Introduction to Smooth Ergodic Theory

Stefano Luzzatto
Abdus Salam International Centre for Theoretical Physics,
luzzatto@ictp.it   http://www.ictp.it/~luzzatto

DRAFT: version April 10, 2017

Contents

1 Introduction

1.1 Dynamical Systems

Let M be a set. A transformation of the space M is simply a map $f : M \to M$. The transformation f is said to be *invertible* if f is a bijection and therefore its inverse f^{-1} can be defined so that $f \circ f^{-1} = f^{-1} \circ f$ =Identity. Any two transformations $f, g : M \to M$ can be applied in sequence to give the *composition* $g \circ f : M \to M$ defined by $g \circ f(x) = g(f(x))$.

Definition 1. An invertible Dynamical System on M is a *group* of transformations of the space M under composition. A non-invertible Dynamical System is a *semi-group* of transformations of the space M under composition.

Example 1. The simplest abstract example of a Dynamical System is given by an arbitrary map $f : M \to M$. We can then define the family of maps $\{f^n\}_{n\in\mathbb{N}}$ where we define f^0 =Identity and f^n is the n'th *iterate* of the map f defined inductively by the relation $f^n = f \circ f^{n-1}$ Then it is easy to see that this family forms a semi-group of transformations of M since it is closed under composition, $f^{n+m} = f^n \circ f^m$, and f^0 is the identity or neutral element of the semi-group. If f is invertible, then the inverse f^{-1} is defined and so are its iterates $f^{-n} = f^{-1} \circ f^{-(n-1)}$ and therefore we can define the family $\{f^n\}_{n\in\mathbb{Z}}$ which is easily seen to be a group under composition. These dynamical systems are sometimes referred to as dynamical systems with *discrete time* because they are parametrised by the sets $\mathbb{N}$ and $\mathbb{Z}$ which are discrete.

Remark 1. Other semigroups and groups of transformations arise naturally in different settings. For example the *flow* defined by an Ordinary Differential Equation is a family $\{\varphi^t\}_{t\in\mathbb{R}}$ which forms a group under composition and is therefore an example of a family of transformations in *continuous time*. Even more general families can be studied for example with *complex time*, parametrised by $\mathbb{C}$. or other parametrised by more abstract groups. In these notes we will focus on the case of dynamical systems in discrete time since there are some very explicit examples and they already contain an extremely rich variety of structures.

In the study of Dynamical Systems we often think of M as a “phase space” of possible states of the system, and the dynamical system as the ”evolution” of the system in time. The most basic and fundamental notion in the theory of dynamical system is that of the *orbit* or *trajectory* of a point or *initial condition* x_0 under the action of the system.

Definition 2. For any $x \in M$ we define the *forward orbit* or *forward trajectory* of x by

$$\mathcal{O}^+(x) := \{f^n(x)\}_{n\in\mathbb{N}}$$

If the dynamical system is invertible we define the *full orbit* or *full trajectory* by

$$\mathcal{O}(x) := \{f^n(x)\}_{n\in\mathbb{Z}}.$$

The notion of an orbit formalises the evolution of an "initial condition" $x \in M$. Notice that the orbit of x is just a sequence of points where $x_0 = x$, $x_1 = f(x), x_2 := f^2(x) = f(f(x)) = f \circ f(x)$ 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.

Remark 2. If f is not invertible, the inverse f^{-1} is not well defined as a map, though for any $n \geq 1$ we can still define the sets

$$f^{-n}(x) := \{y \in M : f^n(y) = x\}$$

and call this set the *$n'th$ preimage* of the point x.

1.2 Topological structures

One of the main goals of the theory of Dynamical Systems can be formulated as the description and classification of the structures associated to dynamical systems and in particular the study of the orbits of dynamical systems. The very simplest, and perhaps one of the most important kinds of orbits is the following.

Definition 3. Let $f : M \to M$ be a map. $x \in M$ is a *fixed point* for f if

$$f(x) = x.$$

If x is a fixed point for f, then it is easy to see that it is a fixed point for every forward iterate of f and therefore a fixed point for the dynamical system generated by f. In this case then the forward orbit reduces to the point x, i.e. $\mathcal{O}^+(x) = \{x\}$.

Definition 4. $x \in M$ is a *periodic point* for f if there exists $k > 0$ such that

$$f^k(x) = x.$$

The minimal $k > 0$ for which the above condition holds is called the *minimal period* of x.

If x is a periodic point with minimal period k then the forward orbit of the point x is just the finite set $\mathcal{O}^+(x) = \{x, f(x), ..., f^{k-1}(x)\}$.

Remark 3. Notice that a fixed point is just a special case of a periodic orbit with $k = 1$ and that any periodic orbit with period k is also a periodic orbit with period any multiple of k.

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 concepts. If the orbit of x is not periodic, then $\mathcal{O}^+(x)$ is a countable set and we need some additional structure on M to describe it. If M is a topological space then we can talk about the *accumulation points* of the orbit of x which describe in some sense the "asymptotic" behaviour of the orbit of x.

