INTRODUCTION TO SMOOTH ERGODIC THEORY LECTURE NOTES

STEFANO LUZZATTO

Contents

Date: DRAFT: version October 3, 2013.

1. Introduction

1.1. **Dynamical Systems.** Let M be a set and

$$f : M \to M$$

be a map. We think of M as a "phase space" of possible states of the system, and the map f as the "law of evolution" of the system. Then, given an "initial condition" $x_0 \in M$ we have a sequence given by $x_1 = f(x), x_2 := f^2(x) = f(f(x)) = f \circ f(x)$ and generally $x_n := f^n(x) = f \circ \cdots \circ f(x)$ given by the n'th composition of the map f with itself.

The main goal of the theory of Dynamical Systems is to describe tand classify the possible structures which arise from the he iteration of such maps. If $x \in M$ then we let

$$\mathcal{O}^+(x) := \{f^n(x)\}_{n \geq 0}$$

denote the (forward) *orbit* or *trajectory* of the point x. The simplest kind of orbit is when x happens to be a *fixed point*, i.e. $f(x) = x$ in which case of course the whole forward orbit reduces to the point x, i.e. $\mathcal{O}^+(x) = \{x\}$. The next simplest kind of orbit is the case in which there exists some $k > 0$ such that $f^k(x) = x$. The point x is then called a *periodic point* and the minimal $k > 0$ for which $f^k(x) = x$ is called the (minimal) *period* of x. Then the forward orbit of the point x is just the finite set $\mathcal{O}^+(x) = \{x, f(x), ..., f^{k-1}(x)\}$. Notice that a fixed point is just a special case of a periodic orbit with $k = 1$. Fixed and periodic orbits are very natural structures and a first approach to the study of dynamical systems is to study the existence of fixed and periodic orbits. Such orbits however generally do not exhaust all the possible structures in the system and we need some more sophisticated tools and concentps.

1.2. **Topological and probabilistic limit sets.** If the orbit of x is not periodic, then $\mathcal{O}^+(x)$ is a countable set and the problem of describing this set becomes non-trivial (in practice even describing a periodic orbit can be non trivially in specific situations, especially if the orbit is large, but at least in these cases it is at least theoretically possible to describe it completely by identifying the finite points on the orbit). Generally we need to have additional properties on the set M. For example if M is a topological space then we can define the *omega-limit* of a point x as

$$\omega(x) := \{y : f^{n_j}(x) \to y \text{ for some } n_j \to \infty\}.$$

The simplest kind of omega-limit is a single point p, so that the iterates $f^n(x)$ converge to p as $n \to \infty$, however omega-limits can also be geometrically extremely complicated sets with fractal structure, or they can even be the entire set if $\mathcal{O}^+(x)$ is dense in M (notice that $\mathcal{O}^+(x)$ can not itself be the entire set Msince this is just a countable set).

Considering in addition the measurable structure on M given by the Borel sigma-algebra, we can also use measures to describe orbits. Indeed, let

$$\mathcal{M} := \{\mu : \mu \text{ is a (Borel) probability measure on } M\}.$$

It is clear that $\mathcal{M} \neq \emptyset$ since it contains for example all Dirac-δ measures. We recall that the Dirac-δ measure δ_x on a point $x \in M$ is defined by $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$

if $x \notin A$. For any $x \in M$ we can define a probabilistic analogue of the ω-limit set by

$$\omega_{prob}(x) := \left\{ \mu \in \mathcal{M} : \frac{1}{n_j} \sum_{i=0}^{n-1} \delta_{f^i(x)} \to \mu \text{ for some } n_j \to \infty \right\}$$

The convergence above is meant in the weak-star topology. We recall that by definition $\mu_{n_j} \to \mu$ if and only if $\int \varphi d\mu_{n_j} \to \int \varphi d\mu$ for all $\varphi \in C^0(M, \mathbb{R})$. In the particular case in which the sequence μ_{n_j} is given by the form above, we have

$$\int \varphi d\left(\frac{1}{n_j} \sum_{i=0}^{n_j-1} \delta_{f^i(x)} \right) = \frac{1}{n} \sum_{i=0}^{n_j-1} \int \varphi d\delta_{f^i(x)} = \frac{1}{n_j} \sum_{i=0}^{n_j-1} \int \varphi(f^i(x)) = \frac{1}{n_j} \sum_{i=0}^{n_j-1} \int \varphi \circ f^i(x)$$

The sum on the right is sometimes called the *time average* of the "observable" φ along the orbit of x. From this, we can rewrite the definition of probabilistic limit set of x as

$$\left\{ \mu \in \mathcal{M} : \frac{1}{n_j} \sum_{i=0}^{n_j-1} \int \varphi \circ f^i(x) \to \int \varphi d\mu \text{ for all } \varphi \in C^0(M, \mathbb{R}), \text{ for some } n_j \to \infty \right\}$$

Since M is compact, $\mathcal{M}$ is also compact and so $\omega_{prob}(x) \neq \emptyset$. However, as we will discuss later, there is a crucial difference between the cases in which the sequence μ_n actually converges, and therefore $\omega_{prob}(x)$ is a single probability measure, and the case in which it has several limit points. We will therefore be particularly interested in establishing situations in which probabilistic omega limit set is a single probability measure. To address this question we start with a probability measure $\mu \in \mathcal{M}$ and define the *"basin of attraction"*

$$\mathfrak{B}_\mu := \left\{ x \in M : \frac{1}{n} \sum_{i=0}^{n-1} \int \varphi \circ f^i(x) \to \int \varphi d\mu \quad \forall\, \varphi \in C^0(M, \mathbb{R}) \right\}.$$

By definition, any point $x \in \mathfrak{B}_\mu$ has the property that its asymptotic distribution in space is well described by the measure μ. We will investigate conditions below which guarantee that $\mathfrak{B}_\mu$ is non-empty and even large in a certain sense.

1.3. **Physical measures.** If the space M of our dynamical systems has an underlying reference measure, such as *Lebesgue measure*, then it is a particularly interesting question to find a measure μ whose basin has positive or even full Lebesgue measure since this means that we are able to describe the the asymptotic distribution of a large set of points. Motivated by this observation we make the following definition.

Definition 1. A probability measure $\mu \in \mathcal{M}$ is called a **physical measure** if $Leb(\mathfrak{B}_\mu) > 0$.

The question which motivates a lot of the material in this course is therefore:

when does a dynamical system admit a physical measure?

and, if it does admit physical measures,

how many physical measures does it have?

Note that by definition a system can have at most a countable number of physical measure, but each of these is in some sense an "attractor" for a certain set of points and therefore the question is about the number of possible asymptotic distributions of "most" points.

This turns out to be a very challenging problem that has not yet been solved in general and there are many examples of systems which do not have physical measures,for example the identity map. However it is "hoped" that this is an exceptional situation.

Conjecture 1 (Palis conjecture)**.** *Most systems have a finite number of physical measures such that the union of their basins has full Lebesgue measure.*

In these notes we give an introduction to some of the results and techniques which have been developed in this direction. In Section 2 we introduce the notion of *invariant* measure and of *ergodic* measure and show that such measures exist under some very mild assumptions on the dynamical systems. In Section 3 we study the basins of attractions of measures which are both invariant and ergodic and prove a fundamental Theorem of Birkhoff that these basins are non-empty and are even "large" in the sense that $\mu(\mathfrak{B}_\mu) = 1$. In Sections 4 and 5 we consider two examples of one-dimensional dynamical systems (irrational circle rotations and piecewise affine full branch expanding maps) in which Lebesgue measure itself is invariant and ergodic and therefore, by Birkhoff's theorem, is a physical measure. In Section 6 we define a much more general class of full branch maps which are not piecewise affine but have a "bounded distortion" property, and show that for these maps Lebesgue measure is still ergodic even if it is not invariant.

2. The Space of Invariant and Ergodic Measures

2.1. **Definitions and basic examples.**

Definition 2. Let $\mu \in \mathcal{M}$.

(1) μ is *f-invariant* if $\mu(f^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}$;
(2) μ is *ergodic* if $f^{-1}(A) = A$ and $\mu(A) > 0$ implies $\mu(A) = 1$ for all $A \in \mathcal{B}$.

These two notions are completely independent of each other, as a measure can be invariant without being ergodic and can be ergodic without being invariant.

Exercise 1. A set $A \subseteq M$ is *fully invariant* if $f^{-1}(A) = A$. Show that if A is fully invariant, letting $A^c := M \setminus A$ denote the complement of A, then $f^{-1}(A^c) = A^c$ and that both $f(A) = A$ and $f(A^c) = A^c$.

Exercise 2. Show that if f is invertible then then μ is invariant if and only if $\mu(f(A)) = \mu(A)$. Find an example of a non-invertible map and a measure μ for which the two conditions are not equivalent.

We will however be particularly interested in measures which are both ergodic and invariant. We give a few simple example here, further examples will be studied in detail in the following sections.

Example 1. Let X be a measure space and $f : X \to X$ a measurable map. Suppose $f(p) = p$. Then the *Dirac* measure

$$\delta_p(A) := \begin{cases} 1 & p \in A \\ 0 & p \notin A \end{cases}$$

is invariant and ergodic. Indeed, let $A \subset I$ be a measurable set. We consider two cases. For the first case, suppose $p \in A$, then $\delta_p(A) = 1$. In this case we also clearly have $p \in f^{-1}(A)$ (notice that p might have multiple preimages, but the point p itself is certainly one of them). Therefore $\delta_p(f^{-1}(A) = 1$, and the result is proved in this case. For the second case, suppose $p \notin A$. Then $\delta_p(A) = 0$ and in this case we also have $p \notin f^{-1}(A)$. Indeed, if we did have $p \in f^{-1}(A)$ this would imply, by definition of $f^{-1}(A) = \{x : f(x) \in A\}$, that $f(p) \in A$ contradicting our assumption. Therefore we have $\delta_p(f^{-1}(A)) = 0$ proving invariance in this case. Ergodicity is trivial in this case.

Example 2. An immediate generalization is the case of a measure concentrated on a finite set of points $\{p_1, \dots, p_n\}$ each of which carries some proportion $\rho_1, \dots, \rho_n$ of the total mass, with $\rho_1 + \cdots + \rho_n = 1$. Then, we can define a measure δ_P by letting

$$\delta_P(A) := \sum_{i : p_i \in A} \rho_i.$$

Then δ_P is invariant if and only if $\rho_i = 1/n$ for every $i = 1, \dots, n$. Also in this case, ergodicity follows automatically.

Example 3. Let $f : M \to M$ be the identity map, then every probability measure is invariant, but the only ergodic measures are the Dirac delta measures on the fixed points.

Example 4. Let $I = [0, 1]$, $\kappa \geq 2$ an integer, and let $f(x) = \kappa x \mod 1$. Then it is easy to see that Lebesgue measure is invariant. It is also ergodic but this is non trivial and we will prove it below. Notice that f also has an infinite number of periodic orbits and thus has an infinite number of ergodic invariant measures. We will show below that it actually has an uncountable number of distinct ergodic non-atomic invariant measures.

Example 5. Let $f : [0, 1] \to [0, 1]$ given by

$$f(x) = \begin{cases} .5 - 2x & \text{if } 0 \leq x < .25 \\ 2x - .5 & \text{if } .25 \leq x < .75 \\ -2x + 2.5 & \text{if } .75 \leq x \leq 1 \end{cases}$$

It is easy to see that the two halves of the interval are each fully invariant and therefore Lebesgue measure is invariant but not ergodic.

2.2. **The spaces of invariant and ergodic probability measures.** Recall that

$$\mathcal{M} := \{\text{probability measure on } \mathcal{M}\}.$$

We let

$$\mathcal{M}_f := \{\mu \in \mathcal{M} : \mu \text{ is } f\text{-invariant}\}$$

and

$$\mathcal{E}_f := \{\mu \in \mathcal{M} : \mu \text{ is } f\text{-invariant and ergodic.}\}$$

Clearly

$$\mathcal{E}_f \subseteq \mathcal{M}_f \subseteq \mathcal{M}.$$

There is no a priori reason for either $\mathcal{E}_f$ or $\mathcal{M}_f$ to be non-empty. However the following results show that this is indeed the case as long as M is compact and f is continuous. They also show that every invariant measure can be written as a combination of ergodic measures.

Theorem 1. *Let M be compact and $f : M \to M$ continuous. Then $\mathcal{M}_f$ is non-empty, convex, and compact. Moreover $\mathcal{E}_f$ is non-empty and $\mu \in \mathcal{E}_f$ if and only if μ is an extremal*[1] *point of $\mathcal{M}_f$.*

An immediate Corollary which follows from Choquet's Theorem for non-empty compact convex sets is the following

Corollary 2.1 (Ergodic decomposition)**.** *Let M be compact and $f : M \to M$ continuous. Then there exists a unique probability measure $\hat{\mu}$ on $\mathcal{M}_f$ such that $\mu(\mathcal{E}_f) = 1$ and such that for all $\mu \in \mathcal{M}_f$ and for all continuous functions $\varphi : M \to \mathbb{R}$ we have*

$$\int_M \varphi d\mu = \int_{\mathcal{E}_f} \left(\int_M \varphi d\nu \right) d\hat{\mu}$$

In the rest of this section we prove Theorem 1.

2.3. **Push-forward of measures.** We start with a key definition and some related results.

Definition 3 (Push-forward of measures)**.** Let

$$f : \mathcal{M} \to \mathcal{M}$$

be the map from the space of probability measures to itself, defined by

$$f_*\mu(A) := \mu(f^{-1}(A)). \tag{1}$$

We call $f_*\mu$ the *push-forward* of μ by f.

Exercise 3. $f_*\mu$ is a probability measure and so the map f_* is well defined.

It follows immediately from the definition that μ is invariant if and only if $f_*\mu = \mu$. Thus the problem of the existence of invariant measures is a problem of the existence of fixed points of f_*. We cannot however apply any general fixed point result, rather we will consider a sequence in $\mathcal{M}$ and show that any limit point is invariant. For any $\mu \in \mathcal{M}$ and any $i \geq 1$ we also let

$$f_*^i\mu(A) := \mu(f^{-i}(A)).$$

We now prove some simple properties of the map f_*.

[1]Recall that $\mathcal{M}_f$ is convex if given any $\mu_0, \mu_1 \in \mathcal{M}_f$, letting $\mu_t := t\mu_0 + (1-t)\mu_1$ for $t \in [0,1]$, then $\mu_t \in \mathcal{M}_f$. Moreover, an extremal point of a convex set A is a point μ such that if $\mu = t\mu_0 + (1-t)\mu_1$ for $\mu_0, \mu_1 \in \mathcal{M}_f$ with $\mu_0 \neq \mu_1$ then $t = 0$ or $t = 1$.

Lemma 2.1. *For all $\varphi \in L^1(\mu)$ we have $\int \varphi d(f_*\mu) = \int \varphi \circ f d\mu$.*

Proof. First let $\varphi = \mathbb{1}_A$ be the characteristic function of some set $A \subseteq X$. Then

$$\int \mathbb{1}_A d(f_*\mu) = f_*\mu(A) = \mu(f^{-1}(A)) = \int \mathbb{1}_{f^{-1}(A)} d\mu = \int \mathbb{1}_A \circ f d\mu.$$

The statement is therefore true for characteristic functions and thus follows for general integrable functions by standard approximation arguments. More specifically, it follows immediately that the result also holds if φ is a simple function (linear combination of characteristic functions). For φ a non-negative integrable function, we use the fact that every measurable function φ is the pointwise limit of a sequence φ_n of simple functions; if f is non-negative then φ_n may be taken non-negative and the sequence $\{\varphi_n\}$ may be taken increasing. Then, the sequence $\{\varphi_n \circ f\}$ is clearly also an increasing sequence of simple functions converging in this case to $\varphi \circ f$. Therefore, by the definition of Lebesgue integral we have $\int \varphi_n d(f_*\mu) \to \int \varphi d(f_*\mu)$ and $\int \varphi_n \circ f d\mu \to \int \varphi \circ f d\mu$ Since we have already proved the statement for simple functions we know that $\int \varphi_n d(f_*\mu) = \int \varphi_n \circ f d\mu$ for every n and therefore this gives the statement. For the general case we repeat the argument for positive and negative parts of φ as usual. □

Corollary 2.2. *$f_* : \mathcal{M} \to \mathcal{M}$ is continuous.*

Proof. Consider a sequence $\mu_n \to \mu$ in $\mathcal{M}$. Then, by Lemma 2.1, for any continuous function $\varphi : X \to \mathbb{R}$ we have

$$\int \varphi d(f_*\mu_n) = \int \varphi \circ f d\mu_n \to \int \varphi \circ f d\mu = \int \varphi d(f_*\mu)$$

which means exactly that $f_*\mu_n \to f_*\mu$ which is the definition of continuity. □

Corollary 2.3. *μ is invariant if and only if $\int \varphi \circ f d\mu = \int \varphi d\mu$ for all $\varphi : X \to \mathbb{R}$ cts.*

Proof. Suppose first that μ is invariant, then the implication follow directly from Lemma 2.1. For the converse implication, we have that

$$\int \varphi d\mu = \int \varphi \circ f d\mu = \int \varphi d f_*\mu$$

for every continuous function $\varphi : X \to \mathbb{R}$. By the Riesz Representation Theorem, measures correspond to linear functionals and therefore this can be restated as saying that $\mu(\varphi) = f_*\mu(\varphi)$ for all continuous functions $\varphi : X \to \mathbb{R}$, and therefore μ and $f_*\mu$ must coincide, which is the definition of μ being invariant. □

2.4. **Existence of invariant measures.**

Proposition 2.1 (Krylov-Boguliobov Theorem)**.** *$\mathcal{M}_f$ is non-empty, convex and compact.*

Proof. Recall first of all that the space $\mathcal{M}$ of probability measures can be identified with the unit ball of the space of functionals $C^*(M)$ dual to the space $C^0(M, \mathbb{R})$ of continuous functions on M. The weak-star topology is exactly the weak topology on the dual space and therefore, by the Banach-Alaoglu Theorem, $\mathcal{M}$ is weak-star compact if M is compact.

