervisor: Tomas Persson

Examiner: Magnus Aspenberg

LUND UNIVERSITY

Centre for Mathematical Sciences

Abstract

Given a discrete dynamical system T, one can ask what the time average of the system will be, that is, what is the average position of $T^n(x)$ for large n? Birkhoff's ergodic theorem, one of the most important results in ergodic theory, says that for an ergodic system on a finite measure space, the time average will in the limit as $n \to \infty$ be equal to the space average for almost all initial values x. In this thesis we study time averages of a dynamical system $T : [0, 1] \to [0, 1]$ that depends on a parameter α. We show that there are values of α for which the points $x, T(x), T^2(x), \ldots$ are equally often in the right half of $[0, 1]$ as on the left half. We also show that for other values of α, the time average never converges, but instead oscillates between being concentrated on the left and right halves of the unit interval. In the process, we also prove the existence of an absolutely continuous invariant ergodic measure for T.

Acknowledgements

I would like to thank my supervisor Tomas Persson of the Centre for Mathematical Sciences at Lund University. He is the one who got me interested in dynamical systems, and he has been incredibly helpful and encouraging during this thesis. I could not have finished it without his support.

Contents

1

Introduction

The study of dynamical systems arose as a way of describing systems in the real world that evolve in time. Perhaps the first example is Newton's classical mechanics, which uses differential equations to describe the movements of the planets. In this case time is continuous, but if time is seen as discrete, a physical system can be modeled by a difference equation

$$x_n = T(x_{n-1}),$$

where x_n is the state at time n and T is a function describing how the system evolves from one time point to the next.

In the field of ergodic theory, one is interested in the typical long-term behaviour of dynamical systems, not the exact behaviour of every single point. A good example is the ergodic hypothesis: the time spent by a system in a given region of the state space is proportional to the volume of this space. For example, the hypothesis says that a gas molecule in a room will be in the left half of the room one half of the time. The hypothesis was proposed by Boltzmann in 1898 in the context of statistical physics. It turned out not to be true for all physical systems, although it is often assumed to be true in statistical mechanics. A classical theorem in ergodic theory, Birkhoff's ergodic theorem, gives conditions under which the ergodic hypothesis is true. Ergodic theory studies statistical properties of dynamical systems by using measure theory, which generalizes the concepts of size, length and volume, and which is also the foundation of probability theory.

In this thesis we will study a specific dynamical system, and decide whether or not the ergodic hypothesis holds. Our system is inspired by another, similar system that has already been well studied. It is often called the LSV map, because it was introduced by Liverani, Saussol and Vaienti in [1]. It is a map $f : [0,1] \to [0,1]$ depending on a parameter $\alpha > 0$ and defined by

$$f(x) = \begin{cases} x(1+2^\alpha x^\alpha) & \text{if } 0 \leq x < 1/2 \\ 2x-1 & \text{if } 1/2 \leq x \leq 1 \end{cases}.$$

The LSV map is a special type of the so-called Pomeau-Manneville maps, introduced in [2]. These maps have two continuous branches, a fixed point at $x = 0$ and a derivative which is greater than 1 except for $f'(0) = 1$. The initial reason for studying them in [2] was that they model the behaviour of fluids in the transition to turbulence.

The LSV map has a nonlinear left branch and an expanding linear right branch. A similar system, but with two nonlinear branches, is the following:

$$T(x) = \begin{cases} x(1+2^\alpha x^\alpha) & \text{if } 0 \leq x < 1/2 \\ 1-(1-x)(1+2^\alpha(1-x)^\alpha) & \text{if } 1/2 \leq x \leq 1 \end{cases}.$$

The left branch of T is the same as the left branch of the LSV map. But instead of a linear right branch, the right branch of T is a mirrored version of the left branch. We will call T the symmetric LSV map. The main question this thesis will answer is: as T is iterated many times, will the points $x, T(x), T^2(x), \ldots$ be evenly spread out on the unit interval, or will they be more concentrated on the left or right part of it? We will prove that for some values of α, they are distributed evenly, but for other α's, the distribution changes with time, oscillating between being concentrated on the left part and being concentrated on the right.

The thesis assumes no knowledge of ergodic theory, or of measure theory. In the next chapter the basics of measure theory are presented, including the Lebesgue integral. The next chapter presents the basics of ergodic theory. The fourth chapter is about the existence and properties of a measure that is preserved by the symmetric LSV map. In the fifth chapter we use the properties of this measure to determine how the points $x, T(x), T^2(x), \ldots$ are distributed.

2

Measure theory

The purpose of measure theory is to formalize notions of size, length and volume. Besides being used in ergodic theory to study dynamical systems, it is the basis of Lebesgue integration, a generalization of Riemann integration, and of probability theory. A more thorough reference, as well as proofs of the theorems presented here, can be found in [3] or [4]. We start with the most natural way to measure the size of a set in $\mathbb{R}^n$, the *Lebesgue measure*.

2.1 Lebesgue measure

Since we will only study transformations on the real line in this thesis, we will only consider the Lebesgue measure on $\mathbb{R}$, although it can easily be extended to higher dimensions. We already know what the natural measure of an interval I is: its length $|I|$. The Lebesgue measure of an interval is indeed its length, but it also assigns a measure to more complicated sets. It does this by approximating a set A with a countable union of intervals. If a sequence $I_1, I_2, \ldots$ of intervals covers a set A, and if it barely covers anything else, then the measure of A should approximately be the sum of the lengths of the intervals.

Definition 2.1. The **Lebesgue outer measure** of a subset A of $\mathbb{R}$ is the number

$$\lambda^*(A) = \inf\left\{\sum_{k=1}^{\infty} |I_k| : A \subset \bigcup_{k=1}^{\infty} I_k \text{ and all } I_k \text{ are bounded intervals}\right\}.$$

Lebesgue outer measure is not exactly the same as Lebesgue measure. The outer measure is defined for all subsets of $\mathbb{R}$, but we do not actually want to be able to measure every subset. A property that a measure should satisfy is the following: if we partition a set A into countably many disjoint subsets, then the measure of A should equal the sum of the measures of the subsets. This is not true for the Lebesgue outer measure λ^*. The solution to this problem is to define which sets can and cannot be measured. A set A in $\mathbb{R}$ is called **Lebesgue measurable**, or just **measurable**, if for every $\epsilon > 0$, there is an open set G_ϵ such that

$$A \subset G_\epsilon \text{ and } \lambda^*(G \setminus A) < \epsilon.$$

In other words, the measurable sets are those which can be well approximated by open sets. Most or all sets in $\mathbb{R}$ that we would call "normal" are measurable, including the following kinds of sets.

Theorem 2.1. *The following kinds of sets are Lebesgue measurable:*

- $\mathbb{R}$ *and the empty set*
- *Open sets*
- *A countable union of measurable sets*
- *A countable intersection of measurable sets*
- *The complement of a measurable set*

Now we can define the Lebesgue measure on $\mathbb{R}$.

Definition 2.2. For a Lebesgue measurable set A in $\mathbb{R}$, the **Lebesgue measure** $\lambda(A)$ is defined by $\lambda(A) = \lambda^*(A)$.

The main reason for working with the Lebesgue measure, rather than the Lebesgue outer measure, is that the measure of a disjoint countable union is the sum of the measures of the sets in the union. This is true for the outer measure only for finite unions.

Theorem 2.2 (Countable additivity)**.** *If* $(A_n)_{n=1}^{\infty}$ *is a sequence of disjoint measurable sets in* $\mathbb{R}$*, then*

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

2.2 General measure spaces

We will in this thesis work with other measures than the Lebesgue measure. When defining a measure on a general set X, there are two things to consider: which sets do we want to measure, and how should they be measured?

The collection of measurable sets should satisfy the following: if we can measure a set, then we can measure its complement, and if we can measure a countable number of sets, then we can measure their union and their intersection. With these requirements, the measurable sets form a so-called σ-algebra.

Definition 2.3. Let X be a nonempty set. A collection $\mathcal{S}$ of subsets of X is called a **σ-algebra** if

(1) $\mathcal{S}$ is nonempty,

(2) $\mathcal{S}$ is closed under complements: if $A \in \mathcal{S}$, then $A^c \in \mathcal{S}$,

(3) $\mathcal{S}$ is closed under countable unions: if $\{A_n\}_{n=1}^{\infty}$ is a collection of sets in $\mathcal{S}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{S}$.

Since $\mathcal{S}$ is nonempty, it contains at least one set A. Hence it must also contain A^c and $A \cup A^c = X$, and $X^c = \varnothing$. Moreover, properties 2 and 3 imply that $\mathcal{S}$ is closed under countable intersections, because if $\{A_n\}_{n=1}^{\infty}$ is a collection of sets in $\mathcal{S}$, then

$$\bigcap_{n=1}^{\infty} A_n = \left(\bigcup_{n=1}^{\infty} A_n^c\right)^c \in \mathcal{S}.$$

Two trivial σ-algebras that exist for any X are $\mathcal{S} = \{\varnothing, X\}$ and $\mathcal{S} = \mathcal{P}(X)$, the power set of X. According to theorem 2.1, the Lebesgue measurable sets form a σ-algebra in $\mathbb{R}$.

If we have a σ-algebra $\mathcal{S}$ on X, and want a σ-algebra $\mathcal{S}_Y$ on a subset Y of X, then we can easily get it by

$$\mathcal{S}_Y = \{A \cap Y : A \in \mathcal{S}\}.$$

In this way, we can define the Lebesgue measurable sets on a proper subset of the real line, for example an interval.

Given a σ-algebra $\mathcal{S}$, we can define a measure on $\mathcal{S}$. The main property that it should have is countable additivity, which is what theorem 2.2 guarantees for the Lebesgue measure.

Definition 2.4. Let $\mathcal{S}$ be a σ-algebra on a set X. A function $\mu : \mathcal{S} \to [0, \infty]$ is called a **measure on $\mathcal{S}$** if

(1) $\mu(\varnothing) = 0$,

(2) μ is countably additive: if $\{A_n\}_{n=1}^{\infty}$ is a collection of disjoint sets in $\mathcal{S}$, then

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

The triple $(X, \mathcal{S}, \mu)$ is called a **measure space**, and the sets in $\mathcal{S}$ are said to be measurable. An example of a measure, besides Lebesgue measure, is the Dirac measure δ_x on $\mathcal{P}(X)$, defined by

$$\delta_x(A) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ otherwise} \end{cases}.$$

This works for any σ-algebra on any set. Another example is the counting measure c on $\mathcal{P}(X)$, which counts the elements in a set, meaning $c(A)$ is the number of elements in A if it is finite, and $c(A) = \infty$ if it is infinite.

We say that a measure μ is **finite** if $\mu(X) < \infty$. If $\mu(X) = 1$, then we call μ a **probability measure**. Of course, any finite measure can be made into a probability measure by rescaling it. If X is a countable union of measurable sets of finite measure, then μ is called **σ-finite**. For example, the Lebesgue measure is σ-finite, but the counting measure on $\mathbb{R}$ is not.

A set of measure zero is called a null set. In measure theory and ergodic theory, many results only hold once one discards a null set from the measure space. A property is said to hold μ-almost everywhere, or just almost everywhere, if all points that do not have the property are contained in a null set. If μ is a probability measure the property is said to hold almost surely. When one describes ergodic theory as the study of *typical* long-term behaviour, one means behaviour that almost every element in the set follows, in this sense.

Suppose $(X, \mathcal{S}, \mu)$ is a probability space, and $A_1, A_2, \ldots$ is a sequence of of events in the space. The event that infinitely many of these events happen can be written

$$\limsup_{n \to \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_n.$$

An element belongs to this set if and only if it belongs to infinitely many A_n's. The Borel–Cantelli lemma is a well known result about whether or not infinitely many events in a sequence happen.

Theorem 2.3 (Borel–Cantelli lemma). *Let $(A_n)_{n=1}^{\infty}$ be a sequence of events in a probability space $(X, \mathcal{S}, \mu)$.*

1. *If*
$$\sum_{n=1}^{\infty} \mu(A_n) < \infty,$$
then $\mu(\limsup A_n) = 0$.

2. *If*
$$\sum_{n=1}^{\infty} \mu(A_n) = \infty$$
and the events are independent, then $\mu(\limsup A_n) = 1$.

Since the Borel–Cantelli lemma was proved, mathematicians have tried to come up with less restrictive conditions than independence for the second part of the lemma. It is well-known that the sets do not need to be independent, only pairwise independent. In the context of dynamical systems T, the sets A_n are often of a specific kind: a sequence $(B_n)_{n=1}^{\infty}$ is given, and one wants to know if the events $A_n = \{T^n(x) \in B_n\}$ happen infinitely often. Independence of the sets A_n is usually too much to hope for in this case, but it might be enough that they get closer and closer to independence, that is, $\mu(A_n \cap A_m) - \mu(A_n)\mu(A_m)$ tends to 0 fast enough as $|n - m|$ gets large. Both the classical Borel–Cantelli lemma and one of its generalizations will be used in chapter 5.