Definition 5. The *omega-limit* set of a point $x \in M$ is

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

If f is invertible, the *alpha-limit* set of a point $x \in M$ is

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

The case in which $\omega(x)$ is the whole space is a special and quite important situation. Thus we make the following definitions.

Definition 6. Let M be a topological space and $f : M \to M$ a map. We say that the orbit of x is *dense* in M if $\omega(x) = M$. We say that f is *transitive* if there exists a point $x \in M$ with a dense orbit. We say that f is *minimal* if every point $x \in M$ has a dense orbit.

1.3 Fundamental examples

We give here three basic examples which we will study in more detail during this course.

1.3.1 Contracting maps

Let M be a metric space with metric $d(\cdot, \cdot)$.

Definition 7. A map $f : M \to M$ is a *contraction* if there exists $\lambda \in [0, 1)$ such that

$$d(f(x)), f(y)) \leq \lambda d(x, y)$$

for all $x, y \in M$.

For contractions we have the following well known result which we formulate here using the notions introduced above.

Proposition 1.1 (Contraction Mapping Theorem)**.** *Let M be a complete metric space and $f : M \to M$ a contraction. Then there exists a unique fixed point $p \in M$ and $\omega(x) = \{p\}$ for all $x \in M$.*

Sketch of proof. Let $x \in M$ be an arbitrary point.

1. The forward orbit $\mathcal{O}^+(x) = \{f^n(x)\}_{n \in \mathbb{N}}$ of x forms a Cauchy sequence.

2. Any accumulation point of $\mathcal{O}^+(x)$ is a fixed point.

3. A contraction can have at most one fixed point.

Therefore every orbit converges to a fixed point and since this point is unique all orbits converge to the same fixed point. □

This result therefore completely describes the asymptotic behaviour of all initial conditions from a topological point of view. The fixed point p has the property that it *attracts* all points in the space. In more general situations we may have the existence of fixed points which do not attract every point but only some points in the space. We formalise this notion as follows.

Definition 8. Let p be a fixed point. The *(topological) basin of attraction of* p is $\mathcal{B}_p := \{x : \omega(x) = \{p\}\}$. The point p is a locally attracting fixed point if $\mathcal{B}_p$ contains a neighbourhood of p and a globally attracting fixed point if $\mathcal{B}_p$ is the whole space.

Remark 4. The notion of an attracting fixed point can be generalized to periodic points. If $p = p_0$ is a periodic point of prime period n with periodic orbit $P = \{p_0, p_1, ..., p_{n-1}\}$ we can define the (topological) basin of attraction as $\mathcal{B}_P = \{x : \omega(x) = P\}$ and say that the orbit is attracting if $\mathcal{B}_P$ contains a neighbourhood $\mathcal{U}$ of P. Notice that in this case, the basin will contain neighbourhoods $\mathcal{U}_i$ of each point $p_i \in P$ made of points whose forward orbits converge to the forward orbit of p_i.

1.3.2 Circle rotations

Let $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$ denote the unit circle. For any $\alpha \in \mathbb{R}$ we define the map $f_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ by

$$f_\alpha(x) = x + \alpha.$$

Then the iterates of f_α have the form

$$f_\alpha^n(x) = x + n\alpha$$

and the dynamics of f_α depends very much on the parameter α.

Proposition 1.2. *The dynamics of f_α exhibits the following dichotomy.*

1. *α is rational if and only if every x is periodic;*
2. *α is irrational if and only if every x is dense.*

Sketch of proof.

1. If $\alpha = p/q$ is rational we have $f_\alpha^q(x) = x + q\alpha = x + p = x \mod 1$ and so every point is periodic (of the same period q). Conversely, if x is periodic, there exists some q such that $f^n(x) = x + q\alpha = x \mod 1$ and so $q\alpha = 0 \mod 1$ and so α must be of the form p/q.

2. If there exists even one point x with dense orbit, then the orbit is not period and so α cannot be rational and so must be irrational. Conversely, suppose α is irrational and let x be an arbitrary point. To show that the orbit of x is dense, carry out the following steps.

   (a) Let $\epsilon > 0$ and cover the circle $\mathbb{S}^1$ with a finite number of arcs of length $\leq \epsilon$. Since α is irrational, the orbit of x cannot be periodic and so is infinite. Therefore there must be at least one such arc which contains at least two point $f_\alpha^m(x), f_\alpha^n(x)$ of the orbit of x.

(b) The map f_α^{n-m} is a circle rotation f_δ for some $\delta \leq \epsilon$.

(c) The orbit of x under f_δ is therefore ϵ-dense.

(d) Every point on the orbit of x under f_δ is also on the orbit of x under f_α and therefore the orbit of x under f_α is also ϵ dense.

(e) ϵ is arbitrary and so the orbit of x is dense.

□

1.3.3 Expanding maps

Let $\kappa \in \mathbb{N}, \kappa \geq 2$ and define $f_\kappa : \mathbb{S}^1 \to \mathbb{S}^1$ by

$$f_\kappa(x) = \kappa x.$$

For simplicity we consider the case $\kappa = 10$ and let $f = f_{10}$ because the dynamics can then be studied very explicitly using the decimal representation of real numbers. The other cases all have very similar dynamics as can be seen by using the base κ representation of real numbers. Several properties of these maps are established in the exercises.