Our strategy therefore is to use the dynamics to define a sequence of probability measures in $\mathcal{M}$ and show that any limit measure of this sequence is necessarily invariant.

For an arbitrary $\mu_0 \in \mathcal{M}$ we define, for every $n \geq 1$,

$$\mu_n = \frac{1}{n} \sum_{i=0}^{n-1} f_*^i \mu_0. \tag{2}$$

Since each $f_*^i \mu_0$ is a probability measure, the same is also true for μ_n. By compactness of $\mathcal{M}$ there exists a measure $\mu \in \mathcal{M}$ and a subsequence $n_j \to \infty$ with $\mu_{n_j} \to \mu$. By the continuity of f_* we have $f_* \mu_{n_j} \to f_* \mu$. and therefore it is sufficient to show that also $f_* \mu_{n_j} \to \mu$. We write

$$\begin{aligned} f_* \mu_{n_j} &= f_* \left(\frac{1}{n_j} \sum_{i=0}^{n_j - 1} f_*^i \mu_0 \right) = \frac{1}{n_j} \sum_{i=0}^{n_j-1} f_*^{i+1} \mu_0 = \frac{1}{n_j} \left(\sum_{i=0}^{n_j-1} f_*^i \mu_0 - \mu_0 + f_*^{n_j} \mu_0 \right) \\ &= \frac{1}{n_j} \sum_{i=0}^{n_j-1} f_*^i \mu_0 - \frac{\mu_0}{n_j} + \frac{f_*^{n_j} \mu_0}{n_j} = \mu_{n_j} + \frac{\mu_0}{n_j} + \frac{f_*^{n_j} \mu_0}{n_j} \end{aligned}$$

Since the last two terms tend to 0 as $j \to \infty$ this implies that $f_* \mu_{n_j} \to \mu$ and thus $f_* \mu = \mu$ which implies that $\mu \in \mathcal{M}_f$. The convexity is an easy exercise. To show compactness, suppose that μ_n is a sequence in $\mathcal{M}_f$ converging to some $\mu \in \mathcal{M}$. Then, by Lemma 2.1 we have, for any continuous function φ, that $\int f \circ \varphi d\mu = \lim_{n\to\infty} \int f \circ \varphi d\mu_n = \lim_{n \to \infty} \int f d\mu_n = \int f d\mu$. Therefore, by Corollary 2.3, μ is invariant and so $\mu \in \mathcal{M}_f$. □

2.5. **Ergodic measures are extremal.**

Proposition 2.2. *$\mu \in \mathcal{E}_f$ if and only if μ is an extremal element of $\mathcal{M}_f$.*

Proof. Suppose first that μ is not ergodic, we will show that it cannot be an extremal point. By the definition of ergodicity, if μ is not ergodic, then there exists a set A with

$$f^{-1}(A) = A, \quad f^{-1}(A^c) = A^c \quad \text{and} \quad \mu(A) \in (0,1).$$

Define two measures μ_1, μ_2 by

$$\mu_1(B) = \frac{\mu(B \cap A)}{\mu(A)} \quad \text{and} \quad \mu_2(B) = \frac{\mu(B \cap A^c)}{\mu(A^c)}.$$

μ_1 and μ_2 are probability measures with $\mu_1(A) = 1, \mu_2(A^c) = 1$, and μ can be written as

$$\mu = \mu(A)\mu_1 + \mu(A^c)\mu_2$$

which is a linear combination of μ_1, μ_2:

$$\mu = t\mu_1 + (1-t)\mu_2 \quad \text{with } t = \mu(A) \text{ and } 1 - t = \mu(A^c).$$

It just remains to show that $\mu_1, \mu_2 \in \mathcal{M}_f$, i.e. that they are invariant. Let B be an arbitrary measurable set. Then, using the fact that μ is invariant by assumption and that

$f^{-1}(A) = A$ we have

$$\mu_1(f^{-1}(B)) := \frac{\mu(f^{-1}(B) \cap A)}{\mu(A)} = \frac{\mu(f^{-1}(B) \cap f^{-1}(A))}{\mu(A)} = \frac{\mu(f^{-1}(B \cap A))}{\mu(A)} = \frac{\mu(B \cap A)}{\mu(A)} = \mu_1(B)$$

This shows that μ_1 is invariant. The same calculation works for μ_2 and so this completes the proof in one direction.

Now suppose that μ is ergodic and suppose by contradiction that μ is not extremal so that $\mu = t\mu_1 + (1-t)\mu_2$ for two invariant probability measures μ_1, μ_2 and some $t \in (0,1)$. We will show that $\mu_1 = \mu_2 = \mu$, thus implying that μ is extremal. We will show that $\mu_1 = \mu$, the argument for μ_2 is identical. Notice first of all that $\mu_1 \ll \mu$ and therefore, by the Radon-Nykodim Theorem, it has a *density* $h_1 := d\mu_1/d\mu$ such that for any measurable set we have $\mu_1(A) = \int_A h_1 d\mu$. The statement that $\mu_1 = \mu$ is equivalent to the statement that $h_1 = 1$ μ-almost everywhere. To show this we define the sets

$$B := \{x : h_1(x) < 1\} \quad \text{and} \quad C := \{x : h_1(x) > 1\}$$

and will show that $\mu(B) = 0$ and $\mu(C) = 0$ implying the desired statement. We give the details of the proof of $\mu(B) = 0$, the argument to show that $\mu(C) = 0$ is analogous. Firstly

$$\mu_1(B) = \int_B h_1 d\mu = \int_{B \cap f^{-1}(B)} h_1 d\mu + \int_{B \setminus f^{-1}B} h_1 d\mu$$

and

$$\mu_1(f^{-1}B) = \int_{f^{-1}B} h_1 d\mu = \int_{B \cap f^{-1}(B)} h_1 d\mu + \int_{f^{-1}B \setminus B} h_1 d\mu$$

Since μ_1 is invariant, $\mu_1(B) = \mu_1(f^{-1}B)$ and therefore,

$$\int_{B \setminus f^{-1}B} h_1 d\mu = \int_{f^{-1}B \setminus B} h_1 d\mu.$$

Notice that

$$\mu(f^{-1}B \setminus B) = \mu(f^{-1}(B)) - \mu(f^{-1}B \cap B) = \mu(B) - \mu(f^{-1}B \cap B) = \mu(B \setminus f^{-1}B).$$

Since $h_1 < 1$ on $B \setminus f^{-1}B$ and and $h_1 \geq 1$ on $f^{-1}B \setminus B$ and the value of the two integrals is the same, we must have $\mu(B \setminus f^{-1}B) = \mu(f^{-1}B \setminus B) = 0$, which implies that $f^{-1}B = B$ (up to a set of measure zero). Since μ is ergodic we have $\mu(B) = 0$ or $\mu(B) = 1$. If $\mu(B) = 1$ we would get

$$1 = \mu_1(M) = \int_M h_1 d\mu = \int_B h_1 d\mu < \mu(B) = 1$$

which is a contradiction. It follows that $\mu(B) = 0$ and this concludes the proof. $\square$

The final statement that $\mathcal{E}_f \neq \emptyset$ follows from the general Krein-Millman Theorem which says that $\mathcal{M}_f$ is therefore the closed convex hull of its extreme elements and thus, in particular, the set $\mathcal{E}_f$ of extreme elements is non-empty.

3. Dynamical implications of invariance and ergodicity

3.1. **Poincaré Recurrence.** Historically, the first use of the notion of an invariant measure is due to Poincaré who noticed the remarkable fact that it implies *recurrence*.

Theorem (Poincaré Recurrence Theorem, 1890)**.** *Let μ be an invariant probability measure and A a measurable set with $\mu(A) > 0$. Then for μ-a.e. point $x \in A$ there exists $\tau > 0$ such that $f^\tau(x) \in A$.*

Proof. Let
$$A_0 = \{x \in A : f^n(x) \notin A \text{ for all } n \geq 1\}.$$
Then it is sufficient to show that $\mu(A_0) = 0$. For every $n \geq 0$, let
$$A_n = f^{-n}(A_0)$$
denote the preimages of A_0. We claim that all these preimages are disjoint, i.e.
$$A_n \cap A_m = \emptyset$$
for all $m, n \geq 0$ with $m \neq n$. Indeed, supppose by contradiction that there exists $n > m \geq 0$ and $x \in A_n \cap A_m$. This implies
$$f^n(x) \in f^n(A_n \cap A_m) = f^n(f^{-n}(A_0) \cap f^{-m}(A_0)) = A_0 \cap f^{n-m}(A_0)$$
But this implies $A_0 \cap f^{n-m}(A_0) \neq \emptyset$ which contradicts the definition of A_0 and this proves disjointness of the sets A_n. From the invariance of the measure μ we have $\mu(A_n) = \mu(A)$ for every $n \geq 1$ and therefore
$$1 = \mu(X) \geq \mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n) = \sum_{n=1}^{\infty} \mu(A).$$
Assuming $\mu(A) > 0$ would lead to a contradiction since the sum on the right hand side would be infinite, and therefore we conclude that $\mu(A) = 0$. □

Exercise 4. The finiteness of the measure μ plays a crucial role in this result. Find an example of an infinite measure space $(\hat{X}, \hat{\mathcal{B}}, \hat{\mu})$ and a measure-preserving map $f : \hat{X} \to \hat{X}$ for which the conclusions of Poincare's Recurrence Theorem do not hold.

Remark 1. It does not follow immediately from the theorem that every point of A returns to A infinitely often. To show that almost every point of A returns to A infinitely often let
$$A'' = \{x \in A : \text{ there exists } n \geq 1 \text{ such that } f^k(x) \notin A \text{ for all } k > n\}$$
denote the set of points in A which return to A at most finitely many times. Again, we will show that $\mu(A'') = 0$. First of all let
$$A''_n = \{x \in A : f^n(x) \in A \text{ and } f^k(x) \notin A \text{ for all } k > n\}$$
denote the set of points which return to A for the last time after exactly n iterations. Notice that A''_n are defined very differently than the A'_n. Then
$$A'' = A''_1 \cup A''_2 \cup A''_3 \cup \cdots = \bigcup_{n=1}^{\infty} A''_n.$$

It is therefore sufficient to show that for each $n \geq 1$ we have $\mu(A''_n) = 0$. To see this consider the set $f^n(A''_n)$. By definition this set belongs to A and consists of points which never return to A. Therefore $\mu(f^n(A''_n)) = 0$. Moreover we have we clearly have

$$A''_n \subseteq f^{-n}(f^n(A''_n))$$

and therefore, using the invariance of the measure we have

$$\mu(A''_n) \leq \mu(f^{-n}(f^n(A''_n))) = \mu(f^n(A''_n)) = 0.$$

3.2. **Birkhoff's Ergodic Theorem.** A couple of decades after Poincaré's theorem, Birkhoff proved the following even more remarkable result which gives some qualitative results about the recurrence in particular in the case in which the invariant measure is also ergodic.

Theorem 2 (Birkhoff, 1920's)**.** *Let M be a measure space, $f : M \to M$ a measurable map, and μ an f-invariant probability measure. Then, for every $\varphi \in L^1(\mu)$ the limit*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ f^i(x)$$

exists for μ almost every x. Moreover, if μ is ergodic, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ f^i(x) = \int \varphi d\mu$$

for μ almost every x.

We can formulate this result informally by saying that when μ is ergodic the *the time averages converge to the space average* as in the following

Corollary 3.1. *Let μ be an f-invariant ergodic probability measure. For any Borel measurable set A and μ-a.e. x we have*

$$\lim_{n\to\infty} \frac{1}{n} \#\{1 \leq j \leq n : f^j(x) \in A\} = \mu(A).$$

Proof. Let $\varphi = \mathbb{1}_A$ be the characteristic function of A. Then

$$\lim_{n\to\infty} \frac{1}{n} \#\{1 \leq j \leq n : f^j(x) \in A\} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_A(f^i(x)) = \int \mathbb{1}_A d\mu = \mu(A).$$

□

We recall that the *basin of attraction* of an arbitrary measure $\mu \in \mathcal{M}$ is the set

$$\mathfrak{B}_\mu := \left\{ x : \frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ f^i(x) \to \int \varphi d\mu \quad \text{for all } \varphi : C^0(M, \mathbb{R}) \right\}.$$

A priori there is no reason for this set to be non-empty.

Corollary 3.2. *Let M be a compact Hausdorff space and μ an f-invariant ergodic probability measure. Then $\mu(\mathfrak{B}_\mu) = 1$.*

3.3. **Existence of the limit for Birkhoff averages.**

Proposition 3.1. *Let μ be an f-invariant measure and $\varphi \in L^1(\mu)$. Then*
$$\varphi_f(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^i(x)$$
exists for μ-almost every x.

We first prove a technical lemma. Let
$$\mathcal{I} := \{A \in \mathcal{B} : f^{-1}(A) = A\}$$
be the collection of fully invariant sets of $\mathcal{B}$ and notice that $\mathcal{I}$ is a sub-σ-algebra. For $\psi \in L^1(\mu)$ let $\psi\mu \ll \mu$ denote the measure which has density ψ with respect to μ, i.e. $\psi = d\psi\mu/d\mu$ is the Radon-Nykodym derivative of $\psi\mu$ with respect to μ. Let $\psi\mu|_{\mathcal{I}}$ and $\mu|_{\mathcal{I}}$ denote the restrictions of these measures to $\mathcal{I}$. Then clearly $\psi\mu|_{\mathcal{I}} \ll \mu|_{\mathcal{I}}$ and therefore the Radon-Nykodim derivative
$$\psi_{\mathcal{I}} := \frac{d\psi\mu|_{\mathcal{I}}}{d\mu|_{\mathcal{I}}}$$
exists. This is also called the conditional expectation of ψ with respect to $\mathcal{I}$

Lemma 3.1. *Suppose $\psi_{\mathcal{I}} < 0$ (resp $\psi_{\mathcal{I}} > 0$). Then*
$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \psi \circ f^k(x) \leq 0 \qquad \left(\textit{resp.} \qquad \liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \psi \circ f^k(x) \geq 0 \right)$$
for μ almost every x.

Proof. Let
$$\Psi_n := \max_{k\leq n} \left\{ \sum_{i=0}^{k-1} \psi \circ f^i \right\} \qquad \text{and} \qquad A := \{x : \Psi_n \to \infty\}$$
Then, for $x \notin A$, Ψ_n is bounded above and therefore
$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \psi \circ f^k \leq \limsup_{n\to\infty} \frac{\Psi_n}{n} \leq 0$$
So it is sufficient to show that $\mu(A) = 0$. To see this, first compare the quantities
$$\Psi_{n+1} = \max_{1\leq k\leq n+1} \left\{ \sum_{i=0}^{k-1} \psi \circ f^i \right\} \qquad \text{and} \qquad \Psi_n \circ f = \max_{1\leq k\leq n} \left\{ \sum_{i=0}^{k-1} \psi \circ f^{i+1} \right\} = \max_{1\leq k\leq n} \left\{ \sum_{i=1}^{k} \psi \circ f^i \right\}$$
The two sums are almost exactly the same except for the fact that Ψ_{n+1} includes the quantity $\psi(x)$ and therefore we have
$$\Psi_{n+1} = \begin{cases} \psi + \Psi_n \circ f & \text{if } \Psi_n \circ f > 0 \\ \psi & \text{if } \Psi_n \circ f < 0 \end{cases}$$

We can write this as

$$\Psi_{n+1} - \Psi_n \circ f = \psi - \min\{0, \Psi_n \circ f\}$$

Then of course, A is forward and backward invariant, and this in particular $A \in \mathcal{I}$ and also $\Psi_n \circ f \to \infty$ on A and therefore $\Psi_{n+1} - \Psi_n \circ f \downarrow \psi$ for all $x \in A$. Therefore, using the invariance of μ, by the Dominated Convergence Theorem, we have

$$\int_A \Psi_{n+1} - \Psi_n d\mu = \int_A \Psi_{n+1} - \Psi_n \circ f d\mu \to \int_A \psi d\mu = \int_A \psi_{\mathcal{I}} d\mu_I$$

By definition we have $\Psi_{n+1} \geq \Psi_n$ and therefore the integral on the right hand side is ≥ 0. Thus if $\psi_{\mathcal{I}} < 0$ this implies that $\mu(A) = \mu_{\mathcal{I}}(A) = 0$. Replacing ψ by $-\psi$ and repeating the argument completes the proof.