2.3 The Lebesgue integral

The first way one learns how to integrate a real-valued function is by the Riemann integral. It is done by approximating the area under a curve with Riemann sums with respect to a partition of the function's domain. By contrast, the Lebesgue integral instead partitions the function's range. It is defined with respect to a measure, and when that measure is the Lebesgue measure, the integral is an extension of the ordinary Riemann integral to many functions that are not Riemann integrable. We will not go into the details of the properties of the Lebesgue integral in this thesis. We will simply define it, and later use some well-known properties of it when we need them.

Take a function $f : A \to B$ as an example, where A and B are intervals on the real line. Partition B into disjoint intervals $B_1, B_2, \ldots, B_n$, and set $A_k = f^{-1}(B_k)$ for $k = 1, 2, \ldots, n$. On each interval B_k, let a_k be the smallest number in B_k. For any set S, define the indicator function
$$\mathbb{I}_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S \end{cases} .$$

Then the function
$$s(x) = \sum_{k=1}^{n} a_k \mathbb{I}_{A_k}$$

approximates f from below, and it is a finite sum of functions that are constant on a subset of A and zero outside of it. When the sets A_k are measurable, such a function is called a **simple** function. We will use approximations by simple functions to define the Lebesgue integral.

The Lebesgue integral is not defined for all real-valued functions, but only for the so-called measurable functions. A function f is said to be **measurable** if the set

$$\{x : f(x) < a\}$$

is measurable for every real number a. One good reason for only considering such functions can be found in example above, where we approximated the function f by a simple function. The sets A_k needed to be measurable, and this requires f to be measurable.

For a measurable set A, the integral of an indicator function $\mathbb{I}_A$ with respect to a measure μ is defined as

$$\int \mathbb{I}_A d\mu = \mu(A).$$

For a simple function, we define the integral as

$$\int \left(\sum_{k=1}^{n} a_k \mathbb{I}_{A_k} \right) d\mu = \sum_{k=1}^{n} \left(a_k \int \mathbb{I}_{A_k} d\mu \right) = \sum_{k=1}^{n} a_k \mu(A_k).$$

Simple functions do not have a unique representation as a sum of indicator functions, but it turns out that the integral does not depend on the representation.

The integral of a non-negative measurable function f is defined by

$$\int f d\mu = \sup \left\{ \int s d\mu : s \text{ is simple and } s \le f \right\}.$$

But we would of course like to be able to integrate functions that take on negative values as well. For an arbitrary measurable function f, split it into $f = f^+ - f^-$, where $f^+(x) = \max(0, f(x))$ and $f^-(x) = \max(-f(x), 0)$. Then f^+ and f^- are non-negative and we can define

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

provided both integrals on the right hand side are finite. If so, we say that f is integrable, otherwise we say that it is non-integrable. The integral of f over a subset A of the measure space is defined by

$$\int_A f d\mu = \int \mathbb{I}_A \cdot f d\mu.$$

3

Ergodic theory

Ergodic theory uses the tools of measure theory to study statistical properties of dynamical systems. We start with some background on the theory of dynamical systems not involving measure theory; for more background on this, see [5]. Then we introduce the ergodic theory of dynamical systems. More on this can be found in [3].

3.1 Discrete dynamical systems

A discrete dynamical system is a set X together with a transformation T on X, that is, a mapping $T : X \to X$. We should think of T as describing how the points in X move in discrete time steps. If, for example, a particle in a room is in position x at time 0, then after one time step it will be at $T(x)$, and after two time steps, it will be at $T^2(x)$, and so on. The set $\{x, T(x), T^2(x), \ldots\}$ is called the **forward orbit** of x.

If a point p satisfies $T(p) = p$, then we call p a **fixed point**. This is because, once the orbit reaches p, it will stay there forever. A natural question to ask is: if we start the system close to a fixed point, will the orbit converge to the fixed point, or will it go away from it? We will restrict X to being a subset of $\mathbb{R}$, so that we can talk about distances and use the derivative.

Definition 3.1. A fixed point p is called **Lyapunov stable** if for all $r > 0$, there is a $\delta > 0$ such that if $|x - p| < \delta$, then

$$|T^k(x) - p| < r \quad \text{for all } k \geq 0.$$

A fixed point p is called **attracting** if p is Lyapunov stable and there is a $\delta_1 > 0$ such that if $|x - p| < \delta_1$, then

$$\lim_{n \to \infty} |T^n(x) - p| = 0.$$

A fixed point p is called **repelling** if there is an $r_1 > 0$ such that if $x \neq p$ and $|x - p| < r_1$, then there is a $k \geq 0$ such that

$$|T^k(x) - p| \geq r_1.$$

Theorem 3.1. *Let I be an interval in $\mathbb{R}$ and let T be a continuously differentiable function from and to I. Let p be a fixed point of T.*

(1) If $|T'(p)| < 1$, then p is attracting.

(2) If $|T'(p)| > 1$, then p is repelling.

Note that the theorem says nothing when $|f'(p)| = 1$. In this case p is called a neutral *fixed point.*

Proof. (1) We will use the mean value theorem to show that, if the initial point is close enough to p, then it will get closer and closer as we iterate. Pick a $\delta > 0$ such that $|T'(x)| < \lambda < 1$ if $|x - p| < \delta$, for some λ. Also pick an initial point $x \in (p - \delta, p + \delta)$. According to the mean value theorem,

$$|T(x) - p| = |T'(\xi)| \cdot |x - p|$$

for some ξ strictly between x and p (here we used that $T(p) = p$). Since $|T'(\xi)| < \lambda$,

$$|T(x) - p| < \lambda|x - p|.$$

And since $\lambda < 1$, $T(x)$ also lies in the interval $(p - \delta, p + \delta)$. Now we can repeat the argument, substituting $T(x)$ for x, which gives

$$|T^2(x) - p| < \lambda|T(x) - p| < \lambda^2|x - p|.$$

Continuing in this way, we get

$$|T^n(x) - p| < \lambda^n|x - p|$$

for all $n \geq 0$. Lyapunov stability follows because $\lambda^n < 1$, and attraction follows because $\lambda^n \to 0$ as $n \to \infty$.

(2) Since $|T'(p)| > 1$, there is a $\delta > 0$ and a λ such that $|T'(x)| > \lambda > 1$ for all $x \in (p - \delta, p + \delta)$. We can use the same argument as for part 1 to show that, as long as $T^n(x)$ stays in $(p - \delta, p + \delta)$,

$$|T^n(x) - p| > \lambda^n|x - p|.$$

But this cannot continue forever. At some point $\lambda^n|x - p|$ must lie outside $(p - \delta, p + \delta)$, and thus p must be repelling.

□

3.2 Measure-preserving transformations

Let $(X, \mathcal{S}, \mu)$ be a measure space and let T be a transformation on X. In the field of ergodic theory, one works with **measurable** transformations, transformations T such that $T^{-1}(A)$ is a measurable set whenever A is a measurable set.

Definition 3.2. If T is measurable and $\mu(A) = \mu(T^{-1}(A))$ for every measurable set A, then T is called **measure-preserving** and μ is called **T-invariant**.

For a measure-preserving transformation, the measure

$$\mu(T^{-n}(A)) = \mu(\{x : T^n(x) \in A\})$$

does not change with time, so the intuition is that, as T acts on all the points in X, the measure of the points lying in A at a given time is always the same. The following theorem gives another characterization of measure-preserving systems, which we will find useful in this thesis.

Theorem 3.2. *A transformation T on a measure space $(X, \mathcal{S}, \mu)$ is measure-preserving if and only if*

$$\int \phi \circ T d\mu = \int \phi d\mu$$

for every μ-integrable function $\phi : X \to \mathbb{R}$.

Proof. For the 'if'-part, take a measurable set A and let $\phi = \mathbb{I}_A$. Note that $\mathbb{I}_A \circ T = \mathbb{I}_{T^{-1}(A)}$. If the integrals of $\mathbb{I}_A \circ T$ and $\mathbb{I}_A$ are equal, then

$$\mu(A) = \int \mathbb{I}_A d\mu = \int \mathbb{I}_A \circ T d\mu = \int \mathbb{I}_{T^{-1}(A)} d\mu = \mu(T^{-1}(A)).$$

For the 'only if'-part, we start by proving it for characteristic functions. If T preserves μ, then for any measurable set A,

$$\int \mathbb{I}_A d\mu = \mu(A) = \mu(T^{-1}(A)) = \int \mathbb{I}_{T^{-1}(A)} d\mu = \int \mathbb{I}_A \circ T d\mu.$$

By linearity of the integral, the result also holds for all simple functions. For a non-negative function ϕ, we use the fact that there exists an increasing sequence $(\phi_n)_{n=1}^{\infty}$ of simple function converging pointwise to ϕ. For each of these, we have

$$\int \phi_n d\mu = \int \phi_n \circ T d\mu.$$

Taking limits on both sides, and switching the order of the limit and the integral using the Monotone Convergence Theorem, we get

$$\int \phi d\mu = \int \phi \circ T d\mu.$$

For a general μ-integrable function ϕ, the result holds for the positive and negative parts of ϕ separately, and so it holds for ϕ as well. □

3.3 Ergodic transformations

One of the most important concepts in ergodic theory is ergodicity. It captures the notion of a transformation $T : X \to X$ that moves X around enough so that you cannot divide X into two parts and study the system separately on those parts. If there was a proper subset A of X such that $T^{-1}(A) = A$, then T restricted to A would be a transformation, as would T restricted to the complement A^c. Therefore, all there was to know about T should be possible to learn by studying those two restrictions. Ergodicity means that this cannot happen, except possibly if A or A^c is a null set.

Definition 3.3. A measure-preserving transformation T is called **ergodic** if for all measurable sets A, $T^{-1}(A) = A$ implies $\mu(A) = 0$ or $\mu(A^c) = 0$.

Then next theorem gives some different characterizations of ergodicity. It also involves the concept of recurrence, which we will just briefly explain. A transformation T on a measure space $(X, \mathcal{S}, \mu)$ is called **recurrent** if for every set A of positive measure and almost every $x \in A$, there is an $n > 0$ such that $T^n(x) \in A$. There is a famous theorem, called the Poincaré recurrence theorem, which says that any measure-preserving transformation on a finite measure space is recurrent. Thus in the finite case, the result below need not mention recurrence. However, it also covers the σ-finite case, when recurrence is not automatic.

Theorem 3.3. *Let $(X, \mathcal{S}, \mu)$ be a σ-finite measure space and let $T : X \to X$ be a measure-preserving transformation. Then the following are equivalent:*

1. *T is recurrent and ergodic.*
2. *For every set A of positive measure, $\mu(X \setminus \bigcup_{n=1}^{\infty} T^{-n}(A)) = 0$.*
3. *For every set A of positive measure and for almost every $x \in X$ there is an integer $n > 0$ such that $T^n(x) \in A$.*
4. *If A and B are sets of positive measure, then there is an integer $n > 0$ such that $T^{-n}(A) \cap B \neq \varnothing$.*
5. *If A and B are sets of positive measure, then there is an integer $n > 0$ such that $\mu(T^{-n}(A) \cap B) > 0$.*

One of the most important results in ergodic theory is the Ergodic theorem. It comes in some different forms, but we will use the one proven by Birkhoff. If $(X, \mathcal{S}, \mu)$ is a probability space, T is a transformation on X and f is an integrable function, then we can define the *time average*

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

and the *space average*

$$\int f d\mu.$$

For general transformations, these need not be related, but according to Birkhoff's ergodic theorem, if T is ergodic, then the time average actually approaches the space average as $n \to \infty$. In other words, ergodic transformations on a probability space satisfy the ergodic hypothesis.

Theorem 3.4 (Birkhoff's Ergodic Theorem)**.** *Let $(X, \mathcal{S}, \mu)$ be a probability space and let T be a measure-preserving transformation on $(X, \mathcal{S}, \mu)$. If $f : X \to \mathbb{R}$ is an integrable function, then the limit*

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

exists for almost every $x \in X$. Denote this limit by $\tilde{f}(x)$. Moreover,

$$\int_X f d\mu = \int_X \tilde{f} d\mu.$$

If T is ergodic, then for almost every $x \in X$,

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

There is an obvious reason for why non-ergodic transformations cannot satisfy the ergodic hypothesis. If there is a set A such that $\mu(A) > 0$, $\mu(A^c) > 0$ and $T^{-1}(A) = A$, then the orbit of an $x \in A$ never reaches A^c, so the time average for $f = \mathbb{I}_{A^c}$ is zero, and cannot get close to the space average $\mu(A^c)$.