1.4 Probabilistic structures

The topological description of the dynamics described above provides a lot of useful and interesting information but in some cases also misses certain key features of the systems under consideration. More specifically, except for the special case of fixed or periodic orbits, it does not contain any information about the *frequency* with which a given orbit visits specific regions of the space. We introduce here some concepts and terminology to formalize this notion.

1.4.1 Probability measures

If M is a metric space then there is also a well defined Borel sigma-algebra and we let

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

Example 2. The simplest example of a Borel measure is the Dirac-δ measure δ_x on a point $x \in M$ defined by

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

for any Borel measurable set $A \subseteq M$. In particular $\mathcal{M} \neq \emptyset$ since it contains for example all Dirac-δ measures. If M has some additional structure, such as that of a Riemannian manifold, then it also contains the normalised *volume* which we generally refer to as *Lebesgue measure*.

1.4.2 Time averages

If M is a metric space and $f : M \to M$ is a map, then for any $x \in M$ we can define the sequence of probability measures

$$\mu_n(x) = \frac{1}{n}\sum_{i=0}^{n-1} \delta_{f^i(x)}. \tag{1}$$

The measure $\mu_n \in \mathcal{M}$ is then just a uniform distribution of mass on the first n points along the orbit. A natural question is whether this sequence converges and to what it converges and what the is the dynamical meaning of this convergence. To study this question, recall that by definition of the weak-star topology, $\mu_n \to \mu$ if and only if $\int \varphi d\mu_n \to \int \varphi d\mu$ for all $\varphi \in C^0(M, \mathbb{R})$. In the particular case in which the sequence μ_n is given by the form above, we have

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

Thus the convergence of the probability measure μ_n to a probability measure μ is equivalent to the convergence of the terms

$$\frac{1}{n}\sum_{i=0}^{n-1} \int \varphi \circ f^i(x) \tag{2}$$

to the average $\int \varphi d\mu$ when φ is a continuous function. The terms (2) are sometimes called the *time averages* of the function φ along the orbit of the point x and the integral $\int \varphi d\mu$ is sometimes called the *space average* of the function φ with respect to μ and so the convergence of one to the other is sometimes referred to as the *time averages converging to the space average*.

The behaviour of time averages constitutes the main object of these notes and one of the main objectives is to give conditions which guarantee their convergence.

1.4.3 Physical measures

If the time averages $\mu_n(x)$ along the orbit of a particular point x converge to some measure μ, it is natural to ask whether there exists any other point y, not belonging to the orbit of x, whose time averages also converge to the same measure μ. Thus we define the "basin of attraction" of a probability measure.

Definition 9. For $\mu \in \mathcal{M}$ let

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

It is easy to check that distinct measures must have disjoint basins (Exercise 13). Moreover, any point for which the time averages converge clearly belongs to the basin of some measure, thus one approach to the problem of studying time averages is to consider measures in $\mathcal{M}$ and their basins. In general the basin of most measures in $\mathcal{M}$ will be empty. In some cases, the basin of a measure may be just a single point (Exercise 4) and in other cases it may be the whole space (Exercise 3). It is natural to be interested in measures with "large" basins since these measures describe the asymptotic distribution of orbits for a large set of points.

Definition 10. A probability measure $\mu \in \mathcal{M}$ is called a **physical measure** if

$$Leb(\mathfrak{B}_\mu) > 0.$$

We can then ask the following questions.

Which dynamical systems admit a physical measure?

and

How many physical measures do they have?

It turns out that these are extremely challenging problems that have not yet been solved in general. There are examples of systems which do not admit any physical measures (Exercise 4) and examples of systems which admit countably many physical measures. However it is generally believed that these are somehow "exceptional" cases.

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.

1.5 Exercises

1.5.1 Fixed and periodic points

Exercise 1. Let f be invertible and suppose that x is a fixed point. Show that x is a fixed point for f^{-1} and conclude that $\mathcal{O}(x) = \{x\}$. Show that if f is not invertible then $x \in f^{-1}(x)$ but $f^{-1}(x)$ may also contain a point $y \neq x$.

Exercise 2. Show that the forward orbit $\mathcal{O}^+(x)$ of a point x is a finite set if and only if x is a periodic or pre-periodic orbit.

1.5.2 Omega limits

Exercise 3. Show that if x is a fixed point, then $\omega(x) = \mathcal{O}^+(x) = \{x\}$ and, more generally, if x is a periodic point of period k then $\omega(x) = \mathcal{O}^+(x)$.

Exercise 4. If f is continuous and $\omega(x) = \{p\}$ then p is a fixed point.

Exercise 5. Show that if M is compact then $\omega(x) \neq \emptyset$ for all $x \in M$. Moreover $\omega(x)$ is *compact* and *forward invariant* ($y \in \omega(x)$ implies $f(y) \in \omega(x)$).

1.5.3 PIecewise expanding maps

Exercise 6. Show that f has an periodic orbits of any given period. Show that f is transitive.