□

Proof of Proposition 3.1 . Let $\psi := \varphi - \varphi_{\mathcal{I}} - \epsilon$. Since $(\varphi_{\mathcal{I}})_{\mathcal{I}} = \varphi_{\mathcal{I}}$ we have $\psi_{\mathcal{I}} = -\epsilon < 0$. Thus, by Lemma 3.1 we have

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \psi \circ f^k(x) = \limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ f^k(x) - \varphi_{\mathcal{I}} - \epsilon \leq 0$$

and therefore

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ f^k(x) \leq \varphi_{\mathcal{I}} + \epsilon$$

for μ almost every x. Now, letting $\psi := -\varphi + \varphi_{\mathcal{I}} + \epsilon$ we have $\psi_{\mathcal{I}} = \epsilon > 0$ and therefore again by Lemma 3.1 we have

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \psi \circ f^k = -\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ f^k + \varphi_{\mathcal{I}} + \epsilon \geq 0$$

which implies

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ f^k \leq \varphi_{\mathcal{I}} + \epsilon$$

for μ almost every x. Since $\epsilon > 0$ is arbitrary we get that the limit exists and

$$\varphi_f := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ f^k = \varphi_{\mathcal{I}}$$

for μ almost every x. □

3.4. **Existence of the limit for ergodic measures.**

Lemma 3.2. *The following two conditions are equivalent:*

(1) *μ is ergodic;*
(2) *if $\varphi \in L^1(\mu)$ satisfies $\varphi \circ f = \varphi$ for μ almost every x. then φ is constant a.e.*

Proof. Suppose first that μ is ergodic and let $\varphi \in L^1$ satisfy $\varphi \circ f = \varphi$. Let

$$X_{k,n} := \varphi^{-1}([k2^{-n}, (k+1)2^{-n})).$$

Since φ is measurable, the sets $X_{k,n}$ are measurable. Moreover, since φ is constant along orbits, the sets $X_{k,n}$ are backward invariant a.e. and thus by ergodicity they have either zero or full measure. Moreover, they are disjoint in n and their union is the whole of $\mathbb{R}$ and so for each n there exists a unique k_n such that $\mu(X_{k_n,n}) = 1$. Thus, letting $Y = \cup_{n\in\mathbb{Z}} X_{k_n,n}$ we have that $\mu(Y) = 1$ and φ is constant on Y. Thus φ is constant *a.e.*.

Conversely, suppose that (2) holds and suppose that $f^{-1}(A) = A$. Let $\mathbb{1}_A$ denote the characteristic function of A. Then clearly $\mathbb{1}_A \in L^1$ and $\mathbb{1}_A \circ f = \mathbb{1}_A$ and so we either have $\mathbb{1}_A = 0$ a.e. or $\mathbb{1}_A = 1$ a.e. which proves that $\mu(A) = 0$ or 1. □

Corollary 3.3. *Let μ be an f-invariant ergodic probability measure and $\varphi \in L^1(\mu)$. Then*

$$\varphi_f(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi \circ f^i(x) = \int \varphi d\mu$$

exists for μ-almost every x.

Proof. Since $\varphi_f = \varphi_{\mathcal{I}}$ a.e., it follows that $\int \varphi_f d\mu = \int \varphi_{\mathcal{I}} d\mu_I = \int \varphi d\mu$ and in particular $\varphi_f \in L^1$. Since φ_f is also invariant along orbits, it follows that it is constant a.e. and therefore $\varphi_f = \int \varphi d\mu$. □

3.5. **Typical points for a measure.**

Proof of Corollary 3.2. By Proposition 3.1 and Corollary 3.3, for every $\varphi \in L^1$, the exceptional set where $\varphi_f \neq \int \varphi d\mu$ has measure 0. This set depends in general on the function φ, but since the union of a countable collection of sets of zero measure has zero measure, given any countable collection $\{\varphi_m\}_{m=1}^{\infty}$ we can find a common set of full measure for which the time averages converge. In particular, since M is compact and Hausdorff, we can choose as such a countable dense subset of continuous functions. Then, for any arbitrary continuous function φ and arbitrary $\epsilon > 0$, choose φ_m such that $\sup_{x\in M} |\varphi(x) - \varphi_m(x)| < \epsilon$. Then we have

$$\frac{1}{n}\sum_{i=0}^{n-1} \varphi \circ f^i(x) = \frac{1}{n}\sum_{i=0}^{n-1} \varphi_m \circ f^i(x) + \frac{1}{n}\sum_{i=0}^{n-1} (\varphi \circ f^i(x) - \varphi_m \circ f^i(x))$$

The first sum converges as $n \to \infty$ and the second sum is bounded by ϵ and therefore all the limit points of the sequence on the left are within ϵ of each other. Since ϵ is arbitrary, this implies that they converge. □

4. Unique Ergodicity and Circle Rotations

We now begin the study of specific classes of dynamical systems. We start with a class of maps, circle rotations, for which Lebesgues measure is invariant and therefore the issue if ergodicity. It is easy to see that Lebesgue measure is not ergodic is the rotation is rational. On the other hand, if the rotation is irrational we have not only that Lebesgue measure is invariant and ergodic but that there are no other invariant probability measures at all. Thus $\mathcal{M}_f = \mathcal{E}_f = \{Lebesgue\}$. Let $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$.

Theorem 3. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be the circle rotation $f(x) = x + \alpha$ with α irrational. Then Lebesgue measure is the unique invariant ergodic probability measure.*

It follows immediately from Birkhoff's ergodic theorem that the orbit $\mathcal{O}^+(x) = \{x_n\}_{n=0}^{\infty}$ of Lebesgue almost every point is *uniformly distributed* in $\mathbb{S}^1$ (with respect to Lebesgue) in the sense that for any arc $(a,b) \subset \mathbb{S}^1$ we have

$$\frac{\#\{0 \le i \le n-1 : x_i \in (a,b)\}}{n} \to m(a,b).$$

As a consequence of the uniqueness of the invariant measure, in the case of irrational circle rotations we get the stronger statement that this property holds for every $x \in \mathbb{S}^1$.

Theorem 4. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be the circle rotation $f(x) = x + \alpha$ with α irrational. Then every orbit is uniformly distributed in $\mathbb{S}^1$.*

We shall prove these results in a more general framework.

Definition 4. We say that a map $f : X \to X$ is *uniquely ergodic* if it admits a unique (ergodic) invariant probability measure.

Trivial examples of uniquely ergodic maps are contraction maps in which every orbit converges to a unique fixed point p, and therefore δ_p is the unique ergodic invariant probability measure.

Theorem 5. *Let $f : X \to X$ be a continuous map of a compact metric space. Suppose there exists a dense set Φ of continuous functions $\varphi : X \to \mathbb{R}$ such that for every $\varphi \in \Phi$ there exists a constant $\bar{\varphi} = \bar{\varphi}(\varphi)$ such that*

$$\frac{1}{n}\sum_{j=0}^{n-1} \varphi \circ f^j \to \bar{\varphi} \quad \textit{uniformly.} \tag{3}$$

Then f is uniquely ergodic. Conversely, suppose that f is uniquely ergodic, then for every continuous function $\varphi : X \to \mathbb{R}$ there exists a constant $\bar{\varphi} = \bar{\varphi}(\varphi)$ such that (3) *holds. Moreover, if μ is the unique invariant probability measure for f, then $\bar{\varphi}(\varphi) = \int \varphi d\mu$.*

Remark 2. The statement also holds for complex valued observables $\varphi : X \to \mathbb{C}$ and we will use the complex valued version in the proof of Theorem 3. We will prove it below for real valued observables but the proox is exactly the same in the complex case.

Lemma 4.1. *Let $f : X \to X$ be a continuous map of a compact metric space. Suppose there exists a dense set Φ of continuous functions $\varphi : X \to \mathbb{R}$ such that for every $\varphi \in \Phi$ there exists a constant $\bar{\varphi} = \bar{\varphi}(\varphi)$ such that* (3) *holds. Then* (3) *holds for every continuous function φ.*

Proof. To simplify the notation we let

$$\mathfrak{B}_n(x,\varphi) := \frac{1}{n}\sum_{i=0}^{n-1} \varphi\circ f^i(x).$$

By assumption, if $\varphi\in\Phi$, there exists a constant $\bar\varphi = \bar\varphi(\varphi)$ such that $\mathfrak{B}_n(x,\varphi)\to\bar\varphi$ uniformly in x. Now let $\psi: X\to\mathbb{R}$ be an arbitrary continuous function. Since Φ is dense, for any $\epsilon>0$ there exists $\phi\in\Phi$ such that $\sup_{x\in X}|\varphi(x)-\psi(x)|<\epsilon$. This implies

$$|\mathfrak{B}_n(x,\varphi)-\mathfrak{B}_n(x,\psi)|<\epsilon$$

for every x,n and therefore

$$\left|\sup_{x,n}\mathfrak{B}_n(x,\psi)-\bar\varphi\right|<\epsilon \quad\text{ and }\quad \left|\inf_{x,n}\mathfrak{B}_n(x,\psi)-\bar\varphi\right|<\epsilon$$

and so in particular

$$\left|\sup_{x,n}\mathfrak{B}_n(x,\psi)-\inf_{x,n}\mathfrak{B}_n(x,\psi)\right|<2\epsilon.$$

Since ϵ is arbitrary, this implies that $\mathfrak{B}_n(x,\psi)$ converges uniformly to some constant $\bar\psi$. Notice that the function φ and therefore the constant $\bar\varphi$ depends on ϵ, so what we have shown here is simply that the inf and the sup are within 2ϵ of each other for arbitrary ϵ and therefore must coincide. This shows that (3) holds for every continuous function. □

Lemma 4.2. *Supppose that for any continuous function $\varphi: X\to\mathbb{R}$ there exists a constant $\bar\varphi$ such that* (3) *holds. Then f is uniquely ergodic.*

Proof. Given a continuous function φ, by Birkhoff's Ergodic Theorem, for every ergodic invariant probability measure μ we have

$$\frac{1}{n}\sum_{j=0}^{n-1}\varphi\circ f^j \to \int\varphi d\mu.$$

Therefore, from (3) we have

$$\int\varphi d\mu = \bar\varphi(\varphi)$$

for *every* ergodic invariant measure μ. This clealry implies unique ergodicity since if μ_1,μ_2 are ergodic invariant probability measures this implies $\int\varphi d\mu_1 = \int\varphi d\mu_2$ for every continuous function φ and this implies $\mu_1=\mu_2$. □

Proof of Theorem 5. Suppose that f is uniquely ergodic and μ is the unique ergodic invariant probability measure. Then by Birkhoff's ergodic Theorem we have

$$\frac{1}{n}\sum_{j=0}^{n-1}\varphi\circ f^j(x) \to \int\varphi d\mu$$

for μ-a.e. x. We let $\bar{\varphi} = \int \varphi d\mu$ and show that (3) holds. Suppose by contradiction that (3) does not hold. Then by the negation of the definition of uniform continuity, there exists a continuous function φ and $\epsilon > 0$ and sequences $x_k \in X$ and $n_k \to \infty$ for which

$$\left| \frac{1}{n} \sum_{i=0}^{n_k - 1} \varphi(f^i(x_k)) - \bar{\varphi} \right| \geq \epsilon$$

Define a sequence of measures

$$\nu_k := \frac{1}{n} \sum_{i=0}^{n_k - 1} f_*^i \delta_{x_k} = \frac{1}{n} \sum_{i=0}^{n_k - 1} \delta_{f^i x_k}.$$

Notice that for any x we have $f_*^i \delta_x = \delta_{f^i(x)}$. Then, for every k we have

$$\int \varphi d\nu_k = \int \varphi d \frac{1}{n_k} \sum_{i=0}^{n_k - 1} \delta_{f^i x_k} = \frac{1}{n_k} \sum_{i=0}^{n_k - 1} \int \varphi d\delta_{f^i x_k} = \frac{1}{n_k} \sum_{i=0}^{n_k - 1} \varphi(f^i(x_k))$$

and therefore

$$\left| \int \varphi d\nu_k - \bar{\varphi} \right| \geq \epsilon.$$

for every k. By the weak-star compactness of the space $\mathcal{M}$ of probability measures, there exists a subsequence $k_j \to \infty$ and a probability measure $\nu \in \mathcal{M}$ such that $\nu_{k_j} \to \nu$ and

$$\left| \int \varphi d\nu - \bar{\varphi} \right| \geq \epsilon.$$

Moreover, arguing as in the proof of the Krylov-Boguliobov Theorem (Proposition 2.1) we get[2] that $\nu \in \mathcal{M}_f$. Thus, by Birkhoff's Ergodic Theorem, for ν-a.e. x the ergodic averages converge to $\int \varphi d\mu \neq \bar{\varphi}$. This implies that $\nu \neq \mu$ contradicting the assumptions of unique ergodicity. □

Proof of Theorem 3. It is sufficient to show that there exists a dense set of continuous functions $\varphi : \mathbb{S}^1 \to \mathbb{C}$ for which the Birkhoff time averages

$$\mathfrak{B}_n(\varphi, x) := \frac{1}{n} \sum_{i=0}^{n-1} \varphi \circ f^i(x)$$

converge uniformly to a constant. For any $m \geq 1$, consider the functions

$$\varphi_m(x) := e^{2\pi i m x} = \cos 2\pi m x + i 2\pi m x$$

[2]Indeed,

$$f_* \nu_{k_j} = f_* \left(\frac{1}{n_{k_j}} \sum_{i=0}^{n_{k_j} - 1} f_*^i \delta_{x_{k_j}} \right) = \frac{1}{n_{k_j}} \sum_{i=0}^{n_{k_j} - 1} f_*^{i+1} \delta_{x_{k_j}} = \frac{1}{n_{k_j}} \sum_{i=0}^{n_{k_j} - 1} f_*^i \delta_{x_{k_j}} + \frac{1}{n_{k_j}} \left(f_*^{n_{k_j}} \delta_{x_{k_j}} - \delta_{x_{k_j}} \right)$$

and therefore $f_* \nu_{k_j} \to \nu$ as $j \to \infty$. Since $\nu_{k_j} \to \nu$ by definition of ν and $f_* \nu_{k_j} \to f_* \nu$ by continuity of f_*, this implies $f_* \nu = \nu$ and thus $\nu \in \mathcal{M}_f$.

and let Φ denote the space of all linear combinations of functions of the form φ_m. By a classical Theorem of Weierstrass, Φ is dense in the soace of all continuous functions, thus it is sufficient to show uniform convergence for functions in Φ. Moreover, notice that for any two continuous functions φ, ψ we have

$$\mathfrak{B}_n(\varphi+\psi, x) := \frac{1}{n}\sum_{i=0}^{n-1}(\varphi+\psi)\circ f^i(x) = \frac{1}{n}\sum_{i=0}^{n-1}(\varphi\circ f^i(x)+\psi\circ f^i(x)) = \mathfrak{B}_n(\varphi, x)+\mathfrak{B}_n(\psi, x).$$

Thus, the *Birkhoff averaging operator* is linear in the observable and therefore to show the statement for all functions in Φ it is sufficient to show it for each φ_m. To see this, notice first of all that

$$\varphi_m \circ f(x) = e^{2\pi im(x+\alpha)} = e^{2\pi im\alpha}e^{2\pi imx} = e^{2\pi im\alpha}\varphi_m(x)$$

and therefore, using $|\varphi_m(x)| = 1$ and the sum $\sum_{j=0}^n x^j = (1-x^{n+1})/(1-x)$ we get

$$\left|\frac{1}{n}\sum_{j=0}^{n-1}\varphi_m\circ f^j(x)\right| = \left|\frac{1}{n}\sum_{j=0}^{n-1}e^{2\pi imj\alpha}\right| = \frac{1}{n}\frac{|1-e^{2\pi imn\alpha}|}{|1-e^{2\pi im\alpha}|} \leq \frac{1}{n}\frac{1}{|1-e^{2\pi im\alpha}|} \to 0$$

The convergence is uniform because the upper bound does not depend on x. Notice that we have used here the fact that α is irrational in an essential way to guarantee that the denominator does not vanish for any m. Notice also that the convergence is of course not uniform (and does not need to be uniform) in m. □

Proof of Theorem 4. Consider an arbitrary arc $[a,b] \subset \mathbb{S}^1$. Then, for any $\epsilon > 0$ there exist continuous functions $\varphi, \psi : \mathbb{S}^1 \to \mathbb{R}$ such that $\varphi \leq \mathbb{1}_{[a,b]} \leq \psi$ and such that $\int \psi - \varphi dm \leq \epsilon$. We then have that

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\mathbb{1}_{[a,b]}(x_j) \geq \liminf_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\varphi(x_j) = \int \varphi dm \geq \int \psi dm - \epsilon \geq \int \mathbb{1}_{[a,b]}(x_j) - \epsilon$$

and

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\mathbb{1}_{[a,b]}(x_j) \leq \limsup_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\psi(x_j) = \int \psi dm \leq \int \varphi dm + \epsilon \leq \int \mathbb{1}_{[a,b]}(x_j) + \epsilon$$

Since ϵ is arbitrary, the limit exists and equals $\int \mathbb{1}_{[a,b]}dm = |b-a|$ and thus the sequence is uniformly distributed. □

4.1. **Benford's distribution.** We give an interesting application of the uniform distribution result above. First of all we define the concept of a leading digit of a number $a \in \mathbb{R}$. We define the *leading digit* of a as the first non-zero digit in the decimal expansion of a. Thus, if $|a| \geq 1$ this is just the first digit of a. If $|a| < 1$ this is the first non-zero digit after the decimal point. We shall use the notation

$$\mathscr{D}(a) = \text{leading digit of } a.$$

Definition 5. We say that the sequence $\{a_i\}_{i=0}^{\infty}$ has a *Benford* distribution if for every $d = 1, \dots, 9$ we have

$$\mathcal{B}(d) := \lim_{n\to\infty} \frac{\#\{0 \le i \le n-1 : \mathscr{D}(a_i) = d\}}{n} = \log_{10}\left(1 + \frac{1}{d}\right).$$

This give the following approximate values:

$\mathcal{B}(1) = 0.301... \approx 30\%$
$\mathcal{B}(2) = 0.176... \approx 17\%$
$\mathcal{B}(3) = 0.124... \approx 12\%$
$\mathcal{B}(4) = 0.096... \approx 9\%$
$\mathcal{B}(5) = 0.079... \approx 8\%$
$\mathcal{B}(6) = 0.066... \approx 7\%$
$\mathcal{B}(7) = 0.057... \approx 6\%$
$\mathcal{B}(8) = 0.051... \approx 5\%$
$\mathcal{B}(9) = 0.045... \approx 4\%$

Notice that

$$\sum_{d=1}^{p} \log_{10}\left(1 + \frac{1}{d}\right) = 1$$

so that $\mathcal{B}(d)$ are the *probabilities* of each digit d occuring as a laeding digit.