4

The symmetric LSV map

The main topic of this thesis is the symmetric LSV map $T : [0,1] \to [0,1]$, depending on a parameter $\alpha > 0$ and defined by

$$T(x) = \begin{cases} x(1 + 2^\alpha x^\alpha) & \text{if } 0 \le x < 1/2 \\ 1 - (1-x)(1 + 2^\alpha (1-x)^\alpha) & \text{if } 1/2 \le x \le 1 \end{cases}.$$

Its graph is shown in figure 4.1, for $\alpha = 1$. (In the future when figures are shown, it will always be with $\alpha = 1$, if nothing else is mentioned. It does not really matter, because the figures will look qualitatively the same if we change α.) It has two fixed points, 0 and 1. The map is symmetric in the sense that $T(1-x) = 1 - T(x)$. The derivative of T is this:

$$T'(x) = \begin{cases} 1 + 2^\alpha(\alpha+1)x^\alpha & \text{if } 0 \le x < 1/2 \\ 1 + 2^\alpha(\alpha+1)(1-x)^\alpha & \text{if } 1/2 < x \le 1 \end{cases}.$$

The fixed points are neutral, i.e. $T'(0) = T'(1) = 1$, so theorem 3.1 cannot tell us anything about their stability. However, the derivative outside the fixed points is greater than 1, so they are actually repelling. This means that the system will not behave in an uninteresting way, that is, it will not simply converge to one of the fixed points. In the next chapter we will study how the points $x, T(x), T^2(x), \ldots$ are distributed along the unit interval. But first, we will investigate the existence of a certain kind of invariant measures for T, called *absolutely continuous* measures. Proving that such a measure exists, and determining what properties it has, is interesting in its own right, but it will also help us when studying time averages of T.

A measure ν is said to be **absolutely continuous** with respect to another measure μ, defined on the same measurable space, if $\mu(A) = 0$ implies $\nu(A) = 0$. According to the Radon–Nikodym theorem [4], if ν is absolutely continuous with respect to μ and the space is σ-finite, then there is a measurable function h, called the density of ν, such that

$$\nu(A) = \int_A h \, d\mu$$

for all measurable sets A. We will use a result of Rychlik [6] about the existence of absolutely continuous invariant measures for certain dynamical systems. Rychlik's result requires the transformation to have two important properties, that we define here.

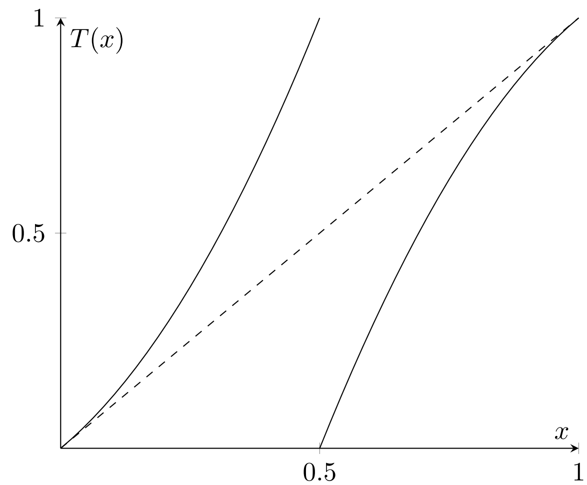

Figure 4.1: Plot of the function T for $\alpha = 1$.

Definition 4.1. A transformation $T : [a, b] \to [a, b]$ is called **piecewise expanding** if $[a, b]$ can be partitioned into a finite or countable set of intervals $\{I_i\}$ such that T is differentiable on the interior of each I_i, and there is a constant $\lambda > 1$ such that $|T'| \geq \lambda$ on the interior of each I_i.

Definition 4.2. Let $f : [a, b] \to \mathbb{R}$. The **variation** $V(f)$ of f is defined as

$$V(f) = \sup \left\{ \sum_{k=1}^{n} |f(x_{k+1}) - f(x_k)| : a \leq x_1 < x_2 < \ldots < x_n \leq b \right\}.$$

A function is said to have **bounded variation** if $V(f) < \infty$, and the set of functions $f : [a, b] \to \mathbb{R}$ of bounded variation is denoted $BV([a, b])$.

The variation of a function f measures how much a point moves in the vertical direction as it is moved along the graph of f. If f oscillates a lot, then it will have a large variation.

Now we can state Rychlik's result.

Theorem 4.1 ([6]). *If T is a piecewise expanding transformation with $1/T'$ of bounded variation, then there exists a T-invariant probability density.*

This theorem cannot be directly applied to the symmetric LSV map T, since $T' = 1$ at the endpoints. What we will instead do is apply it to what is called the *induced map* on a subinterval of $[0, 1]$.

4.1 The induced map

The domain of the induced map is the subinterval $Y = [a, b]$ centered around $1/2$ that is shown in figure 4.2. We choose the endpoints such that $T(a) = b$ and $T(b) = a$. We first

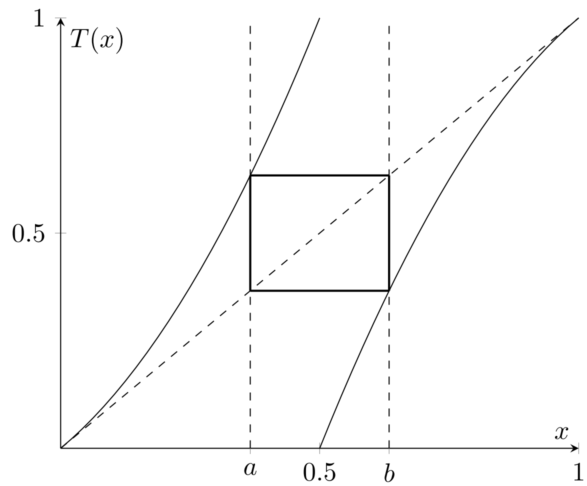

Figure 4.2: The points a and b that are the endpoints of the interval of the induced map. They are defined by $T(a) = b$ and $T(b) = a$.

define the hitting time function R by

$$R(x) = \min\{n \in \mathbb{Z}^+ : T^n(x) \in Y\}.$$

In words, this is the first positive time n at which $T^n(x)$ belongs to Y. This makes sense for all $x \in (0,1) \setminus \{1/2\}$, but when $x \in Y$ we call it the return time, hence the letter R. Then we define the induced map $F : Y \to Y$ by

$$F(x) = T^{R(x)}(x).$$

Technically, $F(1/2)$ is not well-defined, so if we want to be precise, then we should take out $1/2$ from Y. The graph of F is shown in figure 4.3. To understand why F looks like this, we need to look at the graph of T in figure 4.2. First, $F(a) = b$ and $F(b) = a$, and $R(a) = R(b) = 1$. Now let $x_1 = b$. If $x \in Y$ is very close to x_1, then $R(x) = 2$ and $F(x)$ is very close to b. As x decreases, $R(x)$ will keep being equal to 2, and $F(x)$ will decrease continuously until x reaches a value x_2 for which $F(x_2) = a$. For x close to but less than x_2, $R(x) = 3$ and $F(x)$ is close to b. Decreasing x again, $R(x)$ will keep equaling 3, and $F(x)$ will decrease until x reaches a value x_3 where $F(x_3) = a$. As x approaches $1/2$ from b, there will be infinitely many of these discontinuities x_i, and $R(x) = i + 1$ on the interval $[x_{i+1}, x_i)$. Write

$$A_i = [x_{i+1}, x_i).$$

To see that there are infinitely many A_i, we can observe that

$$\lim_{x \to 1/2^+} R(x) = \infty.$$

If there were only finitely many A_i, this limit would be finite. To the left of $1/2$ the same thing happens, giving a corresponding sequence of intervals. So Y has a countable partition $\mathcal{P}$, such that F is differentiable inside each interval of $\mathcal{P}$.

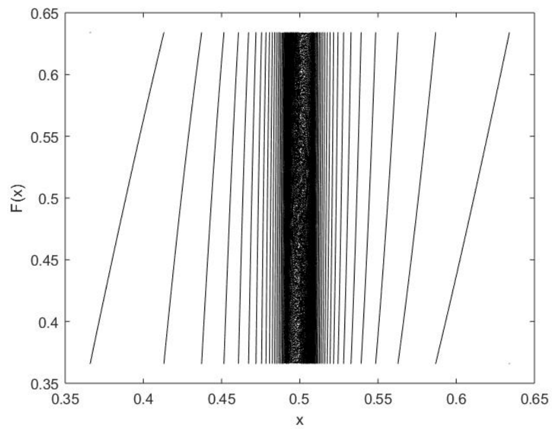

Figure 4.3: The graph of the induced map F.

We will now introduce a partition of $X \setminus Y$ as well, related to $\mathcal{P}$. It will turn out to be important for two things: to prove that F has an invariant measure which is absolutely continuous with respect to the Lebesgue measure, and to determine the finiteness of such a measure for T. Consider just the interval $[0, a]$, as the partition will be mirrored on the right side of Y. Let

$$I_1 = \{x \in [0, a) : T(x) \in Y\} \quad \text{and} \tag{4.1}$$
$$I_{n+1} = \{x \in [0, a) : T(x) \in I_n\} \quad \text{for } n \geq 1. \tag{4.2}$$

Then $\{I_n\}_{n=1}^{\infty}$ is a sequence of intervals that form a partition of $(0, a)$. By definition, $T(I_{n+1}) = I_n$. Moreover, the hitting time of an $x \in I_n$ is $R(x) = n$. The right endpoints a_n of I_n form a sequence defined by

$$a_1 = a,$$
$$a_{n+1} = T^{-1}(a_n) \cap [0, a] \quad \text{for } n > 1.$$

This sequence decreases at the rate $n^{-1/\alpha}$, as the following lemma shows.

Lemma 4.2. *Let $0 < x_1 < 1/2$, let $\hat{T}$ be the map T restricted to $[0, 1/2)$, and let $x_{n+1} = \hat{T}^{-1}(x_n)$ for $n > 0$. Then there are positive constants c_1 and c_2 such that*

$$c_1 n^{-1/\alpha} \leq x_n \leq c_2 n^{-1/\alpha}$$

for all $n > 0$.

Proof. We begin with the inequality $x_n \leq c_2 n^{-1/\alpha}$. The proof of this is mostly the same as the proof of lemma 3.2 in [1]. Let us use induction. By picking a $c_2 \geq x_1$, the inequality becomes true for $n = 1$. Now assume $x_n \leq c_2 n^{-1/\alpha}$. If the inequality does *not* hold for $n + 1$, then $x_{n+1} > c_2(n + 1)^{-1/\alpha}$, implying that

$$x_n = x_{n+1}(1 + 2^\alpha x_{n+1}^\alpha) > c_2(n + 1)^{-1/\alpha}(1 + 2^\alpha c_2^\alpha (n + 1)^{-1}).$$

The aim is now to arrive at a contradiction. Combining $x_n \leq c_2 n^{-1/\alpha}$ with the line above, we get

$$n^{-1/\alpha} > (n+1)^{-1/\alpha}\left(1+\frac{2^\alpha c_2^\alpha}{n+1}\right),$$

or equivalently,

$$\left(1+\frac{1}{n}\right)^{1/\alpha} > 1+\frac{2^\alpha c_2^\alpha}{n+1}.$$

Subtracting 1 and multiplying by $n+1$ gives

$$\left(\left(1+\frac{1}{n}\right)^{1/\alpha} - 1\right)(n+1) > 2^\alpha c_2^\alpha,$$

or

$$\frac{(n+1)^{1/\alpha} - n^{1/\alpha}}{n^{1/\alpha}}(n+1) > 2^\alpha c_2^\alpha. \tag{4.3}$$

The numerator on the left hand side of (4.3) is of the order $n^{1/\alpha-1}$. This means there is a constant K such that $(n+1)^{1/\alpha} - n^{1/\alpha} \leq K n^{1/\alpha-1}$ for all $n \geq 1$. Inserting this into (4.3) gives

$$K\frac{n+1}{n} > 2^\alpha c_2^\alpha.$$

But this is a contradiction if we let $c_2 \geq 2^{1/\alpha-1}K^{1/\alpha}$.

For the inequality $c_1 n^{-1/\alpha} \leq x_n$, we can proceed in the same way as we just did up to (4.3), except that we substitute c_1 for c_2 and reverse all inequalities. We then pick a K such that $(n+1)^{1/\alpha} - n^{1/\alpha} \geq K n^{1/\alpha-1}$ for all $n \geq 1$. Then

$$K\frac{n+1}{n} < 2^\alpha c_1^\alpha,$$

which is a contradiction if $c_1 \leq K^{1/\alpha}/2$. □

It will be convenient to have a short piece of notation for when two sequences grow at the same rate. If $a(n)$ and $b(n)$ are sequences, and there are positive constants c_1 and c_2 such that

$$c_1 \leq \frac{a(n)}{b(n)} \leq c_2,$$

we write $a(n) \sim b(n)$. If only the upper bound holds, we write $a(n) \lesssim b(n)$, and if only the lower bound holds, $a(n) \gtrsim b(n)$. Lemma 4.2 says that $a_n \sim n^{-1/\alpha}$. The lengths of the intervals I_n can then be estimated as

$$|I_n| = a_n - a_{n+1} = T(a_{n+1}) - a_{n+1} = 2^\alpha a_{n+1}^{\alpha+1} \sim (n+1)^{-1-1/\alpha} \sim n^{-1-1/\alpha}. \tag{4.4}$$

Now on to showing that F satisfies the conditions of theorem 4.1. We can use the chain rule to show that F is piecewise expanding. If $x \in A_i$ (an element of $\mathcal{P}$ to the right of $1/2$), then $F(x) = T^{i+1}(x)$. Using the chain rule several times gives

$$F'(x) = T'(T^i(x)) \cdot T'(T^{i-1}(x)) \cdot \ldots \cdot T'(T(x)) \cdot T'(x).$$

Each of these derivatives is at least 1. Moreover, $T'(x) > T'(b)$, so $F'(x) > T'(b) > 1$. The same argument applies to $x < 1/2$ as well, so F is piecewise expanding. It remains to show that $1/F'$ has bounded variation. To do this, we will investigate some *distortion estimates* of F. A distortion estimate is a bound on how much the derivative of a map can vary within a certain region, and such estimates are widely used in the study of dynamical systems.