Exercise 7. Consider the map $f(x) = 10x \bmod 1$. Characterize all fixed, periodic and pre-periodic points of f in terms of their decimal representation. Show in particular that periodic points are dense in $[0, 1]$. Show that f is transitive, i.e. it has dense orbit. Show that it has an infinite number of distinct dense orbits. Show that it has orbits which are neither (pre-)periodic nor dense.

Exercise 8. Consider the questions in the previous exercise for the map $f(x) = 2x \bmod 1$, but this time using binary instead of decimal representation.

Exercise 9. Further generalize the previous two questions establishing similar properties for the map $f(x) = \kappa x \bmod 1$ for an arbitrarry positive integer $\kappa \geq 2$, using the representation of real numbers in base κ.

1.5.4 Convergence of time averages

Exercise 10. If x is a periodic point, then $\mu_n(x)$ converges to the uniform distribution of Dirac delta measures on the points of the periodic orbit. In particular if x is a fixed point, then $\mu_n(x) \to \delta_x$.

Exercise 11. If $\omega(x) = \{p\}$ then $\mu_n(x) \to \delta_p$. In particular, if $f : M \to M$ is a contraction mapping on a complete metric space, them $\mu_n(x) \to \delta_p$ for every $x \in M$.

Exercise 12. Let $f(x) = 10x \mod 1$.

1. Find a point x such that $\omega(x) = [0, 1]$ and $\mu_n(x) \to \delta_0$.

2. Find a point x such that $\omega(x) = [0, 1]$ and $\mu_n(x) \to (\delta_0 + \delta_{1/3})/2$.

3. Find a point x such that $\mu_n(x)$ does not converge.

Exercise 13. Show that if $\mu, \nu \in \mathcal{M}$ and $\mu \neq \nu$ then $\mathcal{B}_{mu} \cap \mathcal{B}_\nu = \emptyset$.

Example 3. If $f : M \to M$ is a contraction on a complete metric space with fixed point p, then δ_p is the only measure which has non-empty basin and $\mathcal{B}_{\delta_p} = M$.

Example 4. If $f(x) = x$ is the identity map, there are an uncountable number of measures with non-empty basin, namely every Dirac-delta measure δ_x has a non-empty basin with $\mathcal{B}_{\delta_x} = \{\delta_x\}$. In particular the identity map has no physical measure.

1.5.5 Harder questions

Exercise 14. * Show that the conclusions above in fact hold for any piecewise affine interval map with a finite number of branches.

2 Invariant Measures

The purpose of this section is to introduce a subset of the space $\mathcal{M}$ of probability measures on M consisting of measures which have particular properties which have some non trivial implications for the dynamics. For many definition or results we just need to the set M to be equipped with a sigma-algebra $\mathcal{B}$ and the map $f : M \to M$ to be measurable. However in some cases we will assume some additional structure and properties. We always let $\mathcal{M}$ denote the space of probability measure defined on the sigma-algebra $\mathcal{B}$ and assume that $\mu \in \mathcal{M}$.

Definition 11. μ is *invariant* if $\mu(f^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{B}$.

Exercise 15. 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 give a few simple examples of invariant measures and then prove a result on the existence of invariant measures and their dynamical implications.

Example 5. Let X be a measure space and $f : X \to X$ a measurable map. Suppose $f(p) = p$. Then the *Dirac* measure δ_p is invariant (Exercise 16).

Example 6. 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 $\rho_1 + \dots + \rho_n = 1$. Then, we can define a measure δ_P by letting

$$\delta_P(A) := \sum_{i:p_i \in A} \rho_i. \tag{3}$$

Then δ_P is invariant if and only if $\rho_i = 1/n$ for every $i = 1, \dots, n$ (Exercise 17).

Example 7. Let $f(x) = x$ be the identity map. Then every probability measure is invariant.

Example 8. Let $f(x) = x+\alpha \mod 1$ be a circle rotation. Then Lebesgue measure is invariant since a circle rotation is essentially a translation and Lebegue measure is translation-invariant. If α is rational then every point is periodic and thus f admits *also* other invariant measures given by the Dirac-delta measure defined in (3) above. This example shows that in general a map might admit many invariant measures.

Example 9. Let $I = [0, 1]$, $\kappa \geq 2$ an integer, and let $f(x) = \kappa x \mod 1$. Then Lebesgue measure is invariant (Exercise 18). Notice that f also has an infinite number of periodic orbits and thus also has an infinite number of invariant measures.

2.1 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 $\mu(A) = 0$. □

Remark 5. 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 $\mu(A''_n) \leq \mu(f^{-n}(f^n(A''_n))) = \mu(f^n(A''_n)) = 0$.

2.2 Convergence of time averages

We now come to one of the fundamental results of the theory which also constitutes the main motivation for the notion of an invariant measure.

Theorem 1 (Birkhoff, 1931)**.** *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.

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. Now 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. 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 $\mathcal{I}$. The proof of Theorem 1 follows easily from the following key technical statement.