Remark 3. Remarkably, this distribution is observed in a variety of real-life data, mostly in case in which there is a large amount of data across several orders of magnitude. It was first observed by American astronomer Simon Newcombe in 1881 when he noticed that the earlier pages of logarithm tables, containing numbers starting with 1, were much more worn that other pages. This was rediscovered by physicist Frank Benford who discovered that a wide amount of data followed this principle.

Proposition 4.1. *Let k be any integer number that is not a power of ten. Then the sequence $\{k^n\}_{n=1}^{\infty}$ satisfies Benford's distribution.*

We prove the Proposition in the following two lemmas.

Lemma 4.3. *Let k be any integer number that is not a power of ten. Then the sequence $\{\log_{10} k^n \bmod 1\}_{n=1}^{\infty}$ is uniformly distributed in $\mathbb{S}^1$.*

Proof of Proposition 4.1. Notice that $\log_{10} k^n = n \log_{10} k$ and therefore it is sufficient to show that the sequence $\{n \log_{10} k \mod 1\}_{i=1}^{\infty}$ is uniformly distributed in $\mathbb{S}^1$. Since k is not a power of 10 the number $\log_{10} k$ is irrational and this sequence can be seen as the sequence of iterates of $x_0 = 0$ under the irrational circle rotation $f(x) = x + \log_{10} k$ and therefore is uniformly distributed. □

Lemma 4.4. *Let $\{a_i\}_{i=1}^{\infty}$ be a sequence of real numbers and suppose that the sequence $\{\log_{10} a_i \mod 1\}_{i=1}^{\infty}$ is uniformly distributed in $\mathbb{S}^1$. Then $\{a_i\}_{i=1}^{\infty}$ satisfies Benford's distribution.*

Proof. Notice first of all that for each a_i we have

$$\mathscr{D}(a_i) = d \quad \Longleftrightarrow \quad d10^j \le a_i < (d+1)10^j \quad \text{for some } j \in \mathbb{Z}$$

Therefore

$$\mathscr{D}(a_i) = d \quad \Longleftrightarrow \quad \log_{10} d + j \le \log_{10} a_1 \le \log_{10}(d+1) + j$$

or

$$\mathscr{D}(a_i) = d \quad \Longleftrightarrow \quad \log_{10} d \le \log_{10} a_i \bmod 1 \le \log_{10}(d+1).$$

By assumption, $\{\log_{10} a_i\}$ is uniformly distributed and therefore

$$\lim_{n\to\infty} \frac{\#\{1 \le i \le n : \mathscr{D}(a_i) = d\}}{n} = \lim_{n\to\infty} \frac{\#\{1 \le i \le n : \log_{10} a_i \bmod 1 \in (\log_{10} d, \log_{10}(d+1)\}}{n}$$
$$= \log \frac{d+1}{d} = \log_{10}\left(1 + \frac{1}{d}\right).$$

□

5. Piecewise affine full branch maps

Definition 6. Let $I \subset \mathbb{R}$ be an interval. A map $f : I \to I$ is a *full branch* map if there exists a finite or countable partition $\mathcal{P}$ of I (mod 0) into subintervals such that for each $\omega \in \mathcal{P}$ the map $f|_{int(\omega)} : int(\omega) \to int(I)$ is a bijection. f is a *piecewise continuous (resp. C^1, C^2, affine) full branch* map if for each $\omega \in \mathcal{P}$ the map $f|_{int(\omega)} : int(\omega) \to int(I)$ is a *homeomorphism (resp. C^1 diffeomorphism, C^2 diffeomorphism, affine)*.

The full branch property is extremely important and useful. It is a fairly strong property but it turns out that the study of many maps which do not have this property can be reduced to maps with the full branch property. In this section we start by studying the case of piecewise affine full branch maps. We will prove the following.

Proposition 5.1. *Let $f : I \to I$ be a piecewise affine full branch map. Then Lebesgue measure is invariant and ergodic.*

Example 6. The simplest examples of full branch maps are the maps $f : [0,1] \to [0,1]$ defined by $f(x) = \kappa x \bmod 1$ for some integer $\kappa \ge 1$. In this case it is almost trivial to check that Lebesgue measure is invariant. In the general case in which the branches have different derivatives and if there are an infinite number of branches it is a simple exercise.

Exercise 5. Let $f : I \to I$ be a piecewise affine full branch map. Then Lebesgue measure is invariant. We write f'_ω to denote the derivative of f on $int(\omega)$. In the general case (even with an infinite number of branches) we have $|\omega| = 1/|f'_\omega|$. Thus, for any interval $A \subset I$ we have

$$|f^{-1}(A)| = \sum_{\omega\in\mathcal{P}} |f^{-1}(A) \cap \omega| = \sum_{\omega\in\mathcal{P}} \frac{|A|}{|f'_\omega|} = |A| \sum_{\omega\in\mathcal{P}} \frac{1}{|f'_\omega|} = |A| \sum_{\omega\in\mathcal{P}} |\omega| = |A|.$$

Thus Lebesgue measure is invariant.

Lemma 5.1. *Let $f : I \to I$ be a continuous (resp. C^1, C^2, affine) full branch map. Then there exists a family of partitions $\{\mathcal{P}^{(n)}\}_{n=1}^{\infty}$ of I (mod 0) into subintervals such that $\mathcal{P}^{(1)} = \mathcal{P}$, each $\mathcal{P}^{(n+1)}$ is a refinement of $\mathcal{P}^{(n)}$, and such that for each $n \geq 1$ and each $\omega^{(n)} \in \mathcal{P}^{(n)}$ the map $f^n : |_{int(\omega^{(n)})} : int(\omega^{(n)}) \to int(I)$ is a homeomorphism (resp. a C^1 diffeomorphism, C^2 diffeomorphism, affine map).*

Proof. For $n = 1$ we let $\mathcal{P}^{(1)} := \mathcal{P}$ where $\mathcal{P}$ is the partition in the definition of a full branch map. Proceeding inductively, suppose that there exists a partition $\mathcal{P}^{(n-1)}$ satisfying the required conditions. Then each $\omega^{(n-1)}$ is mapped by f^{n-1}bijectively to the entire interval I and therefore $\omega^{(n-1)}$ can be subdivided into disjoint subintervals each of which maps bijectively to one of the elements of the original partition $\mathcal{P}$. Thus each of these subintervals will then be mapped under one further iteration bijectively to the entire interval I. These are therefore the elements of the partition $\mathcal{P}^{(n)}$. □

Proof of Proposition 5.1. Let $A \subset [0, 1)$ satisfying $f^{-1}(A) = A$ and suppose that $|A| > 0$. We shall show that $|A| = 1$. Notice first of all that since f is piecewise affine, each element $\omega \in \mathcal{P}$ is mapped affinely and bijectively to I and therefore must have derivative strictly larger than 1 uniformly in ω. Thus the iterates f^n have derivatives which are growing exponentially in n and thus, by the Mean Value Theorem, $|\omega^{(n)}| \to 0$ exponentially (and uniformly). By Lebesgue's density Theorem, for any $\epsilon > 0$ we can find $n = n_\epsilon$ sufficiently large so that the elements of $\mathcal{P}_n$ are sufficiently small so that there exists some $\omega^{(n)} \in \mathcal{P}^{(n)}$ with $|\omega^{(n)} \cap A| \geq (1 - \epsilon)|\omega^{(n)}|$ or, equivalently, $|\omega^{(n)} \cap A^c| \leq \epsilon|\omega^{(n)}|$ or

$$\frac{|\omega^{(n)} \cap A^c|}{|\omega^{(n)}|} \leq \epsilon$$

Since $f^n : \omega^{(n)} \to I$ is an affine bijection we have

$$\frac{|\omega^{(n)} \cap A^c|}{|\omega^{(n)}|} = \frac{|f^n(\omega_n \cap A^c)|}{|f^n(\omega_n)|}.$$

Moreover, $f^n(\omega_n) = I$ and and since $f^{-1}(A) = A$ implies $f^{-1}(A^c) = A^c$ which implies $f^{-n}(A^c) = A^c$ we have

$$f^n(\omega^{(n)} \cap A^c) = f^n(\omega_n \cap f^{-n}(A^c)) = A^c.$$

We conclude that

$$\frac{|A^c|}{|I|} = \frac{|f^n(\omega_n \cap A^c)|}{|f^n(\omega_n)|} = \frac{|\omega^{(n)} \cap A^c|}{|\omega^{(n)}|} \leq \epsilon. \tag{4}$$

This gives $|A^c| \leq \epsilon$ and since ϵ is arbitrary this implies $|A^c| = 0$ which implies $|A| = 1$ as required. □

Remark 4. Notice that the "affine" property of f has been used only in two places: two show that the map is expanding in the sense of Lemma **??**, and in the last equality of (4). Thus in the first place it would have been quite sufficient to replace the affine assumption with a uniform expansivity assumption. In the first place it would be sufficient to have an inequality rather than an equality. We will show below that we can indeed obtain similar results for full branch maps by relaxing the affine assumption.

5.1. **Application: Normal numbers.** The relatively simple result on the invariance and ergodicity of Lebesgue measure for piecewise affine full branch maps has a remarkable application on the theory of numbers. For any number $x \in [0,1]$ and any integer $k \geq 2$ we can write

$$x = \frac{x_1}{k^1} + \frac{x_2}{k^2} + \frac{x_3}{k^3} \dots$$

where each $x_i \in \{0, \dots, k-1\}$. This is sometimes called the expansion of x in base k and is (apart from some exceptional cases) unique. Sometimes we just write

$$x = 0.x_1x_2x_3\dots$$

when it is understood that the expansion is with respect to a particular base k. For the case $k = 10$ this is of course just the well known *decimal expansion* of x.

Definition 7. A number $x \in [0,1]$ is called *normal (in base k)* if its expansion $x = 0.x_1x_2x_3\dots$ in base k contains asymptotically equal proportions of all digits, i.e. if for every $j = 0, \dots, k-1$ we have that

$$\frac{\sharp\{1 \leq i \leq n : x_i = j\}}{n} \to \frac{1}{k}$$

as $n \to \infty$.

Exercise 6. Give examples of normal and non normal numbers in a given base k.

It is not however immediately obvious what proportion of numbers are normal in any given base nor if there even might exist a number that is normal in every base. We will show that in fact Lebesgue almost every x is normal in every base.

Theorem 6. *There exists set $\mathcal{N} \subset [0,1]$ with $|\mathcal{N}| = 1$ such that every $x \in \mathcal{N}$ is normal in every base $k \geq 2$.*

Proof. It is enough to show that for any given $k \geq 2$ there exists a set $\mathcal{N}_k$ with $m(\mathcal{N}_k) = 1$ such that every $x \in \mathcal{N}_k$ is normal in base k. Indeed, this implies that for each $k \geq 2$ the set of points $I \setminus \mathcal{N}_k$ which is not normal in base k satisfies $m(I \setminus \mathcal{N}_k) = 0$. Thus the set of point $I \setminus \mathcal{N}$ which is not normal in every base is contained in the union of all $I \setminus \mathcal{N}_k$ and since the countable union of sets of measure zero has measure zero we have

$$m(I \setminus \mathcal{N}) \leq m\left(\bigcup_{k=2}^{\infty} I \setminus \mathcal{N}_k\right) \leq \sum_{k=2}^{\infty} m(I \setminus \mathcal{N}_k) = 0.$$

We therefore fix some $k \geq 2$ and consider the set $\mathcal{N}_k$ of points which are normal in base k. The crucial observation is that the base k expansion of the number x is closely related to its orbit under the map f_k. Indeed, consider the intervals $A_j = [j/k, (j+1)/k)$ for $j = 0, \dots, k-1$. Then, the base k expansion $x = 0.x_1x_2x_3\dots$ of the point x clearly satisfies

$$x \in A_j \iff x_1 = j.$$

Moreover, for any $i \geq 0$ we have

$$f^i(x) \in A_j \Longleftrightarrow x_{i+1} = j.$$

Therefore the frequency of occurrences of the digit j in the expansion of x is exactly the same as the frequence of visits of the orbit of the point x to A_j under iterations of the map f_k. Birkhoff's ergodic theorem and the ergodicity of Lebesgue measure for f_k implies that Lebesgue almost every orbit spends asymptotically $m(A_j) = 1/k$ of its iterations in each of the intervals A_j. Therefore Lebesgue almost every point has an asymptotic frequence $1/k$ of each digit j in its decimal expansion. Therefore Lebesgue almost every point is normal in base k. □

6. Full branch maps with bounded distortion

We now want to relax the assumption that f is piecewise affine.

Definition 8. A full branch map has *bounded distortion* if

$$\sup_{n\geq 1} \sup_{\omega^{(n)}\in\mathcal{P}^{(n)}} \sup_{x,y\in\omega^{(n)}} \log |Df^n(x)/Df^n(y)| < \infty. \tag{5}$$

Notice that the distortion is 0 if f is piecewise affine so that the bounded distortion property is automatically satisfied in that case.

Theorem 7. *Let $f : I \to I$ be a full branch map with bounded distortion. Then Lebesgue measure is ergodic.*

6.1. Bounded distortion implies ergodicity.

We now prove Theorem 7. For any subinterval J and any $n \geq 1$ we define the distortion of f^n on J as

$$\mathcal{D}(f,^n J) := \sup_{x,y\in J} \log |Df^n(x)/Df^n(y)|.$$

The bounded distortion condition says that $\mathcal{D}(f^n, \omega^{(n)})$ is uniformly bounded. The distortion has an immediate geometrical interpretation in terms of the way that ratios of lengths of intervals are (or not) preserved under f.

Lemma 6.1. *Let $\mathcal{D} = \mathcal{D}(f^n, J)$ be the distortion of f^n on some interval J. Then, for any subinterval $J' \subset J$ we have*

$$e^{-\mathcal{D}}\frac{|J'|}{|J|} \leq \frac{|f^n(J')|}{|f^n(J)|} \leq e^{\mathcal{D}}\frac{|J'|}{|J|}$$

Proof. By the Mean Value Theorem there exists $x \in J'$ and $y \in J$ such that $|Df^n(x)| = |f^n(J')|/|J'|$ and $|Df^n(y)| = |f^n(J)|/|J|$. Therefore

$$\frac{|f^n(J')|}{|f^n(J)|}\frac{|J'|}{|J|} = \frac{|f^n(J')|/|J'|}{|f^n(J)|/|J|} = \frac{|Df^n(x)|}{|Df^n(y)|} \tag{6}$$

From the definition of distortion we have $e^{-\mathcal{D}} \leq |Df^n(x)|/|Df^n(y)| \leq e^{\mathcal{D}}$ and so substituting this into (6) gives

$$e^{-\mathcal{D}} \leq \frac{|f^n(J')|}{|f^n(J)|}\frac{|J|}{|J'|} \leq e^{\mathcal{D}}$$

and rearranging gives the result. □

Lemma 6.2. *Let $f : I \to I$ be a full branch map with the bounded distortion property. Then* $\max\{|\omega^{(n)}|; \omega^{(n)} \in \mathcal{P}^{(n)}\} \to 0$ *as* $n \to 0$

Proof. First of al let $\delta = \max_{\omega\in\mathcal{P}} |\omega| < |I|$ Then, from the combinatorial structure of full branch maps described in Lemma 5.1 and its proof, we have that for each $n \geq 1$ $f^n(\omega^{(n)}) = I$ and that $f^{n-1}(\omega^{(n)}) \in \mathcal{P}$, and therefore $|f^{n-1}(\omega^{(n)})| \leq \delta$ and $|f^{n-1}(\omega^{(n-1)} \setminus \omega^{(n)}|)| \geq |I| - \delta > 0$. Thus, using Lemma 6.1 we have

$$\frac{|\omega^{(n-1)} \setminus \omega^{(n)}|}{|\omega^{(n-1)}|} \geq e^{-\mathcal{D}} \frac{|f^{n-1}(\omega^{(n-1)} \setminus \omega^{(n)}|)|}{|f^{n-1}(\omega^{(n-1)})|} \geq e^{-\mathcal{D}} \frac{|I| - \delta}{|I|} =: 1 - \tau.$$

Then

$$1 - \frac{|\omega^{(n)}|}{|\omega^{(n-1)}|} = \frac{|\omega^{(n-1)}| - |\omega^{(n)}|}{|\omega^{(n-1)}|} = \frac{|\omega^{(n-1)} \setminus \omega^{(n)}|}{|\omega^{(n)}|} \geq 1 - \tau.$$

Thus for every $n \geq 0$ and every $\omega^{(n)} \subset \omega^{(n-1)}$ we have $|\omega^{(n)}|/|\omega^{(n-1)}| \leq \tau$. Applying this inequality recursively then implies $|\omega^{(n)}| \leq \tau|\omega^{(n-1)}| \leq \tau^2|\omega^{(n-2)}| \leq \cdots \leq \tau^n|\omega^0| \leq \tau^n|\Delta|$. □

Proof of Theorem 7. The proof is almost identical to the piecewise affine case. The only difference is when we get to equation (4) where we now use the bounded distortion to get

$$\frac{|I \setminus A|}{|I|} = \frac{|f^n(\omega_n \setminus A)|}{|f^n(\omega_n)|} \leq e^{\mathcal{D}} \frac{|\omega_n \setminus A|}{|\omega_n|} \leq e^{\mathcal{D}}\epsilon. \tag{7}$$

Since ε is arbitrary this implies $m(A^c) = 0$ and thus $m(A) = 1$. □

6.2. **Sufficient conditions for bounded distortion.** In other cases, the bounded distortion property is not immediately checkable, but we give here some sufficient conditions.