4.2 Bounded distortion and Adler's condition

The first distortion estimate we present is just a special case of a later one, but we need to prove the special case first.

Lemma 4.3. *The map F satisfies the following* bounded distortion estimate*: there is a constant K such that for each $I \in \mathcal{P}$ and for all $x, y \in I$,*

$$\frac{1}{K} \leq \frac{F'(x)}{F'(y)} \leq K.$$

Proof. On I, $F = T^n$ for some n. By the chain rule,

$$F'(x) = T'(x) \cdot T'(T(x)) \cdot \ldots \cdot T'(T^{n-1}(x)).$$

The arguments are the same on the right and to the left of $1/2$, so let us assume I lies to the right of $1/2$. For $1 \leq i \leq n-1$, $T^i(x) \in I_{n-i}$, where I_{n-i} the interval consisting of the points in $[0, a]$ whose hitting time to $[a, b]$ is $n - i$, and which was defined in (4.1)-(4.2). Write $I_k = [a_{k+1}, a_k)$, and note that $T(a_{k+1}) = a_k$. Then since T' increases on $[0, a]$,

$$\begin{aligned}\frac{F'(x)}{F'(y)} &= \frac{T'(x) \cdot T'(T(x)) \cdot \ldots \cdot T'(T^{n-1}(x))}{T'(y) \cdot T'(T(y)) \cdot \ldots \cdot T'(T^{n-1}(y))} \\ &\leq \frac{T'(x) \cdot T'(a_{n-1}) \cdot T'(a_{n-2}) \cdot \ldots \cdot T'(a_1)}{T'(y) \cdot T'(a_n) \cdot T'(a_{n-1}) \cdot \ldots \cdot T'(a_2)} \\ &= \frac{T'(x) \cdot T'(a_1)}{T'(y) \cdot T'(a_n)} \leq \sup_{s \in [0,1]} T'(s)^2 = (2+\alpha)^2.\end{aligned}$$

Setting $K = (2 + \alpha)^2$, we have proved the lemma. □

The next distortion estimate, Adler's condition, was introduced by Roy Adler to show the existence of invariant measures for various maps, see for example [7].

Definition 4.3. Let $f : [0, 1] \to [0, 1]$ be a piecewise C^2 function, i.e. there is a finite or countable partition $\mathcal{P}$ of intervals of $[0, 1]$ such that f is twice differentiable on each $I \in \mathcal{P}$. Then f satisfies **Adler's condition** if there is a number $L < \infty$ such that

$$\sup_{I \in \mathcal{P}} \sup_{x,y \in I} \left| \frac{f''(x)}{f'(y)^2} \right| < L.$$

If f were piecewise linear, then the quantity above would be zero, and the condition would be easily satisfied. And if f is not too far from being piecewise linear, the condition should still be satisfied, so Adler's condition is a measure of how close a function is to being piecewise linear.

Lemma 4.4. *The map F satisfies Adler's condition.*

Proof. Take an $I \in \mathcal{P}$ and two points $x, y \in \mathcal{P}$. On I, $F = T^n$ for some n, so we need to calculate the second derivative of T^n. The chain rule and the product rule give

$$(T^n)''(x) = \left(\prod_{i=0}^{n-1} T'(T^i)\right)'(x) = \sum_{i=0}^{n-1}(T'(T^i))'(x)\prod_{\substack{k=0\\k\neq i}}^{n-1} T'(T^i(x))$$
$$= \sum_{i=0}^{n-1} T''(T^i(x))\prod_{j=0}^{i-1} T'(T^j(x))\prod_{\substack{k=0\\k\neq i}}^{n-1} T'(T^i(x)) = \sum_{i=0}^{n-1}\frac{T''(T^i(x))}{T'(T^i(x))}F'(x)\prod_{j=0}^{i-1} T'(T^j(x)).$$

Using this and lemma 4.3,

$$\begin{aligned}\frac{F''(x)}{F'(y)^2} &= \frac{\sum_{i=0}^{n-1} T''(T^i(x))/T'(T^i(x)) \cdot F'(x) \cdot \prod_{j=0}^{i-1} T'(T^j(x))}{F'(y)^2}\\ &\leq K \cdot \frac{\sum_{i=0}^{n-1} T''(T^i(x))/T'(T^i(x)) \cdot \prod_{j=0}^{i-1} T'(T^j(x))}{F'(y)}\\ &\leq K^2 \sum_{i=0}^{n-1}\frac{T''(T^i(x))}{T'(T^i(x))} \cdot \prod_{k=i}^{n-1}\frac{1}{T'(T^k(x))}\\ &\leq C\sum_{i=0}^{n-1}\frac{1}{(T^{n-i})'(T^i(x))}\end{aligned}$$

for some constant C, since T''/T' is bounded. Let us now assume $x > 1/2$; the situation would be symmetrical if $x < 1/2$. The point $T^i(x)$ lies in the interval I_{n-i}. We have $T^{n-i}(I_{n-i}) = [a, b)$, so by the mean value theorem, there is a point $\xi_i \in I_{n-i}$ such that

$$b - a = (T^{n-i})'(\xi_i) \cdot |I_{n-i}|.$$

The argument in the proof of lemma 4.3 can just as well be applied to T^k on I_k, meaning that if $x, y \in I_k$, then

$$\frac{1}{K} \leq \frac{(T^k)'(x)}{(T^k)'(y)} \leq K.$$

Hence $(T^{n-i})'(T^i(x)) \geq (T^{n-i})'(\xi_i)/K$, and so

$$\frac{F''(x)}{F'(y)^2} \leq CK\sum_{i=0}^{n-1}\frac{1}{(T^{n-i})'(\xi_i)} = CK\sum_{i=0}^{n-1}\frac{|I_{n-i}|}{b-a} \leq \frac{CKa}{b-a},$$

where the last inequality is due to the I_k's being disjoint intervals in $[0, a]$. $\square$

Adler's condition can be used to prove a stronger version of the bounded distortion in lemma 4.3. Let

$$\mathcal{P}_n = \left\{\bigcap_{k=0}^{n-1} F^{-k}(I_k) : I_k \in \mathcal{P} \text{ for } k = 0, 1, \ldots, n-1\right\}.$$

Then $\mathcal{P}_n$ consists of the continuity intervals of F^n. The distortion estimate in lemma 4.3 can be improved to hold for F^n on the sets of $\mathcal{P}_n$, independently of n.

Lemma 4.5. *The map F has* uniformly bounded distortion*: there is a constant M such that for all $n > 0$, all $I \in \mathcal{P}_n$ and all $x, y \in I$,*

$$\frac{1}{M} \leq \frac{(F^n)'(x)}{(F^n)'(y)} \leq M.$$

Proof. The first step is to show that $(F^n)''(x)/(F^n)'(x)$ is bounded on $I \in \mathcal{P}_n$. Following essentially the steps in the proof of lemma 4.4, we get

$$\frac{(F^n)''(x)}{(F^n)'(x)} = \sum_{i=0}^{n-1} \frac{F''(F^i(x))}{F'(F^i(x))^2} \cdot \prod_{k=i+1}^{n-1} \frac{1}{F'(F^k(x))}.$$

We have shown that F is piecewise expanding, so $F'(F^k(x)) \geq \lambda$ for some $\lambda > 1$. This and Adler's condition gives

$$\frac{(F^n)''(x)}{(F^n)'(x)} \leq L \sum_{i=0}^{n-1} \frac{1}{\lambda^{n-1-i}} < L \sum_{k=0}^{\infty} \frac{1}{\lambda^k} = \frac{L\lambda}{1-\lambda}.$$

Using this bound, we have, if $x, y \in I \in \mathcal{P}_n$,

$$\begin{aligned} \left|\log \frac{(F^n)'(x)}{(F^n)'(y)}\right| &= |\log|(F^n)'(x)| - \log|(F^n)'(y)|| = \left|\int_y^x \frac{(F^n)''(t)}{(F^n)'(t)} dt\right| \\ &\leq \sup \left|\frac{(F^n)''(t)}{(F^n)'(t)^2}\right| \cdot \left|\int_y^x (F^n)'(t) dt\right| \leq \frac{L\lambda}{1-\lambda} |F^n(x) - F^n(y)| \\ &\leq \frac{L\lambda}{\lambda - 1}. \end{aligned}$$

This means that

$$\exp\left(-\frac{L\lambda}{\lambda-1}\right) \leq \frac{(F^n)'(x)}{(F^n)'(y)} \leq \exp\left(\frac{L\lambda}{\lambda-1}\right),$$

proving the lemma. □

Using Adler's condition and bounded distortion, it is easy to prove that $1/F'$ has bounded distortion, as shown in [8]. Assume $I \in \mathcal{P}$ and $x, y \in I$. Then because of Adler's condition and bounded distortion,

$$\begin{aligned} \left|\frac{1}{F'(x)} - \frac{1}{F'(y)}\right| &= \left|\frac{F'(y) - F'(x)}{F'(x)F'(y)}\right| \leq \int_x^y \frac{|F''(s)|}{|F'(x)F'(y)|} ds = \int_x^y \frac{|F''(s)|}{F'(s)^2} \cdot \frac{F'(s)^2}{|F'(x)F'(y)|} ds \\ &\leq LK^2 |x - y|. \end{aligned}$$

If $x \in I \in \mathcal{P}$ and $y \in J \in \mathcal{P}$, then because of bounded distortion and the mean value theorem,

$$\left|\frac{1}{F'(x)} - \frac{1}{F'(y)}\right| \leq \left|\frac{1}{F'(x)}\right| + \left|\frac{1}{F'(y)}\right| \leq \frac{K|I|}{b-a} + \frac{K|J|}{b-a}.$$

Recall how variation was defined:

$$V(1/F') = \sup \left\{ \sum_{n=1}^{N} \left| \frac{1}{F'(a_{n+1})} - \frac{1}{F'(a_n)} \right| : a \leq a_1 < a_2 < \ldots < a_N \leq b \right\}.$$

Inside of a given interval $I \in \mathcal{P}$, the variation can be at most $LK^2|I|$, because of the first of the two observations above. The second observation means that each interval $I \in \mathcal{P}$ may also contribute at most $2K|I|/(b-a)$ to the total variation through terms $|1/F'(a_k)-1/F'(a_{k+1})|$ where one of a_k and a_{k+1} lies inside I and the other lies outside. Adding the contributions for each $I \in \mathcal{P}$, we get

$$V(1/F') \leq \sum_{I\in\mathcal{P}} LK^2|I| + \frac{2K|I|}{b-a} = LK^2(b-a) + 2K.$$

Here we did not actually need uniformly bounded distortion, but it will be needed later on.

With bounded variation proved, the conditions of Rychlik's theorem 4.1 are satisfied, and so F has an invariant probability measure ν, with a density h.

4.3 The density and the Perron–Frobenius operator

Rychlik proved theorem 4.1 by methods of functional analysis, and one of the main tools he used was the Perron–Frobenius operator, an operator that is commonly used to prove the existence of invariant measures.

Definition 4.4. Let T be a piecewise expanding transformation. The **Perron–Frobenius** operator $P_T : L^1(\lambda) \to L^1(\lambda)$ is defined by

$$P_T f(x) = \sum_{y\in T^{-1}(y)} \frac{f(y)}{|T'(y)|}.$$

The Perron–Frobenius operator is sometimes called the transfer operator. The usefulness of the operator is that $P_T h = h$ if and only if h is a T-invariant density. This is well-known, but let us prove it ourselves. Assume h is a density for the T-invariant measure μ. Then for all measurable sets E,

$$\int_E h(x)dx = \int_{T^{-1}(E)} h(x)dx.$$

Since T is piecewise expanding, the interval $[0, 1]$ can be partitioned into countably many intervals $A_1, A_2, \ldots$ (possibly a finite list) with disjoint interiors inside of which T is injective (if T were not injective on some A_k, then T' would have to be zero somewhere in A_k). The map T can therefore be divided into the maps $T_k = T|_{A_k}$, which are injective. Integration by the substitution $y = T_k(x)$ now gives

$$\begin{aligned}
\int_E h(x)dx &= \int_{T^{-1}(E)} h(x)dx \\
&= \sum_k \int_{T_k^{-1}(E)} h(x)dx \\
&= \sum_k \int_{T_k^{-1}(E)} \frac{h(x)}{|T'(x)|}|T'(x)|dx \\
&= \sum_k \int_{E\cap T_k(A_k)} \frac{h(T_k^{-1}(y))}{|T'(T_k^{-1}(y))|}dy \\
&= \sum_k \int_E \mathbb{I}_{T_k(A_k)} \frac{h(T_k^{-1}(y))}{|T'(T_k^{-1}(y))|}dy.
\end{aligned}$$

This last row is not technically correct, since $T_k^{-1}(y)$ does not exist unless $y \in T_k(A_k)$. But we can easily get around this by extending T_k^{-1} to being identically zero where it would otherwise be undefined. The next step is to interchange the sum and the integral. If the discontinuities of T are finite in number then this is no problem, and if there are infinitely many, then we can for example use the Monotone Convergence Theorem (see [4]). The interchange gives

$$\int_E \left(h(x) - \sum_k \mathbb{I}_{T_k(A_k)} \frac{h(T_k^{-1}(y))}{|T'(T_k^{-1}(y))|} \right) dx = 0.$$

And since the set E can be any measurable set, the integrand has to be identically zero outside of a null set. Thus

$$h(x) = \sum_k \mathbb{I}_{T_k(A_k)} \frac{h(T_k^{-1}(x))}{|T'(T_k^{-1}(x))|} = \sum_{\substack{k \\ T_k^{-1}(x) \text{ exists}}} \frac{h(T_k^{-1}(x))}{|T'(T_k^{-1}(x))|} = \sum_{y\in T^{-1}(x)} \frac{h(y)}{|T'(y)|}.$$

Now we have shown that, if h is a T-invariant density, then $P_T h = h$. For the converse, we reverse the argument.