Lemma 2.1. *Suppose $\psi_{\mathcal{I}} < 0$ (resp $\psi_{\mathcal{I}} > 0$). Then*

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \psi \circ f^k(x) \le 0 \qquad \left(\textit{resp.} \qquad \liminf_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \psi \circ f^k(x) \ge 0 \right)$$

for μ almost every x.

Remark 6. The definition of $\psi_{\mathcal{I}}$ is somewhat abstract and it is not easy to have an intuition for the difference between ψ and $\psi_{\mathcal{I}}$. The main property which we will use is that $\psi_{\mathcal{I}}$ must be constant along orbits, i.e. $\psi_{\mathcal{I}} \circ f = \psi_{\mathcal{I}}$. This follows by the fact that the full orbit of every point $\mathcal{O}(x) = \bigcup_{n\ge 0}\bigcup_{k\ge 0} f^{-n}(f^k(x))$ is an "indecomposable" element of the sigma-algebra $\mathcal{I}$ and thus $\psi_{\mathcal{I}}$ cannot take different values at distinct points of $\mathcal{O}(x)$.

Proof of Theorem 1 . For any $\epsilon > 0$ let $\psi^{\pm} := \varphi - \varphi_{\mathcal{I}} \pm \epsilon$. Since $(\varphi_{\mathcal{I}})_{\mathcal{I}} = \varphi_{\mathcal{I}}$ we have $\psi^{\pm}_{\mathcal{I}} = \pm\epsilon$. Thus, by Lemma 2.1 we have

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

and

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

which imply, respectively,

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

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

$$\varphi_f := \lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \varphi \circ f^k = \varphi_{\mathcal{I}} \tag{4}$$

for μ almost every x. $\square$

Proof of Lemma 2.1. Let

$$\Psi_n := \max_{k \le n} \left\{ \sum_{i=0}^{k-1} \psi \circ f^i \right\} \quad \text{and} \quad A := \{x : \Psi_n \to \infty\}$$

Then, for $x \notin A$, Ψ_n is bounded above and therefore

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \psi \circ f^k \le \limsup_{n\to\infty} \frac{\Psi_n}{n} \le 0$$

So it is sufficient to show that $\mu(A) = 0$. To see this, first compare the quantities

$$\Psi_{n+1} = \max_{1\le k\le n+1} \left\{ \sum_{i=0}^{k-1} \psi \circ f^i \right\} \quad \text{and} \quad \Psi_n \circ f = \max_{1\le k\le n} \left\{ \sum_{i=0}^{k-1} \psi \circ f^{i+1} \right\} = \max_{1\le k\le n} \left\{ \sum_{i=1}^{k} \psi \circ f^i \right\}$$

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 $\Psi_{n+1} \ge \Psi_n$ 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.

□

2.3 Existence of invariant measures

We now prove a general result which gives conditions to guarantee that at least some invariant measure exists. In fact we will take advantage of the topological structure on $\mathcal{M}$ given by the weak-star topology and describe, for a certain class of systems, the structure of the subset of invariant measure. Let

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

Then we have the following

Theorem 2 (Krylov-Boguliobov Theorem)**.** *Suppose $\mathcal{M}$ is a compact metric space and f is continuous. Then $\mathcal{M}_f$ is non-empty, convex*[1]*, compact.*

We start with a key definition and some related results.

Definition 12 (Push-forward of measures)**.** Let $f : \mathcal{M} \to \mathcal{M}$ be the map from the space of probability measures to itself, defined by

$$f_*\mu(A) := \mu(f^{-1}(A)). \tag{5}$$

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

It can be checked that this map is well defined (Exercise 20). It follows immediately from the definition that μ is invariant if and only if $f_*\mu = \mu$, i.e. if μ is a fixed point 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_*.

Lemma 2.2. *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. $\square$

Corollary 2.1. *$f_* : \mathcal{M} \to \mathcal{M}$ is continuous.*

[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 $\mu_t \in \mathcal{M}_f$.

Proof. Consider a sequence $\mu_n \to \mu$ in $\mathcal{M}$. Then, by Lemma 2.2, 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. $\square$

Corollary 2.2. *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 2.2. 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. $\square$

Proof of Theorem 2. 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 $n \geq 1$,

$$\mu_n = \frac{1}{n}\sum_{i=0}^{n-1} f_*^i \mu_0. \tag{6}$$

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 2.2 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 2.2, μ is invariant and so $\mu \in \mathcal{M}_f$. $\square$

2.4 Exercises

Exercise 16. Show that the Dirac-delta measure δ_p on a fixed point is invariant.

Exercise 17. Show that the measure 3 on a periodic orbit is invariant if and only if $\rho_i = 1/n$ for every $i = 1, \ldots, n$.

Exercise 18. Let $I = [0, 1]$, $\kappa \geq 2$ an integer, and let $f(x) = \kappa x \mod 1$. Show that Lebesgue measure is invariant.

Exercise 19. Let $I = [0, 1]$ be a piecewise affine full branch map with an arbitrary finite or countable number of branches. SAhow that Lebesgue measure is invariant.

Exercise 20. $f_*\mu$ is a probability measure and so the map f_* is well defined.

Exercise 21. 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.

3 Ergodic Measures

We now introduce the second fundamental definition.