Definition 9. A full branch map f is *uniformly expanding* if there exist constant $C, \lambda > 0$ such that for all $x \in I$ and all $n \geq 1$ such that $x, f(x), \ldots, f^{n-1}(x) \notin \partial\mathcal{P}$ we have $|(f^n)'(x)| \geq Ce^{\lambda n}$.

Theorem 8. *Let f be a full branch map. Suppose that f is uniformly expanding and that there exists a constant $\mathcal{K} > 0$ such that*

$$\sup_{\omega\in\mathcal{P}} \sup_{x,y\in\omega} |f''(x)|/|f'(y)|^2 \leq \mathcal{K}. \tag{8}$$

Then there exists $\tilde{\mathcal{K}} > 0$ such that for every $n \geq 1, \omega^{(n)} \in \mathcal{P}^{(n)}$ and $x, y \in \omega^{(n)}$ we have

$$\log \frac{|Df^n(x)|}{|Df^n(y)|} \leq \tilde{\mathcal{K}}|f^n(x) - f^n(y)| \leq \tilde{\mathcal{K}}. \tag{9}$$

In particular f satisfies the bounded distortion property.

The proof consists of three simple steps which we formulate in the following three lemmas.

Lemma 6.3. *Let f be a full branch map satisfying* (8). *Then, for all $\omega \in \mathcal{P}$, $x, y \in \omega$ we have*

$$\left| \frac{f'(x)}{f'(y)} - 1 \right| \le \mathcal{K}|f(x) - f(y)|. \tag{10}$$

Proof. By the Mean Value Theorem we have $|f(x) - f(y)| = |f'(\xi_1)||x - y|$ and $|f'(x) - f'(y)| = |f''(\xi_2)||x - y|$ for some $\xi_1, \xi_2 \in [x, y] \subset \omega$. Therefore

$$|f'(x) - f'(y)| = \frac{|f''(\xi_2)|}{|f'(\xi_1)|}|f(x) - f(y)|. \tag{11}$$

Assumption (8) implies that $|f''(\xi_2)|/|f'(\xi_1)| \le \mathcal{K}|f'(\xi)|$ for all $\xi \in \omega$. Choosing $\xi = y$ and substituting this into (11) therefore gives $|f'(x) - f'(y)| = \mathcal{K}|f'(y)||f(x) - f(y)|$ and dividing through by $|f'(y)|$ gives the result. □

Lemma 6.4. *Let f be a full branch map satisfying* (10). *Then, for any $n \ge 1$ and $\omega^{(n)} \in \mathcal{P}_n$ we have*

$$Dist(f^n, \omega^{(n)}) \le \mathcal{K} \sum_{i=1}^{n} |f^i(x) - f^i(y)| \tag{12}$$

Proof. By the chain rule $f^{(n)}(x) = f'(x) \cdot f'(f(x)) \cdots f'(f^{n-1}(x))$ and so

$$\begin{aligned}
\log \frac{|f^{(n)}(x)|}{|f^{(n)}(y)|} &= \log \prod_{i=1}^{n} \frac{|f'(f^i(x))|}{|f'(f^i(y))|} = \sum_{i=0}^{n-1} \log \left| \frac{f'(f^i(x))}{f'(f^i(y))} \right| \\
&= \sum_{i=0}^{n-1} \log \left| \frac{f'(f^i(x))}{f'(f^i(y))} - \frac{f'(f^i(y))}{f'(f^i(y))} + \frac{f'(f^i(y))}{f'(f^i(y))} \right| \\
&= \sum_{i=0}^{n-1} \log \left| \frac{f'(f^i(x)) - f'(f^i(y))}{f'(f^i(y))} + 1 \right| \\
&\le \sum_{i=0}^{n-1} \log \left(\frac{|f'(f^i(x)) - f'(f^i(y))|}{|f'(f^i(y))|} + 1 \right) \\
&\le \sum_{i=0}^{n-1} \frac{|f'(f^i(x)) - f'(f^i(y))|}{|f'(f^i(y))|} \qquad \text{using } \log(1+x) < x \\
&\le \sum_{i=0}^{n-1} \left| \frac{f'(f^i(x))}{f'(f^i(y))} - 1 \right| \le \sum_{i=1}^{n} \mathcal{K}|f^i(x) - f^i(y)|.
\end{aligned}$$

□

Lemma 6.5. *Let f be a uniformly expanding full branch map. Then there exists a constant $\tilde{\mathcal{K}}$ depending only on C, λ, such that for all $n \ge 1$, $\omega^{(n)} \in \mathcal{P}_n$ and $x, y \in \omega^{(n)}$ we have*

$$\sum_{i=1}^{n} |f^i(x) - f^i(y)| \le \tilde{\mathcal{K}}|f^n(x) - f^n(y)|.$$

Proof. For simplicity, let $\tilde{\omega} := (x, y) \subset \omega^{(n)}$. By definition the map $f^n|_{\tilde{\omega}} : \tilde{\omega} \to f^n(\tilde{\omega})$ is a diffeomorphism onto its image. In particular this is also true for each map $f^{n-i}|_{f^i(\tilde{\omega})} : f^i(\tilde{\omega}) \to f^n(\tilde{\omega})$. By the Mean Value Theorem we have that

$$|f^n(x) - f^n(y)| = |f^n(\tilde{\omega})| = |f^{n-i}(f^i(\tilde{\omega})| = |(f^{n-i})'(\xi_{n-i})||f^i(\tilde{\omega})| \geq Ce^{\lambda(n-i)}|f^i(\tilde{\omega})|$$

for some $\xi_{n-i} \in f^{n-i}(\tilde{\omega})$. Therefore

$$\sum_{i=1}^{n} |f^i(x) - f^i(y)| = \sum_{i=1}^{n} |f^i(\tilde{\omega})| \leq \sum_{i=1}^{n} \frac{1}{C} e^{-\lambda(n-i)}|f^n(\tilde{\omega})| \leq \frac{1}{C} \sum_{i=0}^{\infty} e^{-\lambda i}|f^n(x) - f^n(y)|.$$

□

6.3. **The Gauss map.** Before proving Theorems 7 and 8 we consider a specific example to which these results apply. Let $I = [0, 1]$ and define the *Gauss map* $f : I \to I$ by $f(0) = 0$ and

$$f(x) = \frac{1}{x} \mod 1$$

if $x \neq 0$. Notice that for every $n \in \mathbb{N}$ the map

$$f : \left(\frac{1}{n+1}, \frac{1}{n}\right] \to (0, 1]$$

is a diffeomorphism. In particular the Gauss map is a full branch map though it is not piecewise affine. Define the *Gauss measure* μ_G by defining, for every measurable set A

$$\mu_G(A) = \frac{1}{\log 2} \int_A \frac{1}{1+x} dx.$$

Theorem 9. *Let $f : I \to I$ be the Gauss map. Then μ_G is invariant and ergodic.*

We prove this in a sequence of Lemmas. Invariance follows by direct verification.

Lemma 6.6. *μ_G is invariant.*

Proof. It is sufficient to prove invariance on intervals $A = (a, b)$. In this case we have

$$\mu_G(A) = \frac{1}{\log 2} \int_a^b \frac{1}{1+x} dx = \frac{1}{\log 2} \log \frac{1+b}{1+a}$$

Each interval $A = (a, b)$ has a countable infinite of pre-images, one inside each interval of the form $(1/n+1, 1/n)$ and this preimage is given explicitly as the interval $(1/n+b, 1/n+a)$.

Therefore

$$\begin{aligned}\mu_G(f^{-1}(a,b)) = \mu_G\left(\bigcup_{n=1}^{\infty}\left(\frac{1}{n+b},\frac{1}{n+a}\right)\right) &= \frac{1}{\log 2}\sum_{n=1}^{\infty}\log\left(\frac{1+\frac{1}{n+a}}{1+\frac{1}{n+b}}\right)\\ &= \frac{1}{\log 2}\log\prod_{n=1}^{\infty}\left(\frac{n+a+1}{n+a}\frac{n+b}{n+b+1}\right)\\ &= \frac{1}{\log 2}\log\left(\frac{1+a+1}{1+a}\frac{1+b}{1+b+1}\frac{2+a+1}{2+a}\frac{2+b}{2+b+1}\cdots\right)\\ &= \frac{1}{\log 2}\log\frac{1+b}{1+a} = \mu_G(a,b).\end{aligned}$$

□

Lemma 6.7. *The Gauss map is uniformly expanding*

Proof. Exercise □

Lemma 6.8. *Let $f : I \to I$ be the Gauss map. Then* $\sup_{\omega\in\mathcal{P}}\sup_{x,y\in\omega}|f''(x)|/|f'(y)|^2 \le 16$.

Proof. Since $f(x) = x^{-1}$ we have $f'(x) = -x^{-2}$ and $f''(x) = 2x^{-3}$. Notice that both first and second derivatives are monotone decreasing, i.e. take on larger values close to 0. Thus, for a generic interval $\omega = (1/(n+1), 1/n)$ of the partition $\mathcal{P}$ we have $|f''(x)| \le f''(1/(n+1)) = 2(n+1)^3$ and $|f'(y)| \ge |f'(1/n)| = n^2$. Therefore, for any $x, y \in \omega$ we have $|f''(x)|/|f'(y)|^2 \le 2(n+1)^3/n^4 \le 2((n+1)/n)^3(1/n)$. This upper bound is monotone decreasing with n and thus the worst case is $n = 1$ whch gives $|f''(x)|/|f'(y)|^2 \le 16$ as required. □

We remark that Lebesgue measure is not generally invariant if f is not piecewise affine. However the notion of ergodicity still holds and the ergodicity of Lebesgue measure implies the ergodicity any other measure which is absolutely continuous. More generally, we have the following.

Lemma 6.9. *Let $f : I \to I$ be a measurable map and let μ_1, μ_2 be two probability measures with $\mu_1 \ll \mu_2$. Suppose μ_2 is ergodic for f. Then μ_1 is also ergodic for f.*

Proof. Suppose $A \subseteq I$ with $\mu_1(A) > 0$. Then by the absolute continuity this implies $\mu_2(A) > 0$; by ergodicity of μ_2 this implies $\mu_2(A) = 1$ and therefore $\mu_2(I \setminus A) = 0$; and so by absolute continuity, also $\mu_1(I \setminus A) = 0$ and so $\mu_1(A) = 1$. Thus μ_1 is ergodic. □

Proof of Theorem 9. From Lemmas 6.7 and 6.8 we have that Lebesgue measure is ergodic for the Gauss map f. Since the Gauss measure $\mu_G \ll m$ ergodicity of μ_G then follows from Lemma 6.9. □

7. Physical measures for full branch maps

We have proved a general ergodicity result for Lebesgie measure for a relatively large class of maps satisfying the bounded distortion property, but we only have two specific examples of such maps for which we actually have an absolutely continuous invariant measure. In this section we prove that such a measure actually always exists.

Theorem 10. *Let $f : I \to I$ be a full branch map satisfying* (9)*. Then f admits a unique ergodic absolutely continuous invariant probability measure μ. Morever, the density $d\mu/dm$ of μ is Lipschitz continuous and bounded above and below.*

[3] We begin in exactly the same way as for the proof of the existence of invariant measures for general continuous maps and define the sequence

$$\mu_n = \frac{1}{n}\sum_{i=0}^{n} f_*^i m$$

where m denotes Lebesgue measure.

Exercise 7. For each $n \geq 1$ we have $\mu_n \ll m$. *Hint:* by definition f is a C^2 diffeomorphism on (the interior of) each element of the partition $\mathcal{P}$ and thus in particular it is non-singular in the sense that $m(A) = 0$ implies $m(f^{-1}(A) = 0$ for any measurable set A.

Since $\mu_n \ll m$ we can let

$$H_n := \frac{d\mu_n}{d_m}$$

denote the density of μ_n with respct to m. The proof of the Theorem then relies on the following crucial

Proposition 7.1. *There exists a constant $K > 0$ such that*

$$0 < \inf_{n,x} H_n(x) \leq \sup_{n,x} H_n(x) \leq K \tag{13}$$

and for every $n \geq 1$ and every $x, y \in I$ we have

$$|H_n(x) - H_n(y)| \leq K|H_n(x)|d(x,y) \leq K^2 d(x,y). \tag{14}$$

[3]

7.1. **Absolutely continuous invariant ergodic probability measures.** A natural approach to the general study of the existence (or not) of physical measures is to study the conditions which imply that a system has at least one physical measure. Related to this approach is the question of which kind of measures can be physical measures. The easiest example of a physical measure μ is when μ is f-invariant, ergodic and $\mu \ll m$. Then, by the definition of absolute continuity we have $\mu(\mathfrak{B}_\mu) = 1$ implies $m(\mathfrak{B}_\mu) > 0$. Therefore the fact that μ is physical follows in this case directly from Birkhoff's ergodic theorem. In this case, the question of the existence of physical measures therefore reduces to the following question

Question 1. *Does f admit an absolutely continuous, invariant, ergodic probability μ?*

For simplicity we shall often refer to such a measure μ as an *acip*. This will also be the main question we address in these notes. We mention however that singular measures can also be physical measures. For example, suppose p is a fixed point which is *attracting* in the sense that it has a neighbourhood $\mathcal{U}$ such that $f^n(x) \to p$ for all $x \in U$. Then, the measure $\mu = \delta_p$ is ergodic and invariant but Birkhoff's ergodic theorem alone does not imply that it is a physical measure. On the other hand, it is easy to check directly that for all $x \in \mathcal{U}$, the time averages also converge to the space average for all continuous functions φ and so δ_p is indeed a physical measure.

Proof of Theorem assuming Proposition 7.1. The Proposition says that the family $\{H_n\}$ is bounded and equicontinuous and therefore, by Ascoli-Arzela Theorem there exists a subsequence H_{n_j} converging uniformly to a function H satisfying (13) and (14). We define the measure μ by defining, for every measurable set A,

$$\mu(A) := \int_A H dm.$$

Then μ is absolutely continuous with respect to Lebesgue by definition, its density is Lipschitz continuous and bounded above and below, and it is ergodic by the ergodicity of Lebesgue measure and the absolute continuity. It just remains to prove that it is invariant. Notice first of all that for any measurable set A we have

$$\begin{aligned}\mu(A) &= \int_A H dm = \int_A \lim_{n_j\to\infty} H_{n_j} dm = \lim_{n_j\to\infty} \int_A H_{n_j} dm \\ &= \lim_{n_j\to\infty} \mu_{n_j}(A) = \lim_{n_j\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} f_*^i m(A) = \lim_{n_j\to\infty} \frac{1}{n_j}\sum_{i=0}^{n_j-1} m(f^{-i}(A))\end{aligned}$$

For the third equality we have used the dominated convergence theorem to allow us to pull the limit outside the integral. From this we can then write

$$\begin{aligned}\mu(f^{-1}(A)) &= \lim_{n_j\to\infty} \frac{1}{n_j}\sum_{i=0}^{n_j-1} m(f^{-i}(f^{-1}(A)) \\ &= \lim_{n_j\to\infty} \frac{1}{n_j}\sum_{i=1}^{n_j} m(f^{-i}(A) \\ &= \lim_{n_j\to\infty} \left(\frac{1}{n_j}\sum_{i=0}^{n_j-1} m(f^{-i}(A) + \frac{1}{n_j} f^{-n_j}(A) - \frac{1}{n_j} m(A)\right) \\ &= \lim_{n_j\to\infty} \frac{1}{n_j}\sum_{i=0}^{n_j-1} m(f^{-i}(A) \\ &= \mu(A).\end{aligned}$$

This shows that μ is invariant and completes the proof. □

Remark 5. The fact that $\mu_n \ll m$ for every n does not imply that $\mu \ll m$. Indeed, consider the following example. Suppose $f : [0,1] \to [0,1]$ is given by $f(x) = x/2$. We alreeady know that in this case the only physical measure is the Dirac measure at the unique attracting fixed point at 0. In this simple setting we can see directly that $\mu_n \to \delta_0$ where μ_n are the averages defined above. In fact we shall show that stronger statement that $f_*^n m \to \delta_p$ as $n \to \infty$. Indeed, let $\mu_0 = m$. And consider the measure $\mu_1 = f_* m$ which is give by definition by $\mu_1(A) = \mu_0(f^{-1}(A))$. Then it is easy to see that $\mu_1([0,1/2]) = \mu_0(f^{-1}([0.1/2])) = \mu_0([0,1]) = 1$. Thus the measure μ_1 is completely concentrated on the interval $[0,1/2]$. Similarly, it is easy to see that $\mu_n([0,1/2^n]) = \mu_0([0,1]) = 1$ and thus the measure μ_n is completely concetrated on the interval $[0.1/2^n]$. Thus the measures μ_n are concentrated on increasingly smaller neighbourhood of the origin 0. This clearly implies that they are converging in the weak star topology to the Dirac measure at 0.

This counter-example shows that a sequence of absolutely continuous measures does not necessarily converge to an absolutely continuous measures. This is essentially related to the fact that a sequence of L^1 functions (the densities of the absolutely continuous measures μ_n) may not converge to an L^1 function even if they are all uniformly bounded in the L^1 norm.

It just remains to prove Proposition 7.1. We start by finding an explicit formula for the functions H_n.