The Perron–Frobenius operator will be useful for us, for instance in the proof of the following lemma, which says that the F-invariant density h is bounded away from 0 and ∞.

Lemma 4.6. *There are positive constants c_1 and c_2 such that $c_1 \leq h \leq c_2$.*

Proof. Rychlik showed in [6] that h has bounded variation, which obviously implies an upper bound c_2. The lower bound is what requires some work from us. The measure ν is a probability measure, which means that

$$\int_a^b h(x)dx = 1.$$

Since h has bounded variation, the set of discontinuities of h is countable (see for example [9] for a proof of this), hence a Lebesgue null set, and according to Lebesgue's criterion for Riemann integrability [3], h is thus Riemann integrable. The integral above can therefore be seen as an ordinary Riemann integral. That the integral of h is positive implies there is a step function $S \leq h$ that is greater than $\epsilon > 0$ on some entire interval I, which in turn means that $h(x) > \epsilon$ on I. So h is bounded from below on I, and the next step is to somehow extend this to all of Y. We will use the Perron–Frobenius operator for this.

The partition elements of $\mathcal{P}_n$ get arbitrarily small, so there is an n and a $J \in \mathcal{P}_n$ such that $J \subset I$. Since $F^n(J) = Y$, also $F^n(I) = Y$. The measure ν is invariant with respect to F^n, according to the following calculations, where we use the F-invariance of ν:

$$\nu((F^n)^{-1}(E)) = \nu((F^{-n})(E)) = \nu((F^{-n+1})(E)) = \ldots = \nu(F^{-1}(E)) = \nu(E).$$

Hence h is a fixed point of P_{F^n}, meaning that

$$h(x) = \sum_{y \in F^{-n}(x)} \frac{h(y)}{|(F^n)'(y)|}.$$

Since $F^n(I) = Y$, at least one of these y's lie in I. We therefore have the lower bound $c_1 = \epsilon / \sup_{y \in I} |(F^n)'(y)|$. Should $\sup_{y \in I} |(F^n)'(y)| = \infty$, then this would not work, but we can make sure that the supremum is finite. According to the chain rule,

$$(F^n)'(y) = \prod_{i=0}^{n-1} F'(F^i(y)).$$

If the closure of some $F^k(I)$, $0 \leq k \leq n-1$, contains the midpoint 1/2 we have the problem of unbounded $(F^n)'$, but then we could actually cut off a piece of $F^k(I)$ so that there is a positive distance between it and 1/2, and so that $F^n(I)$ still equals Y (this is possible because $F^k(I)$ would have to contain infinitely many intervals of F's partition $\mathcal{P}$). This can in turn be achieved by cutting off a piece of I. Thus $(F^n)'$ becomes bounded on I, and the lower bound c_1 becomes positive. □

That h is bounded away from zero can be used to show that F is ergodic with respect to ν. Assume there is a set A such that $\nu(A) > 0$ and $T^{-1}(A) = A$. Then the measure ν_1 with density $h_1 = h \cdot \mathbb{I}_A$ is also invariant. To see this, take a measurable set E and set $B = E \cap A$ and $C = Y \setminus A$. Then

$$\nu_1(F^{-1}(E)) = \nu_1(F^{-1}(B)) + \nu_1(F^{-1}(C)) = \nu(F^{-1}(B)) + 0 = \nu(B) = \nu_1(B) = \nu_1(E).$$

But lemma 4.6 applies to any F-invariant density, meaning that h_1 must be bounded away from zero. If $\nu(A^c) > 0$ it is obviously not, so $\nu(A^c) = 0$, and F is ergodic.

Rychlik's theorem does not say that the F-invariant density h must be unique, but it is unique in our case — up to a Lebesgue null set of course. This follows from Birkhoff's ergodic theorem. Assume h_1 and h_2 are both F-invariant densities. Then since F is ergodic with respect to h_1 and h_2, the time average must converge to

$$\int_Y h_1(x)\phi(x)dx = \int_Y h_2(x)\phi(x)dx$$

for every function ϕ that is integrable with respect to both. Now assume there is a set A with Lebesgue measure $\lambda(A) > 0$ such that $h_1(x) \neq h_2(x)$ for all $x \in A$. Then either

$A \cap \{h_1 > h_2\}$ or $A \cap \{h_1 < h_2\}$ has positive Lebesgue measure; assume the former. With $B = A \cap \{h_1 > h_2\}$, this means that

$$\int_Y h_1(x)\mathbb{I}_B(x)dx > \int_Y h_2(x)\mathbb{I}_B(x)dx,$$

but this contradicts the line above if we set $\phi = \mathbb{I}_B$.

Let us summarize the facts about ν.

Theorem 4.7. *The measure ν is the unique absolutely continuous F-invariant ergodic measure, and its density h is bounded away from 0 and ∞.*

4.4 Constructing the T-invariant measure

The process of inducing a system on a subset of the domain by using the return time R is common in ergodic theory, as it is sometimes easier to find an invariant measure ν for the induced system. There is a standard way of using ν to construct an invariant measure μ of the original system, which we will follow. It is described by Zweimüller in [10], where μ is defined in the following way (where we extend ν to X by $\nu(X \setminus Y) = 0$):

$$\mu(E) = \sum_{n=0}^{\infty} \nu(\{R > n\} \cap T^{-n}(E)). \tag{4.5}$$

If E is a Lebesgue measurable set in X, then

$$\begin{aligned}
\mu(T^{-1}(E)) &= \sum_{n=0}^{\infty} \nu(\{R > n\} \cap T^{-n-1}(E)) \\
&= \sum_{n=1}^{\infty} \nu(\{R > n\} \cap T^{-n}(E)) + \sum_{n=1}^{\infty} \nu(\{R = n\} \cap T^{-n}(E)) \\
&= \mu(E) - \nu(E \cap Y) + \sum_{n=1}^{\infty} \nu(\{R = n\} \cap T^{-n}(E)) \\
&= \mu(E) - \nu(E \cap Y) + \nu(F^{-1}(E \cap Y)) \\
&= \mu(E),
\end{aligned}$$

proving that μ is T-invariant.

We want to know for which α that μ is finite. First, it follows from the definition of μ that $\mu(Y) = \nu(Y) = 1$. Next, let us calculate $\mu(I_k)$, where $\{I_k\}_{k=1}^{\infty}$ is the partition of $(0, a)$ that was defined in (4.1)-(4.2). Recall that A_k is the interval in $\mathcal{P}$ to the right of $1/2$ that has return time $k+1$. Since $T(A_k) \subset (0, a)$, and since $R(I_k) = k$, we must have $T(A_k) = I_k$. Moreover $T(I_{k+1}) = I_k$, so $T^n(A_{n+k-1}) = T^{n-1}(I_{n+k-1}) = T^{n-2}(I_{n+k-2}) = \ldots = I_k$. This implies, for $n > 0$, that

$$\{R > n\} \cap T^{-n}(I_k) \cap Y = \left(\bigcup_{i=n}^{\infty} A_i \right) \cap T^{-n}(I_k) = A_{n+k-1}.$$

Then

$$\mu(I_k) = \sum_{n=0}^{\infty} \nu(\{R > n\} \cap T^{-n}(I_k)) = \sum_{n=1}^{\infty} \nu(A_{n+k-1}) = \sum_{n=k}^{\infty} \nu(A_n).$$

The lower and upper bounds on the density h mean that $\nu(A_n) \sim |A_n|$, and the fact that T' is bounded between 1 and $2+\alpha$ means that $|A_n| \sim |T(A_n)| = |I_n| \sim n^{-1-1/\alpha}$. Thus $\nu(A_n) \sim n^{-1-1/\alpha}$, and so

$$\mu(I_k) = \sum_{n=k}^{\infty} \nu(A_n) \sim \sum_{n=k}^{\infty} n^{-1-1/\alpha} \sim k^{-1/\alpha}.$$

We have

$$\mu(X) = \mu(Y) + 2\sum_{k=1}^{\infty} \mu(I_k) \sim 1 + 2\sum_{k=1}^{\infty} k^{-1/\alpha},$$

and we can see that $\mu(X) < \infty$ if and only if $\alpha < 1$.

The measure μ is equivalent to the Lebesgue measure λ, by which we mean that $\mu(A) = 0$ if and only if $\lambda(A) = 0$. To see this, first assume $\mu(A) = 0$. It makes no difference if 0 or 1 is in A, since they have zero measure, so assume $0, 1 \notin A$. Set $B = A \setminus Y$ and $C = A \cap Y$. Because $\mu|_Y = \nu$ and ν has a density bounded away from 0, we must have $\lambda(C) = 0$. For B, set $E = T^{-1}(B) \cap Y$. Then we have $\mu(E) = 0$, and $\lambda(E) = 0$, again because of the density h. The set E gets mapped to B by T, and T maps Lebesgue null sets to Lebesgue null sets, meaning that $\lambda(B) = 0$. (A function need not map null sets to null sets, but T does, because it is Lipschitz continuous on the two halves of the unit interval.) Thus every μ-null set is a Lebesgue null set.

Now assume $\lambda(A) = 0$, and set $B = A \setminus Y$ and $C = A \cap Y$. Then $\mu(C) = 0$, again because of the density h. Set $E = T^{-1}(B) \cap Y$. Then $T(E) = B$ (possibly excluding 0 and 1, which are both μ and λ null sets), and since T' is strictly positive and $\lambda(B) = 0$, $\lambda(E) = 0$. Therefore $\mu(E) = 0$, and the T-invariance of μ gives $\mu(B) = 0$, so $\mu(A) = \mu(B) + \mu(C) = 0$. Hence μ and λ are equivalent.

The measure μ is ergodic and recurrent. We can show this using theorem 3.3, which says that ergodicity and recurrence is equivalent to the statement that if $\mu(A) > 0$, then for almost every $x \in X$ there is an $n > 0$ such that $T^n(x) \in A$. Take an arbitrary set A of positive measure. Fix a natural number k, and take an arbitrary point x with $R(x) = k$. Then $T^k(x) \in Y$. Now assume $\mu(A \cap Y) > 0$. We know that F is ergodic with respect to $\nu = \mu|_Y$, which means that for μ-almost every $y \in Y$, $F^m(y) \in A$ for some $m > 0$. And since μ is T-invariant, it is also true that for μ-almost every x with hitting time k, there is an $m > 0$ such that $F^m(T^k(x)) \in A \cap Y$, and hence an $n > 0$ such that $T^n(x) \in A$. If $\mu(A \cap Y) = 0$, then replacing A with $T^{-1}(A) \cap Y$ in the last two sentences again shows that for μ-almost every $x \in X$ with $R(x) = k$ there is an $n > 0$ such that $T^n(x) \in A$. Because $X \setminus \{0, 1, 1/2\}$ is the countable union of the sets $\{R = k\}$, the same holds for μ-almost every $x \in X$.

Let us summarize the facts about μ.

Theorem 4.8. *The measure μ defined in* (4.5) *is T-invariant, ergodic, recurrent and equivalent to the Lebesgue measure. If $\alpha < 1$ it is finite, otherwise it is infinite.*

5

Time averages of the symmetric LSV map

In this chapter we are going to investigate how the points $x, T(x), T^2(x), \dots$ are distributed on the unit interval. To make this more concrete, we first introduce the function

$$\phi(x) = \begin{cases} -1 & \text{if } x \leq 1/2 \\ 1 & \text{if } x > 1/2 \end{cases},$$

which indicates if x is in the left or right part of $[0, 1]$. Then we define the so-called Birkhoff sums

$$S_n^\phi(x) = \sum_{k=0}^{n-1} \phi(T^k(x)) \tag{5.1}$$

and the time averages

$$E_n^\phi(x) = \frac{S_n^\phi(x)}{n}. \tag{5.2}$$

The time average E_n is a measure of how much time the orbit of x spends in the left and right parts of the unit interval. The following theorem is the main result of this chapter, and of this thesis.

Theorem 5.1.