Definition 13. μ is *ergodic* if, for all $A \in \mathcal{B}$, $f^{-1}(A) = A$ and $\mu(A) > 0$ implies $\mu(A) = 1$.

The intuitive meaning of this definition is that the dynamics is "indecomposable", at least as far as the measure μ is concerned. The condition $f^{-1}(A) = A$ is sometimes referred to by saying that the set $A \subseteq M$ is *fully invariant*. In non-invertible maps this is much stronger than assuming forward invariance (Exercise 22).

In the rest of this section we discuss some dynamical consequences of ergodicity. In later sections we address the problem of the existence of ergodic measures and the highly non-trivial and important problems of verifying ergodicity for specific measures.

3.1 Birkhoff's Ergodic Theorem

The definitions of ergodicity and invariance are independent of each other but they both come into their own when they are used together.

Theorem 3 (Birkhoff, 1931). *Let M be a measure space, $f : M \to M$ a measurable map, and μ an f-invariant ergodic probability measure. Then, for every $\varphi \in L^1(\mu)$ the limit*

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

for μ almost every x.

Remark 7. An immediate application of Birkhoff's Ergodic Theorem gives that for any ergodic invariant measure μ, and any Borel measurable set A, letting $\varphi = \mathbb{1}_A$ be the characteristic function of A, we have

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

This means that μ almost every point has the same asymptotic frequency of visits to the set A and this frequence is exactly the *probability* of A.

The proof uses the following simple result.

Lemma 3.1. *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 fully 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 $Y = \cap_{n\in\mathbb{Z}} X_{k_n,n}$ 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. $\square$

Proof of Theorem 3. We use the notation used in the proof of Theorem 1. By definition φ_f is invariant along orbits and so, by ergodicity of μ and Lemma 3.1, it follows that it is constant a.e. Moreover, by (4) in the proof of Theorem 1 we have $\varphi_f = \varphi_{\mathcal{I}}$ a.e and therefore $\varphi_f = \int \varphi_f d\mu = \int \varphi_{\mathcal{I}} d\mu_I = \int \varphi d\mu$. $\square$

3.2 Basins of attraction

Birkhoff's Ergodic Theorem allows us to give a first answer to the question of the existence of non-empty basins of attraction for probability measures.

Corollary 3.1. *If M is a compact Hausdorff space and μ an f-invariant ergodic probability measure, then*

$$\mu(\mathfrak{B}_\mu) = 1.$$

Notice that this does not follow immediately from the previous statements since the set of full measure for which the time averages converge to the space averages depends on the function.

Proof. Since M is compact and Hausdorff we can choose a countable dense subset $\{\varphi_m\}$ of continuous functions and let A be the set of full measure such that the time averages converge for every φ_m. Then, for any arbitrary continuous function φ, every $x \in A$, 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. $\square$

3.3 Ergodic decomposition

We complete this section with a discussion on the existence and structure of the set of ergodic measures. Since we will be mainly interested in measures that are both ergodic and invariant, we consider this set of measures. Let

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

Theorem 4. *$\mu \in \mathcal{E}_f$ if and only if μ is an extremal[2] point of $\mathcal{M}_f$.*

Corollary 4.1. *Let M be compact and $f : M \to M$ continuous. Then $\mathcal{E}_f \neq \emptyset$ and 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*

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

The Corollary follows from standard abstract functional analytic results. More precisely, the fact that $\mathcal{E}_f \neq \emptyset$ follows immediately from the Krein-Millman Theorem which says that every compact convex subset (i.e. $\mathcal{M}_f$) of a locally convex topological vector space (i.e. the set of all Borel measures on M) is the closed convex hull of its extreme elements and thus, in particular, the set $\mathcal{E}_f$ of extreme elements is non-empty. The decomposition follows from Choquet's Theorem which states exactly this decomposition result for general non-empty compact convex sets.

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:

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

[2] Recall that 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 $t = 1$.

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. $\square$

3.4 Exercises

Exercise 22. 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 23. Show that the Dirac-delta measure δ_p on a fixed point is ergodic.

Exercise 24. Show that the Dirac-delta measure δ_P on a periodic orbit is ergodic.

Example 10. Let f be the identity map. The only ergodic measures are the Dirac-delta measures.

Example 11. Let $f : [0,1] \to [0,1]$ given by

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

Show that Lebesgue measure is invariant but not ergodic.

4 Unique Ergodicity

In general a given map may have many invariant measures. However there are certain special, but important, examples of maps which have a unique invariant, and thus ergodic, measure. A trivial example is constituted by contraction maps in which every orbit converges to a unique fixed point p, and therefore δ_p is the unique ergodic invariant probability measure. However there are several other less trivial examples and in this section we discuss some aspects of these examples.

Definition 14. We say that a map $f : X \to X$ is *uniquely ergodic* if it admits a unique (ergodic) invariant probability measure.

4.1 Uniform convergence

Theorem 5. *Let $f : X \to X$ be a continuous map of a compact metric space. Then f is uniquely ergodic if and only if for every continuous function φ, the limit*

$$\varphi_f := \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi \circ f^j \tag{7}$$

exists for every $x \in X$ and is independent of x.