Lemma 7.1. *For every $n \geq 1$ and every $x \in I$ we have*

$$H_n(x) = \frac{1}{n}\sum_{i=1}^{n-1} S_n(x) \quad \textit{where} \quad S_n(x) := \sum_{y=f^{-i}(x)} \frac{1}{|Df^n(y)|}.$$

Proof. It is sufficient to show that S_n is the density of the measure $f_*^n m$ with respect to m, i.e. that $f_*^n m(A) = \int_A S_n dm$. By the definition of full branch map, each point has exactly one preimage in each element of $\mathcal{P}$. Since $f : \omega \to I$ is a diffeomorphism, by standard calculus we have

$$m(A) = \int_{f^{-n}(A)\cap\omega} |Df^n| dm \quad \text{and} \quad m(f^{-n}(A)\cap\omega) = \int_A \frac{1}{|Df^n(f^{-n}(x)\cap\omega)|} dm.$$

Therefore

$$f_*^n m(A) = m(f^{-n}(A)) = \sum_{\omega\in\mathcal{P}_n} m(f^{-n}(A)\cap\omega) = \sum_{\omega\in\mathcal{P}_n} \int_A \frac{1}{|Df^n(f^{-n}(x)\cap\omega)|} dm$$
$$= \int_A \sum_{\omega\in\mathcal{P}_n} \frac{1}{|Df^n(f^{-n}(x)\cap\omega)|} dm = \int_A \sum_{y\in f^{-n}(x)} \frac{1}{|Df^n(y)|} dm = \int_A S_n dm.$$

□

Lemma 7.2. *There exists a constant $K > 0$ such that*

$$0 < \inf_{n,x} S_n(x) \leq \sup_{n,x} S_n(x) \leq K$$

and for every $n \geq 1$ and every $x, y \in I$ we have

$$|S_n(x) = S_n(y)| \leq K|S_n(x)|d(x,y) \leq K^2 d(x,y).$$

Proof. The proof uses in a fundamental way the bounded distortion property (9). Recall that for each $\omega \in \mathcal{P}_n$ the map $f^n : \omega \to I$ is a diffeomorphism with uniformly bounded distortion. This means that $|Df^n(x)/Df^n(y)| \leq \mathcal{D}$ for any $x, y \in \omega$ and for any $\omega \in \mathcal{P}_n$ (uniformly in n). Informally this says that the derivative Df^n is essentially the same at all points of each $\omega \in \mathcal{P}_n$ (although it can be wildly different in principle between different ω's). By the Mean Value Theorem, for each $\omega \in \mathcal{P}_n$, there exists a $\xi \in \omega$ such that $|I| = |Df^n(\xi)||\omega|$ and therefore $|Df^n(\xi)| = 1/|\omega|$ (assuming the length of the entire interval I is normalized to 1). But since the derivative at every point of ω is comparable to that at ξ we have in particular $|Df^n(y)| \approx 1/|\omega|$ and therefore

$$S_n(x) = \sum_{y \in f^{-n}(x)} \frac{1}{|Df^n(y)|} \approx \sum_{\omega \in \mathcal{P}_n} |\omega| \leq K.$$

To prove the uniform Lipschitz continuity recall that the bounded distortion property (9) gives

$$\left|\frac{Df^n(x)}{Df^n(y)}\right| \leq e^{Kd(f^n(x), f^n(y)} \leq 1 + \tilde{K} d(f^n(x), f^n(y)).$$

Inverting x, y we also have

$$\left|\frac{Df^n(y)}{Df^n(x)}\right| \geq \frac{1}{1 + \tilde{K} d(f^n(x), f^n(y))} \geq 1 - \tilde{\tilde{K}} d(f^n(x), f^n(y)).$$

Combining these two bounds we get

$$\left|\frac{Df^n(x)}{Df^n(y)} - 1\right| \leq \hat{K} d(f^n(x), f^n(y))\}.$$

where $\hat{K} = \max\{\tilde{K}, \tilde{\tilde{K}}\}$ For $x, y \in I$ we have

$$\begin{aligned}
|S_n(x) - S_n(y)| &= \left| \sum_{\tilde{x} \in f^{-n}(x)} \frac{1}{|Df^n(\tilde{x})|} - \sum_{\tilde{y} \in f^{-n}(y)} \frac{1}{|Df^n(\tilde{y})|} \right| \\
&= \left| \sum_{i=1}^{\infty} \frac{1}{|Df^n(\tilde{x}_i)|} - \sum_{i=1}^{\infty} \frac{1}{|Df^n(\tilde{y}_i)|} \right| \qquad \text{where } f^n(\tilde{x}_i) = x, \ f^n(\tilde{y}_i) = y \\
&\leq \sum_{i=1}^{\infty} \left| \frac{1}{|Df^n(\tilde{x})|} - \frac{1}{|Df^n(\tilde{y})|} \right| = \sum_{i=1}^{\infty} \frac{1}{|Df^n(\tilde{x}_i)|} \left| 1 - \frac{Df^n(\tilde{x}_i)}{Df^n(\tilde{y}_i)} \right| \\
&\leq \hat{K} \sum_{i=1}^{\infty} \frac{1}{|Df^n(\tilde{x}_i)|} d(f^n(\tilde{x}_i), f^n(\tilde{y}_i)) \leq \hat{K} \sum_{i=1}^{\infty} \frac{1}{|Df^n(\tilde{x}_i)|} d(x, y) = \hat{K} S_n(x) d(x, y).
\end{aligned}$$

□

Proof of Proposition 7.1. This Lemma clearly implies the Proposition since

$$\begin{aligned}
|H_n(x) - H_n(y)| &= |\frac{1}{n} \sum S_i(x) - \frac{1}{n} \sum S_i(y)| \leq \frac{1}{n} \sum |S_i(x) - S_i(y)| \\
&\leq \frac{1}{n} \sum K S_i(x) d(x, y) = H_n K d(x, y) \leq K^2 d(x, y).
\end{aligned}$$

□

8. Invariant measures via conjugacy

Now that we have a general theorem for the existence of physical measures for full branch maps with bounded distortion, we can use these results to obtain physical measures for maps which are not full branch or do not satisfy the bounded distortion conditions. We will describe two techniques to achieve this. The first one is to relate two dynamical systems via the notion of conjugacy.

Definition 10. Let X, Y be two metric spaces and $f : X \to X$ and $g : Y \to Y$ be two maps. We say that f and g are *conjugate* if there exists a bijection $h : X \to Y$ such that $h \circ f = g \circ h$. or, equivalently, $f = h^{-1} \circ g \circ h$.

A conjugacy h maps orbits of f to orbits of g.

Exercise 8. Show that if f, g are conjugate, then $f^n(x) = h^{-1} \circ g^n \circ h(x)$ for every $n \geq 1$.

In particular a conjugacy naps fixed points to fixed points and periodic points to corresponding periodic points. However, without additional assumptions on the regularity of h it may not preserve additional structure. We that f, g are *(Borel) measurably conjugate* if h, h^{-1} are (Borel) measurable, *topologically conjugate* if h is a homeomorphism, and C^r conjugate, $r \geq 1$, if h is a C^r diffeomorphism.

Exercise 9. Show that conjugacy defines an equivalence relation on the space of all dynamical systems. Show that measurable, topological, and C^r conjugacy, each defines an equivalence relation on the space of dynamical systems.

Measurable conjugacies map sigma-algebras to sigma-algebras and therefore we can define a map

$$h_* : \mathcal{M}(X) \to \mathcal{M}(Y)$$

from the space $\mathcal{M}(X)$ of all probability measures on X to the space $\mathcal{M}(Y)$ of all probabiulity measures on Y, by

$$h_*\mu(A) = \mu(h^{-1}(A)).$$

Lemma 8.1. *Suppose f, g are measurably conjugate. Then*

(1) *$h_*\mu$ is invariant under g if and only if μ is invariant under f.*
(2) *$h_*\mu$ is ergodic for g if and only if μ is ergodic for f.*

Proof. Exercise. (Hint: Indeed, for any measurable set $A \subseteq Y$ we have $\mu_Y(g^{-1}(A)) = \mu_X(h^{-1}(g^{-1}(A))) = \mu_X((h^{-1} \circ g^{-1})(A)) = \mu_X((g \circ h)^{-1}(A)) = \mu_X((h \circ f)^{-1}(A)) = \mu_X(f^{-1}(h^{-1}(A))) = \mu_X(h^{-1}(A) = \mu_Y(A)$.). For ergodicity, let $A \subset Y$ satisfy $g^{-1}(A) = A$. Then, it's preimage by the conjugacy satisfies the same property, i.e. $f^{-1}(h^{-1}(A)) = h^{-1}(A)$. Thus, by the ergodicity of μ we have either $\mu(h^{-1}(A)) = 0$ or $\mu(h^{-1}(A)) = 1$. □

We can therefore find invariant measure for a dynamical systems if we have information about invariant measures for conjugate systems. We will give two applications of this strategy.

8.1. **The Ulam-von Neumann map.** Define the *Ulam-von Neumann* map $f : [-2, 2] \to [-2, 2]$ by

$$f(x) = x^2 - 2.$$

Proposition 8.1. *The measure μ defined by*

$$\mu(A) = \frac{2}{\pi} \int_A \frac{1}{\sqrt{4 - x^2}} dm$$

is invariant and ergodic for f.

The invariance of μ can in principle be checked directly by computing explicity preimages of intervals, as for the Gauss map. The ergodicity however is non-trivial. We use the pull-back strategy to get both at the same time, and also to explain how the measure μ is computed in the first place.

Consider the piecewise affine *tent map* $T : [0, 1] \to [0, 1]$ defined by

$$T(z) = \begin{cases} 2z, & 0 \le z < \frac{1}{2} \\ 2 - 2z, & \frac{1}{2} \le z \le 1. \end{cases}$$

Lemma 8.2. *The map $h : [0, 1] \to [-2, 2]$ defined by*

$$h(z) = 2 \cos \pi z.$$

is a conjugacy between f and T.

Proof. Notice that h is a bijection and both h and h^{-1} are smooth in the interior of their domains of definition. Moreover, if $y = h(z) = 2 \cos \pi z$, then $z = h^{-1}(y) = \pi^{-1} \cos^{-1}(y/2)$. Therefore

$$\begin{aligned} h^{-1}(f(h(x))) &= \frac{1}{\pi} \cos^{-1}\left(\frac{f(h(x))}{2}\right) = \frac{1}{\pi} \cos^{-1}\left(\frac{(2\cos \pi x)^2 - 2}{2}\right) \\ &= \frac{1}{\pi} \cos^{-1}(2\cos^2 \pi x - 1) = \frac{1}{\pi} \cos^{-1}(\cos 2\pi x) = T(x). \end{aligned}$$

For the last equality, notice that for $x \in [0, 1/2]$ we have $2\pi x \in [0, \pi]$ and so $\pi^{-1} \cos^{-1}(\cos 2\pi x) = 2x$. On the other hand, for $x \in [1/2, 1]$ we have $2\pi x \in [\pi, 2\pi]$ and so $\cos^{-1}(\cos 2\pi x) = -\cos^{-1}(\cos(2\pi x - 2\pi)) = -\cos^{-1}(\cos 2\pi(x-1)) = -2\pi(x-1)$ and therefore $\pi^{-1} \cos^{-1}(\cos 2\pi x) = -2(x-1) = -2x - 2$. $\square$

Thus, any ergodic invariant measure for T can be "pulled back" to an ergodic invariant measure for f using the conjugacy h.

Proof of Proposition 8.1. We will show that $\mu = h_* m$ which implies immediately that it is ergodic and invariant since Lebegue measure m is ergodic and invariant for T. Using the explicit form of h^{-1} and differentiating, we have

$$(h^{-1})'(x) = \frac{1}{\pi} \frac{-1}{\sqrt{1 - \frac{x^2}{4}}} = \frac{2}{\pi} \frac{-1}{\sqrt{4 - x^2}}$$

and therefore, for aninterval $A = (a, b)$ we have, using the fundamental theorem of calculus,

$$h_* m(A) = m(h^{-1}(A) = \int_a^b |(h^{-1})'(x)| dx = \frac{2}{\pi} \int_a^b \frac{1}{\sqrt{4 - x^2}} dx.$$

□

8.2. **Uncountably many non-atomic ergodic measures.** We now use the pull-back method to show construct an uncountable family of ergodic invariant measures. We recall that a measure is called *non-atomic* if there is no individual point which has positive measure.

Proposition 8.2. *The interval map $f(x) = 2x \mod 1$ admits an uncountable family of non-atomic, mutually singular, ergodic measures.*

We shall construct these measures quite explicitly and thus obtain some additional information about their properties. the method of construction is of intrinsic interest. For each $p \in (0, 1)$ let $I^{(p)} = [0, 1)$ and define the map $f_p : I^{(p)} \to I^{(p)}$ by

$$f_p = \begin{cases} \frac{1}{p} x & \text{for } 0 \le x < p \\ \frac{1}{1-p} x - \frac{p}{1-p} & \text{for } p \le x < 1. \end{cases}$$

Lemma 8.3. *For any $p \in (0, 1)$ the maps f and f_p are topologically conjugate.*

Proof. This is a standard proof in topological dynamics and we just give a sketch of the argument here because the actual way in which the conjugacy h is constructed plays a crucial role in what follows. We use the *symbolic dynamics* of the maps f and f_p. Let

$$I_0^{(p)} = [0, p) \quad \text{and } I_1^{(p)} = (p, 1].$$

Then, for each x we define the symbol sequence $(x_0^{(p)} x_1^{(p)} x_2^{(p)} \dots) \in \Sigma_2^+$ by letting

$$x_i^{(p)} = \begin{cases} 0 \text{ if } f^i(x) \in I_0^{(p)} \\ 1 \text{ if } f^i(x) \in I_1^{(p)}. \end{cases}$$

This sequence is well defined for all points which are not preimages of the point p. Moreover it is unique since every interval $[x, y]$ is expanded at least by a factor $1/p$ at each iterations and therefore $f^n([x, y])$ grows exponentially fast so that eventually the images of $f^n(x)$ and $f^n(y)$ must lie on opposite sides of p and therefore give rise to different sequences. The map $f : I \to I$ is of course just a special case of $f_p : I^{(p)} \to I^{(p)}$ with $p = 1/2$. We can therefore define a bijection

$$h_p : I^{(p)} \to I$$

which maps points with the same associated symbolic sequence to each other and points which are preimages of p to corresponding preimages of $1/2$.

Exercise 10. Show that h_p is a conjugacy between f and f_p.

Exercise 11. Show that h_p is a homeomorphism. *Hint:* if x does not lie in the pre-image of the discontinuity ($1/2$ or p depending on which map we consider) then sufficiently close points y will have a symbolic sequence which coincides with that of x for a large number of terms, where the number of terms can be made arbitrarily large by choosing y sufficiently close to x. The corresponding points therefore also have symbolic sequences which coincide for a large number of terms and this implies that they must be close to each other.

From the previous two exercises it follows that h is a topological conjugacy. □

Since $h_p : I^{(p)} \to I$ is a topological conjugacy, it is also in particular measurable conjugacy and so, letting m denote Lebesgue measure, we define the measure

$$\mu_p = h_* m.$$

By Proposition 5.1 Lebesgue measure is ergodic and invariant for f_p and so it follows from Lemma 8.1 that μ_p is ergodic and invariant for f.

Exercise 12. Show that μ_p is non-atomic.

Thus it just remains to show that the μ_p are mutually singular.

Lemma 8.4. *The measures in the family* $\{\mu_p\}_{p\in(0,1)}$ *are all mutually singular.*

Proof. The proof is a straightforward, if somewhat subtle, application of Birkhoff's Ergodic Theorem. Let

$$A_p = \{x \in I \text{ whose symbolic coding contain asymptotically a proportion } p \text{ of 0's}\}$$

and

$$A_p^{(p)} = \{x \in I^{(p)} \text{ whose symbolic coding contain asymptotically a proportion } p \text{ of 0's}\}$$

Notice that by the way the coding has been defined the asymptotic propertion of 0's in the symbolic coding of a point x is exactly the asymptotic relative frequency of visits of the orbit of the point x to the interval I_0 or $I_0^{(p)}$ under the maps f and f_p respectively. Since Lebesgue measure is invariant and ergodic for f_p, Birkhoff implies that the relative frequence of visits of Lebesgue almost every point to $I_0^{(p)}$ is asymptotically equal to the Lebesgue measure of $I_0^{(p)}$ which is exactly p. Thus we have that

$$m(A_p^{(p)}) = 1.$$

Moreover, since the conjugacy preserves the symbolic coding we have

$$A_p = h(A_p^{(p)}).$$

Thus, by the definition of the pushforward measure

$$\mu_p(A_p) = m(h^{-1}(A_p)) = m(h^{-1}(h(A_p^{(p)})) = m(A_p^{(p)}) = 1.$$

Since the sets A_p are clearly pairwise disjoint for distinct values of p it follows that the measures μ_p are mutually singular.

□

Remark 6. This example shows that the conjugacies in question, even though they are homeomorphisms, are *singular* with respect to Lebesgue measure, i.e. thay maps sets of full measure to sets of zero measure.

9. Invariant measures via inducing

It is not always easy or even possible to conjugate an unnkown dynamical system to a known one and thus the method of pull-back described in the previous section has limited applicability. We describe here another method which, on the other hand, turns out to be quite generally applicable. This is based on the general notion of *inducing*.

Definition 11. Let $f : X \to X$ be a map, $\Delta \subseteq X$ and $\tau : \Delta \to \mathbb{N}$ a function such that $f^{\tau(x)}(x) \in \Delta$ for all $x \in \Delta$. Define a map $F : \Delta \to \Delta$ by

$$F(x) := f^{\tau(x)}(x).$$

The map F is called the *induced map* of f for the *return time* function τ.