1. *If* $0 < \alpha < 1$, *then* $E_n^\phi(x) \to 0$ *as* $n \to \infty$ *for Lebesgue-almost every* x.
2. *If* $\alpha \geq 1$, *then*

$$\liminf_{n\to\infty} E_n^\phi(x) = -1 \quad \textit{and} \quad \limsup_{n\to\infty} E_n^\phi(x) = 1$$

for Lebesgue-almost every x.

The second item in the theorem says that, if $\alpha \geq 1$, then the time average oscillates, but it says even more. It says that one can find time points N such that the orbit has spent almost all of its time up to N in the left half of the unit interval, as well as time points where it has spent almost all of its time in the right half.

Theorem 5.1 has an interesting corollary, involving what are called physical measures. If a transformation on a probability space is ergodic, then the time average converges to the space average for almost every initial point. However, there are systems f where, although the time average does not converge to the space average for almost every x, it does so for a quite large set of initial points. An f-invariant measure μ is called a **physical measure** if there is a set U with positive Lebesgue measure such that for each $x \in U$ and each continuous $\phi : X \to \mathbb{R}$,

$$\frac{1}{n}\sum_{k=0}^{n-1}\phi(f^k(x)) \to \int f\,d\mu.$$

Every invariant ergodic probability measure which is absolutely continuous with respect to the Lebesgue measure is physical, since U can be taken to be X minus a null set. An example of a physical measure which is not absolutely continuous is the Dirac measure δ_p where p is an attracting fixed point. Here we can take U to be a small enough interval around p that every $x \in U$ converges to p, making the time average converge to $\phi(p) = \int \phi d\delta_p$. It is not known exactly which systems have a physical measure, but theorem 5.1 implies that T has no physical measure if $\alpha \geq 1$. For the time average never converges according to the theorem, which means it certainly cannot converge to any space average. A small detail here is that ϕ in the definition of a physical measure is continuous, while in theorem 5.1 it is not. But this can be easily fixed by modifying the discontinuous ϕ so that it increases continuously from -1 to 1 in a small interval around 1/2.

5.1 Simulations

Before trying to prove theorem 5.1, it is a good idea to look at some simulations, and see if they agree with the theorem. I decided to iterate T a hundred million times and plot the time average as it evolved, and I did this for nine different values of α: 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8 and 2. The results are shown in figure 5.1. The x-axis is logarithmic, since the ups and downs of the time average seem to occur at intervals that get exponentially longer with time. For all $\alpha < 1$ the time average does appear to converge, while for $\alpha \geq 1$ it does not, at least not before 10^8 iterations. Thus the graphs agree with theorem 5.1. Contrast this with the time average of the LSV map, shown for the same α's in figure 5.2. Here the time average always converges, though not to zero.

Let us give a heuristic explanation for why the time average oscillates for the symmetric LSV map T, and why it does not oscillate for the ordinary LSV map. It has to do with the neutral fixed points 0 and 1. If x lies very close to one of these, then because $T'(x) \approx 1$, $T(x) \approx x$, and the orbit of x will stay close to the fixed point for a long time, which gives many contributions of the same sign to the time average, enough to substantially change it. Because of the symmetry of T, the orbit should get close to 0 about as often as it gets close to 1, causing the time average to go back and forth. As time increases, the orbit needs to pass closer and closer to the fixed points to cause a noticeable change in the time average. Apparently, if $\alpha < 1$, the orbit does not tend to pass close enough to the fixed points. Regarding the LSV map, 1 is not a neutral fixed point, so it does not counteract the influence of the neutral fixed point 0.

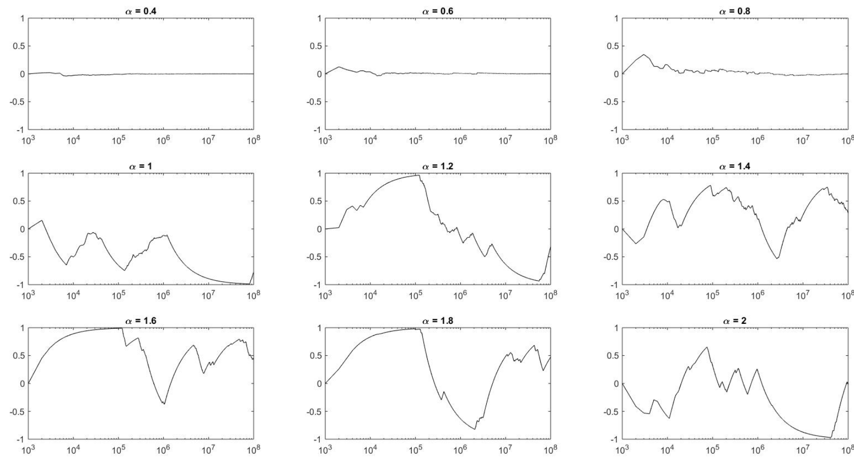

Figure 5.1: The time average of T as a function of time for nine different values of α. Note the logarithmic x-axes.

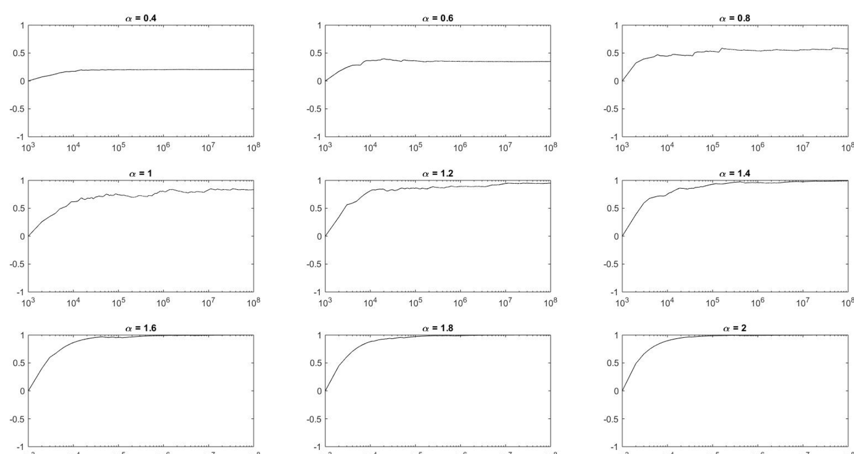

Figure 5.2: The time average of the LSV map as a function of time for nine different values of α. Note the logarithmic x-axes.

5.2 Converging time average

The first part of theorem 5.1, that the time average converges to 0 when $\alpha < 1$, is quite easy to prove using the results of the previous chapter. The measure μ is T-invariant, ergodic, and if $\alpha < 1$ it is finite, and after rescaling, a probability measure. Obviously ϕ is also integrable in this case, so Birkhoffs ergodic theorem (3.4) applies. The only thing left to prove is that the integral of ϕ is 0. It seems obvious, since T is symmetric around 1/2, but we should prove it. It is because μ is symmetric in the sense that $\mu(A) = \mu(1 - A)$, which in turn is because $h(x) = h(1-x)$. We can use the Perron-Frobenius operator to show this. Since h is a fixed point of the Perron-Frobenius operator P_F, and since $F(x) = 1 - F(1-x)$,

$$\begin{aligned} h(1-x) &= \sum_{F(1-x)=y} \frac{h(y)}{|F'(y)|} = \sum_{F(1-x)=1-y} \frac{h(1-y)}{|F'(1-y)|} = \sum_{1-F(1-x)=y} \frac{h(1-y)}{|F'(y)|} \\ &= \sum_{F(x)=y} \frac{h(1-y)}{|F'(y)|} = P_F(h(1-x)). \end{aligned}$$

Thus $h(1-x)$ is an F-invariant density, and since there is only one, $h(x) = h(1-x)$ for Lebesgue-almost every $x \in Y$. The symmetry of μ now follows from how it was constructed. According to Birkhoff's ergodic theorem,

$$E_n^\phi = \frac{1}{n} \sum_{k=0}^{n-1} \phi(T^k(x)) \to \int \phi \, d\mu = 0$$

for μ-almost all $x \in X$. And because μ is equivalent to the Lebesgue measure, $E_n^\phi \to 0$ for Lebesgue-almost every $x \in X$ as well.

5.3 Oscillating time average

With the case $\alpha < 1$ taken care of, we will now turn to the case $\alpha \geq 1$, and prove the second part of theorem 5.1, oscillating behaviour of the time average. It is, however, easier to do this by proving oscillating behaviour for a time average of F. Instead of the function ϕ which was used for E_n^ϕ, we use the function $\psi : Y \to \mathbb{R}$ defined by

$$\psi(x) = \begin{cases} R(x) - 2 & \text{if } x \leq 1/2 \\ -R(x) + 2 & \text{if } x > 1/2 \end{cases}.$$

The point of using this function is that it has the same sign as ϕ, but it weights the argument according to its return time to Y, so that

$$\psi(x) = \sum_{k=0}^{R(x)-1} \phi(T^k(x)). \tag{5.3}$$

Now we define the Birkhoff sums

$$S_n^\psi(x) = \sum_{k=0}^{n-1} \psi(F^k(x)),$$

and the time averages $E_n^\psi(x) = S_n^\psi(x)/n$. The notation is perhaps a bit unclear, since it does not mention F, but it should not cause any confusion. Because of (5.3), the Birkhoff sums S_n^ϕ and S_n^ψ are related in the following way:

$$S_n^\psi(x) = \sum_{k=0}^{n-1}\left(\sum_{i=0}^{R(F^k(x))-1}\phi(T^i(F^k(x)))\right) = S_m^\phi(x),$$

where $m > n$. Therefore, if $E_n^\psi(x)$ oscillates, then so does $E_n^\phi(x)$. This is of course only valid for $x \in Y$, but since almost every orbit of X eventually reaches Y, it is no problem to restrict x to Y.

The strategy to show oscillating behaviour will be to prove that the individual terms of S_n^ψ get big enough often enough to single-handedly change the sign of E_n^ψ. More specifically, we will prove that $|\psi(F^n(x))| - S_n^\psi(x) > 0$ infinitely often, both for $F^n(x) > 1/2$ and $F^n(x) < 1/2$, and that this difference becomes very large. We will then show why this implies that $\liminf E_n^\phi = -1$ and $\limsup E_n^\phi = 1$.

It will turn out to be easier to analyze a modified version of the Birkhoff sums S_n^ψ, where ψ is replaced by non-negative functions with finite expectation. Firstly, consider replacing ψ by the return time R, which is equal to $|\psi| + 2$. If $R(F^n(x)) > S_n^R(x)$ infinitely often, then also $|\psi(F^n(x))| > S_n^\psi(x)$ infinitely often. However, the expectation of the terms of S_n^R is by F-invariance equal to the expectation of R, and

$$\begin{aligned}\mathbb{E}(R) = \int_Y R(x)h(x)dx &\sim \int_Y R(x)dx = 2\sum_{k=1}^{\infty}\int_{A_k} R(x)dx \\ &\sim \sum_{k=1}^{\infty}(k+1)k^{-1/\alpha-1} \sim \sum_{k=1}^{\infty} k^{-1/\alpha},\end{aligned}$$

so if $\alpha \geq 1$, then $\mathbb{E}(R) = \infty$ (recall that the A_k's are the intervals of the right part of the partition $\mathcal{P}$). But if we set R to zero if it gets too large, then it will have finite expectation. More precisely, we define the "cut-off" return times

$$R_k = R \cdot \mathbb{I}_{\{R<a_k\}}$$

for some numbers a_k, so that $\mathbb{E}(R_k) < \infty$. The numbers a_k can be chosen so that $R \circ F^k = R_k \circ F^k$ almost surely for all but a finite number of k's. At this point, it is necessary to divide the calculations according to whether $\alpha = 1$ or $\alpha > 1$. First assume the latter; the former case will be dealt with below. For $\alpha > 1$ and $k > 1$, set

$$a_k = k^\alpha(\log k)^q$$

for some $q > \alpha$. (For $k = 0, 1$, we can set $a_k = 1$ for example, or something else finite. Some of the statements below will only hold for $k > 1$, but the first two terms will not matter). Then

$$\nu(\{R \circ F^k > a_k\}) = \nu(\{R > a_k\}) \sim \sum_{n=\lceil a_k\rceil}^{\infty}\nu(A_n) \tag{5.4}$$

$$\sim \sum_{n=\lceil a_k\rceil}^{\infty} n^{-1/\alpha-1} \sim a_k^{-1/\alpha} = \frac{1}{k(\log k)^{q/\alpha}} \tag{5.5}$$

($\lceil a \rceil$ denotes the smallest integer that is at least as big as a), and so

$$\sum_{k=0}^{n-1} \nu(\{R \circ F^k > a_k\}) \sim \sum_{k=2}^{n-1} \frac{1}{k(\log k)^{q/\alpha}}. \tag{5.6}$$

This series is known to converge when $q/\alpha > 1$, for the following reason: the primitive function of $1/(x(\log x)^{q/\alpha})$ is $(\log x)^{1-q/\alpha}/(1-q/\alpha)$ which goes to 0 as $x \to \infty$, and therefore the integral test implies that the series is convergent.