Proof of Theorem 5. Supppose first that for any continuous function φ the limit φ_f exists for every x and is independent of x. By Birkhoff's Ergodic Theorem, for every ergodic invariant probability measure μ we have $\varphi_f = \int \varphi d\mu$ for μ a.e. x. But then 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 $\mu_1 = \mu_2$. Thus f has only one ergodic invariant probability measure and so is uniquely ergodic.

Conversely, suppose that f is uniquely ergodic and μ is the unique ergodic invariant probability measure. By Birkhoff's Ergodic Theorem, $\varphi_f(x) = \int \varphi d\mu$ for μ-a.e. x. We need to show that that this actually holds for *every* x, i.e. that the sequence

$$\frac{1}{n}\sum_{j=0}^{n-1} \varphi \circ f^j \to \varphi_f \tag{8}$$

as continuous functions, and thus uniformly. Suppose by contradiction that (8) 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

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

Define a sequence of measures

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

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

$$\left| \int \varphi d\nu_k - \varphi_f \right| \geq \epsilon.$$

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

$$\left| \int \varphi d\nu - \varphi_f \right| \geq \epsilon. \tag{9}$$

Moreover, arguing as in the proof of the Krylov-Boguliobov Theorem 2 we get[3] that $\nu \in \mathcal{M}_f$. But, by unique ergodicity we must have $\nu = \mu$ and so $\int \varphi d\nu = \int \varphi d\mu = \varphi_f$ contradicting (9) and therefore (8) and this completing the proof. □

[3]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 $\nu \in \mathcal{M}_f$.

For future reference we remark that one direction of the Theorem above only needs to be verified for a dense subset of continuous functions since it implies the statement for all continuous functions.

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 such that for every $\varphi \in \Phi$ the limit φ_f exists for every x and is independent of x. Then the same 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 (8) holds for every continuous function. □

4.2 Circle rotations

Proposition 4.1. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be the circle rotation $f(x) = x + \alpha$ with α irrational. Then f is uniquely ergodic.*

Proof. For any $m \geq 1$, consider the functions

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

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 space 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

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

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

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

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. $\square$

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

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

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 $x \in \mathbb{S}^1$.

Proposition 4.2. *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 $\mathbb{S}^1$.*

Proof. 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. $\square$

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

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

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 8. 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 4.3. *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 4.2. *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 4.3. 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. $\square$

Lemma 4.3. *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

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

Therefore

$$\mathscr{D}(a_i) = d \iff \log_{10} d + j \leq \log_{10} a_1 \leq \log_{10}(d+1) + j$$

or

$$\mathscr{D}(a_i) = d \iff \log_{10} d \leq \log_{10} a_i \bmod 1 \leq \log_{10}(d+1).$$

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

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

$\square$

5 Full Branch Maps

Definition 16. 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)*.

5.1 Invariance and Ergodicity of Lebesgue measure

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 12. The simplest examples of full branch maps are the maps $f : [0,1] \to [0,1]$ defined by $f(x) = \kappa x \mod 1$ for some integer $\kappa \geq 1$. 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 25. 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

$$\frac{|A^c|}{|I|} = \frac{|f^n(\omega_n \cap A^c)|}{|f^n(\omega_n)|} = \frac{|\omega^{(n)} \cap A^c|}{|\omega^{(n)}|} \le \epsilon. \tag{10}$$

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

Remark 9. 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 (10). 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.

5.2 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 $k \ge 2$ 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 17. 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

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

as $n \to \infty$.

Exercise 26. 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 $j = 0, \ldots, k-1$. Then, the base k expansion $x = 0.x_1x_2x_3\ldots$ of the point x clearly satisfies

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

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. $\square$

5.3 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 5.2. *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

Definition 18. 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 27. 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 28. 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 5.2. *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 $\mu(h^{-1}(A)) = 1$. □

Example 13. Define the *Ulam-von Neumann* map $f : [-2, 2] \to [-2, 2]$ by

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

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

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

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 $\pi^{-1}\cos^{-1}(\cos 2\pi x) = -2(x-1) = -2x-2$.

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

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

Thus $\mu = h_*m$ is invariant and ergodic for f.

We now apply the method of defining measures via conjugacy to piecewise affine maps. 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 5.3. *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 $(x_0^{(p)}x_1^{(p)}x_2^{(p)}\ldots) \in \Sigma_2^+$ 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 29. Show that h_p is a conjugacy between f and f_p.

Exercise 30. 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 5.2 that μ_p is ergodic and invariant for f.

Exercise 31. Show that μ_p is non-atomic.

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

Lemma 5.4. *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.

$\square$

Remark 10. 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.