If $\Delta = X$ then any function τ can be used to define an induced map, on the other hand, if Δ is a proper subset of X then the requirement $f^{\tau(x)}(x) \in \Delta$ is a non-trivial restriction. The map $F : \Delta \to \Delta$ can be considered as a dynamical system in its own right and therefore has its own dynamical properties which might be, at least a priori, completely different from those of the original map f. However it turns out that there is a close relation between certain dynamical properties of F, in particular invariant measures, and dynamical properties of the original map f, and we analyze this relationship below. We recall that a map $f : X \to X$ is *non-singular* with respect to a measure μ if it maps positive measure sets to positive measure sets: $\mu(A) > 0$ implies $\mu(f(A)) > 0$ or, equivalently, $m(A) = 0$ implies $m(f^{-1}(A)) = 0$.

Theorem 11. *Let $f : X \to X$ be a map and $F := f^\tau : \Delta \to \Delta$ the induced map on some subset $\Delta \subseteq X$ corresponding to the return time function $\tau : \Delta \to \mathbb{N}$. Let $\hat{\mu}$ be a probability measure on Δ and suppose that $\hat{\tau} := \int \tau d\hat{\mu} < \infty$. Then*

$$\mu := \frac{1}{\hat{\tau}} \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} f_*^i(\hat{\mu}|_{\Delta_n})$$

is a probability measure on X. Moreover,

(1) *If $\hat{\mu}$ is invariant for F then μ is invariant for f.*
(2) *If $\hat{\mu}$ is ergodic for F then μ is ergodic for f.*

Suppose additionally that there exists a reference measure m on X and that f is non-singular with respect to m. Then

(3) *If $\hat{\mu} \ll m$ then $\mu \ll m$.*

As an immediate consequence we have the following

Corollary 9.1. *Suppose $f : X \to X$ is non-singular with respect to Lebesgue measure and there exists a subset $\Delta \subseteq X$ and an induced map $F : \Delta \to \Delta$ which admits an* *invariant, ergodic, absolutely continuous probability measure $\hat{\mu}$ for which the return time τ is integrable, then f admits an invariant ergodic absolutely continuous probability measure.*

Remark 7. Notice that Theorem 11 and its Corollary are quite general and in particular apply to maps in arbitrary dimension.

Proof of Theorem 11. For convenience, we introduce the notation

$$\Delta_n := \{x \in \Delta : \tau(x) = n\}.$$

By the measurability of τ each Δ_n is a measurable set. To prove that μ is a probability measure observe first of all that for a measurable set $B \subseteq X$, we have $f_*^i(\hat{\mu}|_{\Delta_n})(B) = \hat{\mu}|_{\Delta_n}(f^{-i}(B)) = \hat{\mu}(f^{-i}(B) \cap \Delta_n)$ and therefore

$$\mu(B) := \frac{1}{\hat{\tau}} \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} f_*^i(\hat{\mu}|_{\Delta_n})(B) = \frac{1}{\hat{\tau}} \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} \hat{\mu}(f^{-i}(B) \cap \Delta_n).$$

This shows that μ is a well defined measure and also shows the way the measure is constructed by "spreading" the measure $\hat{\mu}$ around using the dynamics. To see that μ is a probability measure we write

$$\mu(X) = \frac{1}{\hat{\tau}} \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} \hat{\mu}(f^{-i}(X) \cap \Delta_n) = \frac{1}{\hat{\tau}} \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} \hat{\mu}(\Delta_n) = \frac{1}{\hat{\tau}} \sum_{n=1}^{\infty} n\hat{\mu}(\Delta_n) = \frac{1}{\hat{\tau}} \int \tau d\hat{\mu} = 1$$

To prove (1), suppose that $\hat{\mu}$ is F-invariant, and therefore $\hat{\mu}(B) = \hat{\mu}(F^{-1}(B))$ for any measurable set B. We will show first that

$$\sum_{n=1}^{\infty} \hat{\mu}(B \cap \Delta_n) = \sum_{n=1}^{\infty} \hat{\mu}(f^{-n}(B) \cap \Delta_n)) \tag{15}$$

Since the sets Δ_n are disjoint and their union is Δ, the sum on the right hand side is exactly $\hat{\mu}(B)$. So we just need to show that the sum on the right hand side is $\hat{\mu}(F^{-1}(B))$. By the definition of F we have

$$F^{-1}(B) = \{x \in \Delta : F(x) \in B\} = \bigcup_{n=1}^{\infty} \{x \in \Delta_n : f^n(x) \in B\} = \bigcup_{n=1}^{\infty} (f^{-n}(B) \cap \Delta_n).$$

Since the Δ_n are disjoint, for any measure $\hat{\mu}$ we have

$$\hat{\mu}(F^{-1}(B)) = \hat{\mu}(\bigcup_{n=1}^{\infty} (f^{-n}(B) \cap \Delta_n)) = \sum_{n=1}^{\infty} \hat{\mu}(f^{-n}(B) \cap \Delta_n)).$$

This proves (15) and therefore implies

$$\begin{aligned}
\mu(f^{-1}(B)) &= \sum_{n=1}^{\infty}\sum_{i=0}^{n-1}\hat{\mu}(f^{-i}(f^{-1}(B))\cap\Delta_n)\\
&= \sum_{n=1}^{\infty}\hat{\mu}(f^{-1}(B)\cap\Delta_n)+\hat{\mu}(f^{-2}(B)\cap\Delta_n)+\cdots+\hat{\mu}(f^{-n}(B)\cap\Delta_n)\\
&= \sum_{n=1}^{\infty}\sum_{i=1}^{n-1}\hat{\mu}(f^{-i}(B)\cap\Delta_n)+\sum_{n=1}^{\infty}\hat{\mu}(f^{-n}(B)\cap\Delta_n)\\
&= \sum_{n=1}^{\infty}\sum_{i=1}^{n-1}\hat{\mu}(f^{-i}(B)\cap\Delta_n)+\sum_{n=1}^{\infty}\hat{\mu}(B\cap\Delta_n)\\
&= \sum_{n=1}^{\infty}\sum_{i=0}^{n-1}\hat{\mu}(f^{-i}(B)\cap\Delta_n)\\
&= \mu(B).
\end{aligned}$$

This shows that μ is invariant and thus completes the proof of (1). To prove (2), assume that $\hat{\mu}$ is ergodic. Now let $B \subseteq X$ satisfy $f^{-1}(B) = B$ and $\mu(B) > 0$. We will show that $\mu(B) = 1$ thus implying that μ is ergodic. Let $\hat{B} = B \cap \Delta$. We first show that

$$F^{-1}(\hat{B}) = \hat{B} \quad \text{and} \quad \hat{\mu}(\hat{B}) = 1. \tag{16}$$

Indeed, $f^{-1}(B) = B$ implies $f^{-n}(\hat{B}) = f^{-n}(B) \cap f^{-n}(\Delta) = B \cap f^{-n}(\Delta)$ and therefore

$$F^{-1}(\hat{B}) = \bigcup_{n=1}^{\infty}(f^{-n}(\hat{B})\cap\Delta_n) = \bigcup_{n=1}^{\infty}(B\cap f^{-n}(\Delta)\cap\Delta_n) = \bigcup_{n=1}^{\infty}(B\cap\Delta_n) = B\cap\Delta = \hat{B}$$

where the third equality follows from the fact that $\Delta_n := \{x : \tau(x) = n\} \subseteq \{x : f^n(x) \in \Delta\} = f^{-n}(\Delta)$. Now, from the definition of μ we have that $f^{-1}(B) = B$ and $\mu(B) > 0$ imply $\hat{\mu}(B \cap \Delta_n) = \hat{\mu}(f^{-i}(B) \cap \Delta_n) > 0$ for some $n > i \geq 0$ and therefore $\hat{\mu}(\hat{B}) = \hat{\mu}(B \cap \Delta) > 0$. Thus, by the ergodicity of $\hat{\mu}$, we have that $\hat{\mu}(B \cap \Delta) = \hat{\mu}(\hat{B}) = 1$ and this proves (16), and thus in particular, letting $B^c := X \setminus B$ denote the complement of B, we have that $\hat{\mu}(B^c \cap \Delta) = 0$ and therefore

$$\mu(B^c) := \sum_{n=1}^{\infty}\sum_{i=0}^{n-1}\hat{\mu}(f^{-i}(B^c)\cap\Delta_n) = \sum_{n=1}^{\infty}\sum_{i=0}^{n-1}\hat{\mu}(B^c\cap\Delta_n) = 0$$

This implies that $\mu(B) = 1$ and thus completes the proof of (2). Finally (3) follows directly from the definition of μ. □

9.1. **Physical measure for intermittency maps.** We give a relatively simple but non-trivial application of the method of inducing. Let $\gamma \geq 00$ and consider the map $f_\gamma : [0,1] \to [0,1]$ given by

$$f_\gamma(x) = x + x^{1+\gamma} \mod 1.$$

For $\gamma > 0$ this can be thought of as a perturbation of the map $f(x) = 2x \bmod 1$ (for $\gamma = 0$) (though it is a C^0 perturbation and *not* a C^1 perturbation). It is a full branch map, but it fails to satisfy both the uniform expansivity and the bounded distortion condition since

$$f'_\gamma(x) = 1 + (1+\gamma)x^\gamma$$

and so in particular for the fixed point at the origin we have $f'(0) = 1$ and thus $(f^n)'(0) = 1$ for all $n \geq 1$. Nevertheless we will still be able to prove the following:

Theorem 12. *For any $\gamma \in [0,1)$ the map f_γ admits a unique ergodic absolutely continuous invariant probability measure.*

We first construct the full branch induced map, then show that it satisfies the uniform expansivity and distortion conditions and finally check the integrability of the return times. Let $x_1 := 1$, let x_2 denote the point in the interior of $[0,1]$ at the boundary between the two domains on which f is smooth, and let $\{x_n\}_{n=3}^\infty$ denote the branch of pre images of x_2 converging to the fixed point at the origin, so that we have $x_n \to 0$ monotonically and and $f(x_{n+1}) = x_n$. For each $n \geq 1$ we let $\Delta_n = (x_{n+1}, x_n]$. Then, the intervals Δ_n form a partition of $\Delta := (0,1]$ and there is a natural induced map $F : \Delta \to \Delta$ given by $F|_{\Delta_n} = f^n$ such that $F : \Delta_n \to \Delta$ is a C^1 diffeomorphism.

Lemma 9.1. *F is uniformly expanding.*

Proof. Exercise. □

It remains to show therefore that F has bounded distortion and that the inducing times are integrable. For both of these results we need some estimates on the size of the partition elements Δ_n. To simplify the exposition, we shall use the following notation. Given two sequences $\{a_n\}$ and $\{b_n\}$ we use the notation $a_n \approx b_n$ to mean that there exists a constant C such that $C^{-1}b_n \leq a_n \leq Cb_n$ for all n and $a_n \lesssim b_n$ to mean that $a_n \leq Cb_n$ for all n.

Lemma 9.2. $x_n \approx 1/n^{\frac{1}{\gamma}}$ *and* $|\Delta_n| \approx 1/n^{\frac{1}{\gamma}+1}$.

Proof. First of all notice that since $x_n = f(x_{n+1}) = x_{n+1} + x_{n+1}^{1+\gamma}$ we have

$$|\Delta_n| = |x_n - x_{n+1}| = x_{n+1}^{1+\gamma}$$

Also, the ratio between x_n and x_{n+1} is bounded since

$$x_n/x_{n+1} = (x_{n+1} + x_{n+1}^{1+\gamma})/x_{n+1} = 1 + x_{n+1}^\gamma \to 1$$

as $n \to \infty$. So in fact, up to a uniform constant independent of n we have

$$|\Delta_n| \approx x_{(n)}^{1+\gamma} \quad \text{for } any \ x_{(n)} \in \Delta_n. \tag{17}$$

Now consider the sequence $y_k = 1/k^{1/\gamma}$ and let $J_k = [y_{k+1}, y_k]$. Then, considering the function $g(x) = 1/x^{1/\gamma}$ we have $g'(x) \approx 1/\gamma x^{\frac{1}{\gamma}+1}$ and a straightforward application of the Mean Value Theorem gives

$$|J_k| = |y_k - y_{k+1}| = \left|\frac{1}{k^\gamma} - \frac{1}{(k+1)^\gamma}\right| = |g(k) - g(k+1)| \approx \frac{1}{k^{\frac{1}{\gamma}+1}} = \left(\frac{1}{k^{\frac{1}{\gamma}}}\right)^{1+\gamma} = y_k^{1+\gamma}$$

Similarly as above we have

$$y_k/y_{k+1} = ((k+1)/k)^{1+\gamma} \to 1$$

as $k \to \infty$, and therefore, up to a constant independent of k we have

$$|J_k| \approx y_{(k)}^{1+\gamma} \quad \text{for } any \ y_{(k)} \in J_k \tag{18}$$

Combining (17) and (18) we see that if $\Delta_n \cap J_k \neq \emptyset$ then $|\Delta_n| \approx |J_k|$. This means that there is a uniform bound on the number of intervals that can overlap each other which means that the sequences x_n, y_n have the same asymptotics and so $x_n \approx y_n = 1/n^{\frac{1}{\gamma}}$ and in particular $|\Delta_n| \approx x_n^{1+\gamma} = 1/n^{\frac{1}{\gamma}+1}$. □

9.2. **Distortion estimates.**

Lemma 9.3. *There exists a constant $\mathcal{D} > 0$ such that for all $n \geq 1$ and all $x, y \in \Delta_n$*

$$\left|\log \frac{Df^n(x)}{Df^n(y)}\right| \leq \mathcal{D}|f^n(x) - f^n(y)|.$$

Proof. We start with the standard inequality

$$\left|\log \frac{Df^k(x)}{Df^k(y)}\right| \leq \sum_{i=0}^{k-1} |\log Df(f^i(x)) - \log Df(f^i(y))| \leq \sum_{i=0}^{k-1} \frac{D^2 f(\xi_i)}{Df(\xi_i)} |f^i(x) - f^i(y)|$$

for some $\xi_i \in (f^i(x), f^i(y))$, where we have used here the Mean Value Theorem and the fact that $D(\log Df) = D^2 f/Df$. Since $x, y \in \Delta_n$ then $x_i, y_i \in \Delta_{n-i}$ and so, by the previous Lemma we have

$$|f^i(x) - f^i(y)| \leq |\Delta_{n-i}| \leq 1/(n-i)^{\frac{1}{\gamma}+1}.$$

Moreover, by the definition of f we have

$$Df(x) = 1 + (1+\gamma)x^\gamma \quad \text{and} \quad D^2 f(x) = \gamma(1+\gamma)x^{\gamma-1}$$

and therefore, from the fact that $\xi_i \in \Delta_{n-i}$ we have

$$\xi_i \approx \frac{1}{(n-i)^{\frac{1}{\gamma}}}, \quad Df(\xi_i) \approx 1 + \frac{1}{n-i}, \quad D^2 f(\xi_i) \approx \frac{1}{(n-i)^{1-\frac{1}{\gamma}}}$$

we get

$$\left|\log \frac{Df^k(x)}{Df^k(y)}\right| \leq \sum_{i=0}^{k-1} \frac{D^2 f(\xi_i)}{Df(\xi_i)} |f^i(x) - f^i(y)| \lesssim \sum_{i=1}^{k-1} \frac{(n-i)^{\frac{1}{\gamma}-1}}{(n-i)^{\frac{1}{\gamma}+1}} \leq \sum_{i=1}^{\infty} \frac{1}{i^2}.$$

This gives a uniform bound for the distortion but not yet in terms of the distance as required in the Lemma. For this we now take advantage of the distortion bound just obtained to get

$$\frac{|x-y|}{|\Delta_n|} \approx \frac{|f^i(x) - f^i(y)|}{|\Delta_{n-i}|} \approx \frac{|f^n(x) - f^n(y)|}{|\Delta|}$$

to get in particular

$$|f^i(x) - f^i(y)| \approx |\Delta_{n-i}||f^n(x) - f^n(y)|.$$

Repeating the calculation above with this new estimate we get

$$\left|\log \frac{Df^k(x)}{Df^k(y)}\right| \leq \sum_{i=0}^{k-1} \frac{D^2 f(\xi_i)}{Df(\xi_i)} |\Delta_{n-i}||f^n(x) - f^n(y)| \lesssim |f^n(x) - f^n(y)|$$

□

Lemmas 9.1 and 9.3 imply that the map $F : \Delta \to \Delta$ has a unique ergodic absolutely continuous invariant probability measure $\hat{\mu}$. To get the corresponding measure for μ it only remains to show that $\int \tau d\hat{\mu} < \infty$. We also know however that the density $d\hat{\mu}/dm$ of $\hat{\mu}$ with respect to Lebesgue measure μ is Lipschitz and in particular bounded, and therefore it is sufficient to show that $\int \tau dm < \infty$.

Lemma 9.4. *For $\gamma \in (0,1)$, the induced map F has integrable inducing times. Moreover, for every $n \geq 1$ we have*

$$m(\{x : \tau(x) \geq n\}) = \sum_{j=n}^{\infty} m(\Delta_n) \lesssim \frac{1}{n^{\frac{1}{\gamma}}}$$

Proof. From the estimates obtained above we have that $|\Delta_n| \approx n^{-(\frac{1}{\gamma}+1)}$. Therefore

$$\int \tau dx \lesssim \sum_n n|\Delta_n| \approx \sum_n \frac{1}{n^{\frac{1}{\gamma}}}.$$

The sum on the right converges whenever $\gamma \in (0,1)$ and this gives the integrability. The estimate for the tail follows by standard methods such as the following

$$\sum_{j=n}^{\infty} |\Delta_n| \lesssim \sum_{j=n}^{\infty} \frac{1}{n^{\frac{1}{\gamma}+1}} \leq \int_{n-1}^{\infty} \frac{1}{x^{\frac{1}{\gamma}+1}} dx \approx \left[\frac{1}{x^{\frac{1}{\gamma}}}\right]_{n-1}^{\infty} = \frac{1}{(n-1)^{\frac{1}{\gamma}}} \approx \frac{1}{n^{\frac{1}{\gamma}}}.$$

□

Inducing is a very powerful method for constructing invariant ergodic probability measures which are absolutely continuous with respect to Lebesgue measure. In this application above we constructed a uniformly expanding full branch induced map with bounded distortion since we have proved that such maps have ergodic acip's.