Now the Borel–Cantelli lemma tells us that with probability 1, $R \circ F^k > a_k$ for only a finite number of k's. Thus $R \circ F^k = R_k \circ F^k$ for all large enough k almost surely. Therefore, instead of working with the Birkhoff sums S_n^R, let us work with the sums

$$S_n(x) = \sum_{k=0}^{n-1} R_k(F^k(x)),$$

where we avoid the superscript for simplicity. For almost every $x \in Y$ there is an n_0 such that $R \circ F^k = R_k \circ F^k$ if $k > n_0$, which means that $S_n(x) = S_n^R(x) - C$ for some constant C (depending on x, but not on n) . Hence it suffices to show that $R_n \circ F^n(x) - S_n(x) > 0$ infinitely often and that this difference becomes arbitrarily large, since then the constant C will not matter. For simpler notation, write $X_k = R_k \circ F^k$. We will first show that the events

$$B_n = \{S_n < u_1\mathbb{E}(S_n) < u_2\mathbb{E}(S_n) < X_n\}$$

with positive probability happen infinitely often, where $u_1 < u_2$. This is the reason for working with sums of finite expectation.

If the events B_n were independent and their measures summed up to infinity, then the result would follow from the classical Borel–Cantelli lemma. However, the B_n's are not independent, so it is necessary to use a generalization of the Borel–Cantelli lemma. We will use one proven by Petrov.

Theorem 5.2 ([11]). *Let μ be a probability measure and let $A_1, A_2, \ldots$ be a sequence of events satisfying $\sum_{k=1}^{\infty} \mu(A_k) = \infty$. Assume $H \geq 1$. If*

$$\liminf_n \frac{\sum\limits_{i,j=1}^{n} \mu(A_j \cap A_i) - H\mu(A_i)\mu(A_j)}{\left(\sum\limits_{i=1}^{n} \mu(A_i)\right)^2} \leq 0,$$

then $\mu(\limsup A_n) \geq 1/H$.

This is a more general version of a much earlier result from a paper in 1959 by Erdős and Rényi, in which they proved the special case of $H = 1$ [12]. For the case $H = 1$, it is clear to see that the numerator estimates how close the events A_i and A_j are to being independent; if all events are pairwise independent, then the numerator will be zero.

Let us first show that $\sum_n B_n = \infty$. The expectation of X_n is

$$\mathbb{E}(X_n) \sim \sum_{k=1}^{\lceil a_n \rceil} k \cdot \nu(A_k) \sim \sum_{k=1}^{\lceil a_n \rceil} k^{-1/\alpha} \sim \int_1^{a_n} x^{-1/\alpha} dx \sim a_n^{-1/\alpha+1} - 1 \sim n^{\alpha-1}(\log n)^{q(1-1/\alpha)}. \tag{5.7}$$

The expectation of S_n then becomes

$$\mathbb{E}(S_n) = \sum_{k=0}^{n-1} \mathbb{E}(X_k) \sim \sum_{k=1}^{n-1} k^{\alpha-1}(\log k)^{q(1-1/\alpha)} \leq \sum_{k=1}^{n-1} n^{\alpha-1}(\log n)^{q(1-1/\alpha)} \tag{5.8}$$

$$= n^{\alpha}(\log n)^{q(1-1/\alpha)}. \tag{5.9}$$

The events B_n consist of two separate events,

$$E_n = \{u_2\mathbb{E}(S_n) < X_n\}$$

and

$$F_n = \{S_n < u_1\mathbb{E}(S_n)\}.$$

Moreover, E_n can be written as $E_n = F^{-n}(K_n)$, where

$$K_n = \{x \in Y : u_2\mathbb{E}(S_n) < R(x) < a_n\}.$$

The events E_n and F_n are unfortunately not independent, but they are close enough, in the sense of the following lemma.

Lemma 5.3. *$\nu(B_n) \sim \nu(E_n) \cdot \nu(F_n)$.*

Proof. First, note that S_n is constant och each $I \in \mathcal{P}_n$. This means that F_n is a union of elements in $\mathcal{P}_n$. Next, let $I \in \mathcal{P}_n$. Because of uniformly bounded distortion, i.e.

$$\frac{1}{M} \leq \frac{(F^n)'(x)}{(F^n)'(y)} \leq M$$

(lemma 4.5), the length of $I \cap E_n$ can be bounded between two extremes. The one extreme is that $F^n|_I$ is a straight line with some slope t_1 on $I \setminus E_n$, and a straight line with slope Mt_1 on $I \cap E_n$, in which case $|I \cap E_n| = |K_n|/(Mt_1)$. The other extreme is that $F^n|_I$ is a straight line with some slope t_2 on $I \cap E_n$, and a straight line with slope Mt_2 on $I \setminus E_n$, meaning $|I \cap E_n| = |K_n|/t_2$. The slopes t_1 and t_2 must satisfy $|Y|/(M|I|) \leq t_1, t_2 \leq |Y|/|I|$. Therefore,

$$\frac{|K_n| \cdot |I|}{M|Y|} \leq |I \cap E_n| \leq \frac{M \cdot |K_n| \cdot |I|}{|Y|}.$$

Since these bounds hold for every n and every $I \in \mathcal{P}_n$, and since F_n is a union of intervals in $\mathcal{P}_n$,

$$|B_n| \sim |F_n| \cdot |K_n|.$$

Finally, the F-invariance of ν and the fact that $\nu(A) \sim |A|$ for all measurable sets A imply that $\nu(B_n) \sim \nu(E_n) \cdot \nu(F_n)$. □

The next step in order to show that $\sum_n \nu(B_n) = \infty$ is then to estimate $\nu(E_n)$ and $\nu(F_n)$. Let us start with the latter. We will use Markov's inequality, which states that if X is a non-negative random variable and $t > 0$, then $\nu(\{X > t\}) \leq \mathbb{E}(X)/t$ [4]. It implies that

$$\nu(F_n) = 1 - \nu(\{S_n > u_1\mathbb{E}(S_n)\}) \geq 1 - \frac{\mathbb{E}(S_n)}{u_1\mathbb{E}(S_n)} = 1 - \frac{1}{u_1}.$$

If we pick an $u_1 > 1$, then $\nu(F_n)$ is bounded away from zero.

The probability $\nu(E_n)$ can be estimated as follows:

$$\nu(E_n) = \nu\left(\{u_2\mathbb{E}(S_n) < X_n\}\right) \sim \sum_{k=\lceil u_2\mathbb{E}(S_n)\rceil}^{\lceil a_n\rceil} \nu(A_k) \sim \sum_{k=\lceil u_2\mathbb{E}(S_n)\rceil}^{\lceil a_n\rceil} k^{-1-1/\alpha} \tag{5.10}$$

$$\sim \int_{u_2\mathbb{E}(S_n)}^{a_n} x^{-1-1/\alpha}dx \sim (u_2\mathbb{E}(S_n))^{-1/\alpha} - a_n^{-1/\alpha} \tag{5.11}$$

$$\gtrsim n^{-1}(\log n)^{-(q-q/\alpha)/\alpha} - n^{-1}(\log n)^{-q/\alpha} \tag{5.12}$$

$$\sim n^{-1}(\log n)^{-(q-q/\alpha)/\alpha}. \tag{5.13}$$

Using lemma 5.3,

$$\sum_{k=1}^{n} \nu(B_n) \sim \sum_{k=1}^{n} \nu\left(E_k\right)\nu(F_k) \geq \left(1 - \frac{1}{u_1}\right)\sum_{k=1}^{n} \nu\left(E_k\right) \gtrsim \sum_{k=1}^{n} k^{-1}(\log k)^{-(q-q/\alpha)/\alpha}. \tag{5.14}$$

If we pick the q such that $(q - q/\alpha)/\alpha = 1$, then this series diverges as $n \to \infty$. There is enough margin to pick such a q, since the only previous requirement was that $q > \alpha$.

If $\alpha = 1$, we modify the calculations by first setting

$$a_n = n^2.$$

The aim is still to show that $\sum_n B_n = \infty$, and we do it in basically the same way. It is the labeled calculations (5.4)-(5.14) that need to be changed, and the rest of the arguments stay the same. We get

$$\nu(\{R \circ F^k > a_k\}) \sim \frac{1}{a_k} = \frac{1}{k^2},$$

and so

$$\sum_{k=1}^{n} \nu(\{R \circ F^k > a_k\}) \sim \sum_{k=1}^{n} \frac{1}{k^2},$$

which converges. This again shows that $R \circ F^k = R_k \circ F^k$ for all large enough k almost surely.

The expected value of X_n for $\alpha = 1$ is calculated as

$$\mathbb{E}(X_n) \sim \sum_{k=1}^{\lceil a_n\rceil} \frac{1}{k} \sim \int_1^{a_n} \frac{1}{x}dx = \log a_n = 2\log n.$$

Thus the expected value of S_n becomes

$$\mathbb{E}(S_n) = \sum_{k=1}^{n} \mathbb{E}(X_k) \sim \sum_{k=1}^{n} \log k \leq n\log n.$$

Then, reusing the calculations in (5.10)-(5.11),

$$\nu(E_n) \sim \frac{1}{u_2\mathbb{E}(S_n)} - \frac{1}{a_n} \gtrsim \frac{1}{n\log n} - \frac{1}{n^2} \sim \frac{1}{n\log n}. \tag{5.15}$$

Then

$$\sum_{k=1}^{n} \nu(B_n) \sim \sum_{k=1}^{n} \nu\left(E_k\right) \nu(F_k) \geq \left(1 - \frac{1}{u_1}\right) \sum_{k=1}^{n} \nu\left(E_k\right) \gtrsim \sum_{k=1}^{n} \frac{1}{n \log n},$$

which diverges. Now all the differences between the cases $\alpha = 1$ and $\alpha > 1$ have been taken care of.

Let us estimate $\nu(B_i \cap B_j) - H\nu(B_i)\nu(B_j)$ for some H. For this, it does not matter if $\alpha = 1$ or $\alpha > 1$. According to lemma 5.3, there is a constant c such that $\nu(B_n) \geq c\nu(E_n)\nu(F_n)$ for all n. Hence

$$\begin{aligned} \nu(B_i \cap B_j) - H\nu(B_i)\nu(B_j) &\leq \nu(E_i \cap E_j) - H\nu(B_i)\nu(B_j) \\ &\leq \nu(E_i \cap E_j) - c^2 H\nu(E_i)\nu(E_j)\nu(F_i)\nu(F_j). \end{aligned}$$

Since

$$\nu(F_n) \geq 1 - \frac{1}{u_1},$$

we can pick an H big enough that $c^2 H\nu(F_i)\nu(F_j) \geq 1$. Then we get

$$\nu(B_i \cap B_j) - H\nu(B_i)\nu(B_j) \leq \nu(E_i \cap E_j) - \nu(E_i)\nu(E_j).$$

The right hand side can be written as

$$\nu(E_i \cap E_j) - \nu(E_i)\nu(E_j) = \int (\mathbb{I}_{K_j} \circ T^j) \cdot (\mathbb{I}_{K_i} \circ T^i) d\nu - \int \mathbb{I}_{K_j} d\nu \int \mathbb{I}_{K_i} d\nu,$$

which, if $j > i$, is the same as

$$\int (\mathbb{I}_{K_j} \circ T^{j-i}) \cdot \mathbb{I}_{K_i} d\nu - \int \mathbb{I}_{K_j} d\nu \int \mathbb{I}_{K_i} d\nu.$$

Kim [13] used a result of Rychlik [6] to show that if T is piecewise expanding and $1/T'$ has bounded variation, then there are constants $C > 0$ and $0 < r < 1$ such that for all $n > 0$, all $f \in L^1(\nu)$ and all ϕ with bounded variation,

$$\int (f \circ T^n) \cdot \phi \, d\nu - \int f d\nu \int \phi \, d\nu \leq C r^n \|f\|_1 \cdot \|\phi\|_{BV}. \tag{5.16}$$

Here the BV-norm is defined by $\|\phi\|_{BV} = \|\phi\|_1 + V(\phi)$, and it can be proven that it really is a norm. We have $\|\mathbb{I}_{K_n}\|_1 = \nu(K_n) = \nu(E_n) \to 0$ and $V(\mathbb{I}_{K_n}) = 4$ (K_n consists of two intervals), so the BV-norm becomes:

$$\|\mathbb{I}_{K_n}\|_{BV} = \|\mathbb{I}_{K_n}\|_1 + V(\mathbb{I}_{K_n}) \sim 1.$$

So by (5.16) there exists a $C > 0$ and an r, $0 < r < 1$, such that, if $j > i$,

$$\int (\mathbb{I}_{K_j} \circ T^{j-i}) \cdot \mathbb{I}_{K_i} d\nu - \int \mathbb{I}_{K_j} d\nu \int \mathbb{I}_{K_i} d\nu \leq C r^{j-i} \nu(E_j).$$

Then we get

$$\begin{aligned}\sum_{i,j=1}^{n} \nu(B_i \cap B_j) - H\nu(B_i)\nu(B_j) &\leq \sum_{i,j=1}^{n} \nu(E_i \cap E_j) - \nu(E_i)\nu(E_j) \\ &= 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \nu(E_i \cap E_j) - \nu(E_i)\nu(E_j) + \sum_{i=1}^{n} \nu(E_i) - \nu(E_i)^2 \\ &\leq 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} Cr^{j-i}\nu(E_j) + \sum_{i=1}^{n}\nu(E_i) \leq 2\sum_{i=1}^{n} C\frac{r}{1-r}\nu(E_i) + \sum_{i=1}^{n}\nu(E_i) \\ &= D\sum_{i=1}^{n}\nu(E_i),\end{aligned}$$

where $D = 1 + 2Cr/(1-r)$.

For the numerator in theorem 5.2,

$$\left(\sum_{i=1}^{n}\nu(B_i)\right)^2 \geq \left(\sum_{i=1}^{n} c\nu(E_i)\nu(F_i)\right)^2 \geq \left(\sum_{i=1}^{n} c\frac{1}{1-u_1}\nu(E_i)\right)^2 = \left(\frac{c}{1-u_1}\right)^2\left(\sum_{i=1}^{n}\nu(E_i)\right)^2.$$

Hence

$$\frac{\sum_{i,j=1}^{n} \nu(B_i \cap B_j) - H\nu(B_i)\nu(B_j)}{\left(\sum_{i=1}^{n}\nu(B_i)\right)^2} \leq \frac{D(1-u_1)^2}{c^2}\cdot\frac{1}{\sum_{i=1}^{n}\nu(E_i)} \to 0 \quad \text{as } n\to\infty.$$

This proves that the B_n's satisfy the conditions of theorem 5.2, and therefore $\nu(\limsup B_n) \geq 1/H > 0$.

Although $\limsup B_n$ does not necessarily have probability 1, we can use ergodicity to show that something similar to B_n almost surely happens infinitely often. Because F is ergodic, theorem 3.3 says that for almost every $x \in Y$ there is a $k > 0$ such that $F^k(x) \in \limsup B_n$. This means that

$$S_n(x) - S_k(x) < u_1\mathbb{E}(S_{n-k}) < u_2\mathbb{E}(S_{n-k}) < X_n(x) \tag{5.17}$$

infinitely often for $n > k$. Since $S_k(x)$ does not depend on n, this is similar enough to B_n.

Important for theorem 5.1 is that (5.17) happens infinitely often both for $F^n(x) > 1/2$ and $F^n(x) < 1/2$. This can be shown by replacing the events B_n by $B_n^r = B_n \cap \{F^n(x) \in (1/2, b)\}$ or by $B_n^\ell = B_n \cap \{F^n(x) \in (a, 1/2)\}$, which by the symmetry of F are both half the measure of B_n. This changes nothing essential, so by the same arguments as for B_n, B_n^r and B_n^ℓ both happen infinitely often with positive probability. Thus (5.17) ν-almost surely happens infinitely often for both $F^n(x) < 1/2$ and $F^n(x) > 1/2$.

Now let us see what this says about E_n^ϕ. Take an n where (5.17) happens, and set $m = \sum_{i=0}^{n-1} R(F^i(x)) = S_n^R(x)$; this is the m for which $S_n^\psi(x) = S_m^\phi(x)$. Let $N = m + R(F^n(x))$, which for large n is equal to $m + X_n(x)$, and consider $E^\phi_N(x)$. Assume $\psi(F^n(x)) > 0$, i.e. $F^n(x) < 1/2$. Then $R(F^n(x)) = \psi(F^n(x)) + 2$, and by using this and the fact that $S^\psi_n(x) \geq -S^R_n(x) + 2n$, we get

$$
\begin{aligned}
E^\phi_N(x) = \frac{S^\phi_N(x)}{N} &= \frac{S^\psi_n(x) + \psi(F^n(x))}{X_n(x) + S^R_n(x)} \geq \frac{-S^R_n(x) + 2n + X_n(x) - 2}{X_n(x) + S^R_n(x)} \\
&\geq \frac{X_n(x) - S^R_n(x)}{X_n(x) + S^R_n(x)} = 1 - \frac{2S^R_n(x)}{X_n(x) + S^R_n(x)} = 1 - \frac{2(S_n(x) + C)}{X_n(x) + S_n(x) + C} \\
&\geq 1 - \frac{2(S_n(x) + C)}{X_n(x) + S_n(x)},
\end{aligned}
$$

where C is the number such that $S^R_n(x) = S_n(x) + C$ for all large enough n. Combining this with (5.17), we get

$$
E^\phi_N(x) \geq 1 - \frac{2(S_n(x) + C)}{X_n(x) + S_n(x)} \geq 1 - \frac{2(S_n(x) + C)}{u_2\mathbb{E}(S_{n-k}) + S_n(x)}.
$$

Once $u_2\mathbb{E}(S_{n-k})$ becomes larger than C, the right hand side will be decreasing as a function of $S_n(x)$. By (5.17), $S_n(x) < S_k(x) + u_1\mathbb{E}(S_{n-k})$, so

$$
E^\phi_N(x) \geq 1 - \frac{2(u_1\mathbb{E}(S_{n-k}) + S_k(x) + C)}{(u_1 + u_2)\mathbb{E}(S_{n-k}) + S_k(x)} \geq 1 - \frac{2u_1}{u_1 + u_2} - \frac{2(S_k(x) + C)}{(u_1 + u_2)\mathbb{E}(S_{n-k})}
$$

In conclusion: if (5.17) holds with $F^n(x) < 1/2$ and a large enough n, then $E^\phi_N(x) \geq 1 - u_1/(u_1 + u_2) - c_n$, where $c_n \to 0$. By increasing u_2 and n, we can make $E^\phi_N(x)$ as close to 1 as we want. By doing a similar argument for $F^n(x) > 1/2$, we can in that case also make $E^\phi_N(x)$ as close to -1 as we want. Since (5.17) happens infinitely often with both $F^n(x) < 1/2$ and $F^n(x) > 1/2$, this proves that $\liminf E^\phi_n(x) = -1$ and $\limsup E^\phi_n(x) = 1$ for μ-almost every x. And since μ is equivalent to the Lebesgue measure, this proves theorem 5.1.

We mention a connection with a result from Galatolo, Holland, Persson and Zhang [14]. They investigate growth speeds of Birkhoff sums, and of the maximum process

$$
M^\phi_n(x) = \max_{0 \leq k \leq n-1} \phi(T^k(x))
$$

for maps T preserving a probability measure μ and non-integrable functions ϕ: $\int \phi d\mu = \infty$. They study a certain kind of piecewise expanding dynamical systems called Gibbs–Markov systems — see their paper for the definition. For $\alpha \geq 1$, proposition 2.8 in their paper about Gibbs–Markov systems applies to F, and says that for each $\epsilon > 0$ and almost all $x \in Y$, there is an n_0 such that for all $n > n_0$,

$$
n^\alpha (\log n)^{-\alpha-\epsilon} \leq M^R_n(x) < S^R_n(x) \leq n^\alpha (\log n)^{\alpha+\epsilon}.
$$

In particular, since $M^R_n(x)$ grows to infinity, $R(F^k(x))$ is greater than all previous terms infinitely often. We have proven something even stronger for F, namely that $R(F^k(x))$ infinitely often is greater than the sum of all previous terms.

5.4 Frequency of oscillations

We have shown that the events $C_n = \{X_n > S_n\}$ happen infinitely often almost surely, but not how frequently they occur. They may occur quite often, or extremely seldom. The sum $\sum_{k=1}^n \mathbb{I}_{C_k}(x)$ counts how many times C_k happens up to time n. A natural hypothesis is that the number of C_k's that occur before time n is approximately equal to the expected value of this sum, or in other words, that we almost surely have

$$\frac{\sum_{k=1}^n \mathbb{I}_{C_k}(x)}{\mathbb{E}\left(\sum_{k=1}^n \mathbb{I}_{C_k}\right)} = \frac{\sum_{k=1}^n \mathbb{I}_{C_k}(x)}{\sum_{k=1}^n \nu(C_k)} \to 1 \quad \text{as } n \to \infty. \tag{5.18}$$

This, unfortunately, seems to be hard to prove. We will therefore content ourselves with proving it for the events $E_n = \{X_n > u_2\mathbb{E}(S_n)\}$ instead of C_n. This is not as interesting a result, since it does not show how often oscillations occur, but it is at least an indication that (5.18) might be correct.

The classical Borel–Cantelli lemma gives a condition for the events in a sequence $(A_n)_{n=1}^\infty$ to occur infinitely often. When studying dynamical systems, one is often interested in whether the event $F^n(x) \in A_n$ occurs infinitely often. Assume $\sum_n \nu(A_n) = \infty$. If $F^n(x) \in A_n$ infinitely often almost surely, then $(A_n)_{n=1}^\infty$ is called a Borel–Cantelli sequence, BC for short. If

$$\frac{\sum_{k=1}^n \mathbb{I}_{A_k} \circ F^k(x)}{\sum_{k=1}^n \nu(A_k)} \to 1$$

for almost every x, then $(A_n)_{n=1}^\infty$ is called a Strong Borel–Cantelli (SBC) sequence. An SBC sequence is a BC sequence where we know how often $F^n(x) \in A_n$. The terms BC and SBC sequence were introduced by Chernov and Kleinbock in [15]. We have shown that K_n is a BC sequence (recall that $F^{-n}(K_n) = E_n$), but it is also an SBC sequence. This follows immediately from the following theorem proved by Kim [13]. Just as above, $\|f\|_{BV} = \|f\|_1 + V(f)$.

Theorem 5.4 ([13])**.** *Let T be a piecewise expanding map with $1/|T'|$ of bounded variation. Assume T has a unique absolutely continuous invariant measure μ with density bounded away from 0. If $(A_n)_{n=1}^\infty$ is a sequence of sets with $\sum \mu(A_n) = \infty$ and $\|\mathbb{I}_{A_n}\|_{BV} < M$ for all n for some M, then $(A_n)_{n=1}^\infty$ is an SBC sequence.*

The map F satisfies the conditions of the theorem, as do the sets K_n, since K_n is just two intervals. Hence K_n is an SBC sequence, and therefore we almost surely have

$$\frac{\sum_{k=1}^n \mathbb{I}_{E_k}(x)}{\sum_{k=1}^n \nu(E_k)} = \frac{\sum_{k=1}^n \mathbb{I}_{K_k} \circ F^k(x)}{\sum_{k=1}^n \nu(K_k)} \to 1 \quad \text{as } n \to \infty.$$

How fast does the denominator grow? For $\alpha > 1$ and $\alpha = 1$ equations (5.13) and (5.15) respectively give the lower bound $\nu(E_n) \gtrsim n^{-1}(\log n)^{-1}$, and it is not too hard to show that $\gtrsim$ can be replaced by $\sim$. Thus

$$\sum_{k=1}^n \nu(E_k) \sim \sum_{k=1}^n \frac{1}{k \log k} \sim \int_1^n \frac{1}{x \log x} dx = \log(\log n).$$

Hence $\sum_{k=1}^n \mathbb{I}_{E_k}(x)$ grows at the very slow rate of $\log(\log n)$. In other words, E_n occurs extremely rarely. A further research topic might be to show this result for C_n instead of E_n.

6

Bibliography

[1] Carlangelo Liverani, Benoît Saussol, and Sandro Vaienti. A probabilistic approach to intermittency. *Annalen der Physik*, 19:671–685, 1999.

[2] Yves Pomeau and Paul Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Communications in Mathematical Physics*, 74:189–197, 1980.

[3] Cesar E. Silva. *Invitation to Ergodic Theory*. American Mathematical Society, 2008.

[4] Donald L. Cohn. *Measure Theory, Second Edition*. Birkhäuser, 2013.

[5] R Clark Robinson. *An introduction to Dynamical Systems: Continuous and Discrete, Second Edition*. American Mathematical Society, 2012.

[6] Marek Rychlik. Bounded variation and invariant measures. *Studia Mathematica*, 76: 69–80, 1983.

[7] Roy Adler. F-expansions revisited. *Springer lecture notes*, 318:1–5, 1973.

[8] Chinmaya Gupta, Matthew Nicol, and William Ott. A borel-cantelli lemma for non-uniformly expanding dynamical systems. *Nonlinearity*, 23, 2010.

[9] Noella Grady. Functions of bounded variation. `https://www.whitman.edu/Documents/Academics/Mathematics/grady.pdf`.

[10] Roland Zweimüller. Invariant measures for general(ized) induced transformations. *Proceedings of the American Mathematical Society*, 133:2283–2295, 2005.

[11] Valentin V. Petrov. A generalization of the borel–cantelli lemma. *Statistics Probability Letters*, 67:233–239, 2004.

[12] Paul Erdős and Alfréd Rényi. On cantor's series with convergent $\sum 1/q_n$. *Ann. Univ. Sci. Budapest Sect. Math.*, 2:93–109, 1959.

[13] Dong Han Kim. The dynamical borel-cantelli lemma for interval maps. *Discrete and Continuous Dynamical Systems*, 17:891–900, 2007.

[14] Stefano Galatolo, Mark Holland, Tomas Persson, and Yiwei Zhang. Birkhoff sums of infinite observables and anomalous time-scaling of extreme events in infinite systems, 2018. arXiv:1810.10742.

[15] Nikolai Chernov and Dmitry Kleinbock. Dynamical borel-cantelli lemmas for gibbs measures. *Israel J. Math.*, 122:1–27, 2001.