Theorem 13 (Rychlik)**.** *Let $F : [0,1] \to [0,1]$ be a piecewise C^1 on a countable partition $\mathcal{P}$ and suppose that*

$$\sum_{\omega \in \mathcal{P}} var\left(\frac{\mathbb{1}_\omega}{F'}\right) < \infty.$$

Then F admits a unique ergodic absolutely continuous invariant measure μ with density of bounded variation.

It may be easier in some cases to induce to a Rychlik map than to a full branch map. On the other hand, even though it may be difficult to construct explicitly in certain situations, it turns out that the existence of a uniformly expanding full branch induced map with bounded distortion is quite general and indeed almost a necessary condition for the existence of ergodic invariant absolutely continuous invariant measures.

Definition 12. Let M be a Riemannian manifold. We say that an invariant probability measure μ is *expanding* if all its Lyapunov exponents are positive, i.e. for μ-almost every x and every $v \in T_xM \setminus \{0\}$,

$$\lambda(x,v) := \limsup_{n\to\infty} \frac{1}{n} \log \|Df^n(x)v\| > 0. \tag{19}$$

We say that μ is *regularly expanding* if it is expanding and in addition we have

$$\log \|Df^{-1}\| \in L^1(\mu). \tag{20}$$

Theorem 14 (Alves, Dias, Luzzatto, 2010)**.** *Let $f : M \to M$ be a C^2 map with a non-degenerate critical set. Then f admits a uniformly expanding full branch induced map with bounded distortion if and only if it admits an ergodic regularly expanding acip.*

10. Mixing and Decay of Correlations

Ergodicity is only the beginning of the story in terms of the statistical properties of dynamical systems. The asymptotic statistical distribution given by Birkhoff's ergodic theorem provides some information about the dynamic, but does not distinguish, for example, between such quite different dynamical systems as an irrational circle rotation and a piecewise affine uniformly expanding map, both of which admit Lebesgue measure as an invariant ergodic measure.

Definition 13. For measurable functions $\varphi, \psi : M \to \mathbb{R}$ we define the *correlation function*

$$\mathcal{C}_n(\varphi,\psi) = \left| \int \psi(\varphi \circ f^n) d\mu - \int \psi d\mu \int \varphi d\mu \right|$$

We say that the correlation function $\mathcal{C}_n(\varphi,\psi)$ *decays* if $\mathcal{C}_n(\varphi,\psi) \to 0$ as $n \to \infty$. In the special case in which φ, ψ are characteristic functions of sets A, B we can write

$$\begin{aligned}
\mathcal{C}_n(\mathbb{1}_A, \mathbb{1}_B) &= \left| \int \mathbb{1}_A(\mathbb{1}_B \circ f^n) d\mu - \int \mathbb{1}_A d\mu \int \mathbb{1}_B d\mu \right| \\
&= \left| \int \mathbb{1}_{A\cap f^{-n}(B)} d\mu - \int \mathbb{1}_A d\mu \int \mathbb{1}_B d\mu \right| \\
&= |\mu(A \cap f^{-n}(B)) - \mu(A)\mu(B)|.
\end{aligned}$$

Definition 14. We say that an invariant probability measure μ is *mixing* if

$$\mathcal{C}_n(\mathbb{1}_A, \mathbb{1}_B) = |\mu(A \cap f^{-n}(B)) - \mu(A)\mu(B)| \to 0$$

as $n \to \infty$, for all measurable sets $A, B \subseteq M$.

Mixing has some natural geometrical and probabilistic interpretations. Indeed, dividing through by $\mu(B)$ we get

$$\left| \frac{\mu(A \cap f^{-n}(B))}{\mu(B)} - \mu(A) \right| \to 0$$

as $n \to \infty$, for all measurable sets $A, B \subseteq M$, with $\mu(B) \neq 0$. Geometrically, one can think of $f^{-n}(B)$ as a "redistribution of mass" and the mixing condition says that for large n the proportion of $f^{-n}(B)$ which intersects A is just proportional to the measure of A. In other words $f^{-n}(B)$ is spreading itself uniformly with respect to the measure μ. A more probabilistic point of view is to think of $\mu(A \cap f^{-n}(B))/\mu(B)$ as the conditional probability of having $x \in A$ given that $f^n(x) \in B$. The mixing condition then says that this probability converges to the probability of A, i.e., asymptotically, there is no causal relation between the two events. This is why we say that a mixing system exhibits *stochastic-like* or *random-like* behaviour.

Example 7. It is easy to verify that an irrational circle rotation is not mixing. On the other hand, the map $2x$ mod 1 is mixing, though this is not completely trivial to verify.

A natural question is on the *rate of decay* of the correlations function. In general this will depend on the functions φ, ψ.

Definition 15. Given classes $\mathcal{B}_1, \mathcal{B}_2$ of functions and a sequence $\{\gamma_n\}$ of positive numbers with $\gamma_n \to 0$ as $n \to \infty$ we say that the correlation function $C_n(\varphi, \psi)$ decays for functions $\varphi \in \mathcal{B}_1, \psi \ \in \mathcal{B}_2$ at the rate given by the sequence $\{\gamma_n\}$ if, for any $\varphi, \psi \in \mathcal{B}$ there exists a constant $C = C(\varphi, \psi) > 0$ such that

$$\mathcal{C}_n(\varphi, \psi) \leq C\gamma_n$$

for all $n \geq 1$.

For example, if $\gamma_n = e^{\gamma n}$ we say that the correlation decays exponentially, if $\gamma_n = n^-\gamma$ we say that the correlation decays polhynomially. The key point here is that the *rate*, i.e. the sequence $\{\gamma_n\}$ is not allowed to depend on the functions but only on the function class. Thus the rate of decays becomes in some sense an intrinsic property of the system. It is not always possible to obtain decay at any specific rate if the class of observables is too large. For example, if we choose $\mathcal{B}_1 = \mathcal{B}_2 = \mathcal{L}^1$ or any other space of functions that includes characteristic functions, then given any rate, it is possible to find subsets A, B such that the correlation function $\mathcal{C}_n(\mathbb{1}_A, \mathbb{1}_B)$ of the corresponding characteristic functions decays at a slower rate. It is however possible to prove that many piecewise uniformly expanding one-dimensional maps exhibit exponential decay of correlations for relatively large spaces of functions such as Hölder continuous functions or functions of bounded variation. About a decade ago, L.-S. Young, showed that the method of inducing can be used also to study the decay of correlations of maps which may not be uniformly expanding but which admit good induced maps. More precisely, suppose that Suppose that $f : M \to M$ admits an induced uniformly expanding full branch map $F = f^\tau : \Delta \to \Delta$ satisfying the bounded distortion property. We have see above that F admits a unique ergodic *acip* $\hat{\mu}$ with bounded density. If the return times are Lebesgue integrable $\int \tau dm < \infty$ then there exists an ergodic *acip* μ for f. The rate of decay of correlation of μ is captured by the rate of decay of the tail of the return time function. More precisely, let

$$\hat{\Delta}_n := \{x \in \Delta : \tau(x) \geq n\}.$$

Theorem 15. *The rate of decay of correlation with respect to μ for Hölder continuous functions is determined by the rate of decay of $|\hat{\Delta}_n|$: if $|\hat{\Delta}_n| \to 0$ exponentially, then the rate of decay is exponential, if $|\hat{\Delta}_n| \to 0$ polynomially, then the rate of decay is polynomial.*

These general results indicate that the rate of decay of correlations is linked to what is in effect the *geometrical structure* of f as reflected in the tail of the return times for the induced map F. From a technical point of view they shift the problem of the statistical properties of f to the problem of the geometrical structure of f and thus to the (still highly non-trivial) problem of showing that f admits an induced Markov map and of estimating the tail of the return times of this map. The construction of an induced map in certain examples is relatively straightforward and essentially canonical but the most interesting constructions require statistical arguments to even show that such a map exists and to estimate the tail of the return times. In these cases the construction is not canonical and it is usually not completely clear to what extent the estimates might depend on the construction.

11. Abundance of maps with absolutely continuous invariant probability measures

We have seen that, at least in theory, the method of inducing is a very powerful method for constructing invariant measures and studying their statistical properties. The question of course is whether the method is really applicable and more generally if there are many maps with acip's. Let $C^2(I)$ denote the family of C^2 maps of the interval. We say that $c \in I$ is a *critical point* if $f'(c) = 0$. In principle, critical points constitute a main obstruction to the construction and estimates we have carried out above, since they provide the biggest possible contraction. If a critical point is periodic of period k, then c is a fixed point for f^k and $(f^k)'(c) = 0$ and so c is an attracting periodic orbit. On the other hand we have already seen that maps with critical points can have acid's as in the case of the map $f(x) = x_2 - 2$ which is smoothly conjugate to a piecewise affine "tent map". This map belongs to the very well studied quadratic family

$$f_a(x) = x^2 + a.$$

It turns out that any interesting dynamics in this family only happens for a bounded interval

$$\Omega = [-2, a^*]$$

of parameter values. For this parameter interval we define

$$\Omega^+ := \{a \in \Omega : f_a \text{ admits an ergodic acip } \mu\}$$

and

$$\Omega^- := \{a \in \Omega : f_a \text{ admits an attracting periodic orbit}\}.$$

Over the last 20 years or so, there have been some quite remarkable results on the structure of these sets. First of all, if $a \in \Omega^+$ then μ is the unique physical measure and $m(\mathcal{B}_\mu) = 1$, i.e. the time average of *Lebesgue* almost every x for a function φ converge to $\int \varphi d\mu$. On the other hand, if $a \in \Omega^-$ then the Dirac measure $\delta_{\mathcal{O}^+(p)}$ on the attracting periodic orbit is the unique physical measure and $m(\mathcal{B}_{\delta_{\mathcal{O}^+(p)}}) = 1$, Lebesgue almost every x converges to the orbit of p. Thus in particular we have

$$\Omega^+ \cap \Omega^- = \emptyset.$$

Moreover, we also have the following results:

Theorem 16. (1) *Lebesgue almost every $a \in \Omega$ belongs to either Ω^+ or Ω^-;*
(2) *Ω^- is open and dense in Ω;*
(3) *$m(\Omega^+) > 0$.*

The last of these statements is actually the one that was proved first, by Jakabson in 1981. He used precisely the method of inducing to show that there are a positive Lebesgue measure set of parameters for which there exists a full branch induced map with exponential tails (and this exponential decay of correlations).

Appendix A. Additional Remarks

A.1. **Singular physical measures and hyperbolicity.** A much more sophisticated version of this argument can be applied in the case that $\Lambda \subset M$ is a certain kind of *chaotic attractor* such that $m(\Lambda) = 0$ but such that the topological basin of attraction of Λ has positive measure. Then any ergodic invariant probability measure is necessarily supported on Λ and therefore is singular with respect to Lebesgue and so once again Birkhoff's ergodic theorem cannot be used to prove that μ is a physical measure. The question therefore is whether the points in the basin which are topologically attracted to Λ are also "probabilistically attracted" to μ, in the sense that, as they get closer and closer to Λ, their time averages converge to those of points in Λ. This can be shown to be the case if Λ satisfies some *hyperbolicity* conditions, which imply the existence of a *foliation* of *stable manifolds*. This implies that points in the topological basin are not justgenerically attracted to the attractor Λ but are actually attracted to the orbit of a specific point on the attractor, and thus their asymptotic time averages will be the same as those of the point on the attractor which they are attracted to. This strategy has actually been implemented successfully for the most famous chaotic attractor of all, the Lorenz attarctor, given by a relatively simple system of ODE's The Lorenz equations were introduced by the metereologist E. Lorenz in 1963, as an extremely simplified model of the Navier-Stokes equations for fluid flow.

$$\begin{cases} \dot{x}_1 & = 10(x_2 - x_1) \\ \dot{x}_2 & = 28x_1 - x_2 - x_1x_3 \\ \dot{x}_3 & = x_1x_2 - 8x_3/3. \end{cases}$$

This is a very good example of a relatively simple ODE which is quite intractable from many angles. It does not admit any explicit analytic solutions; the topology is extremely complicated with infinitely many periodic solutions which are knotted in many different ways (there are studies from the point of view of knot theory of the structure of the periodic solutions in the Lorenz equations); numerical integration has very limited use since nearby solutions diverge very quickly. Lorenz noticed that the dynamics of the Lorenz equations is chaotic in that it exhibits *sensitive dependence on initial conditions*, nearby solutions diverge exponentially fast and seem to have completely independent futures. This makes it extremely difficult to follow a specific solution with any real accuracy for any reasonably long time. However we have the following

Theorem 17 (1963-2000, combinations of several results by different people)**.** *The Lorenz equations admit a physical measure μ whose basin $\mathfrak{B}_\mu$ has full measure.*

This remark implies in particular that the asymptotic probability distribution is determined by μ and is therefore *independent* of x for every $x \in \mathfrak{B}_\mu$. This contrast with the sensitive dependence on initial conditions but does not contradict it. The sensistive dependence on initial conditions says that two initial conditions diverge from each other and therefore are hard to follow, but this says that nevertheless, thei asymptotic distribution is the same.

Appendix B. Lack of convergence of probabilistic limits

The probabilistic ω-limit set of course related to the topological ω-limit set, but there are some important and non-trivial subtleties. Notice first of all that it tracks the amount of time that the orbit spends in different regions of the space, so that if the proportion of points of the orbit in a certain region of space is positive but tends to zero, then the limit measures will not give positive measure to that region. We give two examples.

Example 8. Suppose $f : M \to M$ is a continuous map, p is a fixed point, and $f^n(x) \to p$ as $n \to \infty$. Then it is easy to verify that $\omega(x) = \{p\}$ and $\omega_{prob}(x) = \{\delta_p\}$, which is exactly what we expect.

Example 9. On the other hand, consider a situation with a heteroclinic cycle connecting two fixed points, a very common situation both in continuous time and discrete time systems. Suppose that f has two hyperbolic fixed points p_A, p_B whose separatrices connect the two points defining a closed topological disk as in the following picture.

Suppose that area enclosed by the fixed points p_A, p_B and the separatrices contains a fixed point P which is repelling and that all trajectories spiral away from P and accumulate on the boundary of the disk. Under suitable conditions on the eigenvalues of the points p_A and p_B the situation is the following. Suppose we have a point inside the domain bounded by this cycle whose orbit spirals outwards towards the cycle. Then it is quite easy to verify that the topological ω-limit set is just the entire cycle. There are on the other hand various possibilities for the probabilistic ω-limit set, depending on the specific values of the eigenvalues of the linearizations at the fixed points. It is fairly easy to see that the orbit will spend much longer time near the fixed points, since the orbit is very slow near the fixed points and so spends a large amount of time there, but takes only a fixed bounded number of iterations to get from one fixed point to the other. However there are two crucially different possible situations:

(1) $\omega_{prob}(x) = \{\mu\}$ where μ is some convex combination $\eta\delta_{p_1} + (1-\eta)\delta_{p_2}$ of the Dirac-δ measures in the fixed points for some $\eta \in [0, 1]$.

(2) $\omega_{prob}(x) = \{\delta_{x_1}, \delta_{x_2}\}$ so that the sequence μ_n does not converge but has both Dirac-δ measures as limit points.

These two cases look quite similar but they are dramatically different. The first case says that the sequence μ_n does actually converge and that the orbit asymptotically spends a certain proportion of time near p_1 and a certain proportion near p_2. On the other hand, the second case says that the sequence μ_n *does not converge* and has both Dirac-δ measures as limits. This means, that there is some subsequence $n_j \to \infty$ such that $\mu_{n_j} \to \delta_{p_1}$ and another subsequence $n_k \to \infty$ such that $\mu_{n_k} \to \delta_{p_2}$. But this means that if we look at the finite piece of orbit up to time n_j, *most* points along the orbit, e.g. 99% of points, will be close to δ_{p_1}. On the other hand if we look at finite pieces of orbit up to some time n_k then *most* points along the orbit, e.g. 99% of points, will be close to δ_{p_2}. So the "probability" of being close to p_1 or to p_2 depends on the size of the piece of initial trajectory. It is relatively easy to show that this situation can indeed occur, by choosing the eigenvalues of the linearization at the fixed points appropriately. This situation means that *the averages do not converge* ! This would be like tossing a coin and having for example that if you toss is ten times you are more likely to get heads, if you toss it one hundered times, you are more likely to get tails, if you toss it one thousand times you are more likely to get heads, etc.

References

[1] Anatole Katok and Boris Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, vol. 54, Cambridge University Press, Cambridge, 1995. With a supplementary chapter by Katok and Leonardo Mendoza.

[2] Ricardo Mañé, *Ergodic theory and differentiable dynamics*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 8, Springer-Verlag, Berlin, 1987. Translated from the Portuguese by Silvio Levy.

[3] Karl Petersen, *Ergodic theory*, Cambridge Studies in Advanced Mathematics, vol. 2, Cambridge University Press, Cambridge, 1983.

[4] Ya. G. Sinaĭ, *Topics in ergodic theory*, Princeton Mathematical Series, vol. 44, Princeton University Press, Princeton, NJ, 1994.

[5] Peter Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York, 1982.

Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, Trieste, Italy

E-mail address: `luzzatto@ictp.it`

URL: `http://www.ictp.it/~luzzatto`