6 Distortion

6.1 The Gauss map

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

Lemma 6.1. *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

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

□

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

Definition 19. A full branch map has *bounded distortion* if

$$
\sup_{n\geq 1}\sup_{\omega^{(n)}\in\mathcal{P}^{(n)}}\sup_{x,y\in\omega^{(n)}}\log|Df^n(x)/Df^n(y)| < \infty. \tag{11}
$$

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

Theorem 8. *Let $f : I \to I$ be a full branch map with bounded distortion. Then Lebesgue measure is ergodic.*

Lemma 6.2. *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. □

6.2 Bounded distortion implies ergodicity

We now prove Theorem 8. 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 6.3. *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

$$\frac{|f^n(J')|}{|f^n(J)|} \frac{|J'|}{|J|} = \frac{|f^n(J')|/|J'|}{|f^n(J)|/|J|} = \frac{|Df^n(x)|}{|Df^n(y)|} \tag{12}$$

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 (12) gives

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

and rearranging gives the result. $\square$

Lemma 6.4. *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 6.3 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 $|\omega^{(n)}| \leq \tau|\omega^{(n-1)}| \leq \tau^2|\omega^{(n-2)}| \leq \cdots \leq \tau^n|\omega^0| \leq \tau^n|\Delta|$. $\square$

Proof of Theorem 8. The proof is almost identical to the piecewise affine case. The only difference is when we get to equation (10) where we now use the bounded distortion to get

$$\frac{|I \setminus A|}{|I|} = \frac{|f^n(\omega_n \setminus A)|}{|f^n(\omega_n)|} \leq e^{\mathcal{D}} \frac{|\omega_n \setminus A|}{|\omega_n|} \leq e^{\mathcal{D}} \epsilon. \tag{13}$$

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

6.3 Sufficient conditions for bounded distortion

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

Definition 20. 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 9. *Let f be a full branch map. Suppose that f is uniformly expanding and that there exists a constant $\mathcal{K} > 0$ such that*

$$\sup_{\omega \in \mathcal{P}} \sup_{x,y \in \omega} |f''(x)|/|f'(y)|^2 \leq \mathcal{K}. \tag{14}$$

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

$$\log \frac{|Df^n(x)|}{|Df^n(y)|} \leq \tilde{\mathcal{K}}|f^n(x) - f^n(y)| \leq \tilde{\mathcal{K}}. \tag{15}$$

In particular f satisfies the bounded distortion property.

Lemma 6.5. *The Gauss map is uniformly expanding and satisfies* (14).

Proof. We leave the verification that the Gauss map is uniformly expanding as an exercise. 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 $|f''(x)| \leq f''(1/(n+1)) = 2(n+1)^3$ and $|f'(y)| \geq |f'(1/n)| = n^2$. Therefore, for any $x, y \in \omega$ we have $|f''(x)|/|f'(y)|^2 \leq 2(n+1)^3/n^4 \leq 2((n+1)/n)^3(1/n)$. This upper bound is monotone decreasing with n and thus the worst case is $n = 1$ whch gives $|f''(x)|/|f'(y)|^2 \leq 16$ as required. $\square$

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

Lemma 6.6. *Let f be a full branch map satisfying* (14)*. Then, for all $\omega \in \mathcal{P}$, $x, y \in \omega$ we have*

$$\left|\frac{f'(x)}{f'(y)} - 1\right| \leq \mathcal{K}|f(x) - f(y)|. \tag{16}$$

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

$$|f'(x) - f'(y)| = \frac{|f''(\xi_2)|}{|f'(\xi_1)|}|f(x) - f(y)|. \tag{17}$$

Assumption (14) implies that $|f''(\xi_2)|/|f'(\xi_1)| \leq \mathcal{K}|f'(\xi)|$ for all $\xi \in \omega$. Choosing $\xi = y$ and substituting this into (17) therefore gives $|f'(x) - f'(y)| = \mathcal{K}|f'(y)||f(x) - f(y)|$ and dividing through by $|f'(y)|$ gives the result. $\square$

Lemma 6.7. *Let f be a full branch map satisfying* (16)*. Then, for any $n \geq 1$ and $\omega^{(n)} \in \mathcal{P}_n$ we have*

$$Dist(f^n, \omega^{(n)}) \leq \mathcal{K} \sum_{i=1}^{n} |f^i(x) - f^i(y)| \tag{18}$$

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

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

□

Lemma 6.8. *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 $n \geq 1$, $\omega^{(n)} \in \mathcal{P}_n$ and $x, y \in \omega^{(n)}$ we have*

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

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

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

□

7 Physical measures

We have proved above that Lebesgue measure is ergodic for every full branch map with bounded distortion. In particular this implies ergodicity for any absolutely continuous probability measure and thus for the Gauss measure which is invariant for the Gauss map. In general most maps do not have explicit formulae fot eh invariant measure but we show here that any full branch map with bounded distortion does have an ergodic invariant probability measure which is absolutely continuous with respect to Lebesgue and thus a physical measure.

Theorem 10. *Let $f : I \to I$ be a full branch map satisfying* (15)*. 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.*

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 32. 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.

Remark 11. 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.

The proof of the Theorem then relies on the following crucial

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

$$0 < \inf_{n,x} H_n(x) \leq \sup_{n,x} H_n(x) \leq K \tag{19}$$

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

$$|H_n(x) - H_n(y)| \leq K|H_n(x)|d(x,y) \leq K^2 d(x,y). \tag{20}$$

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 (19) and (20). 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. $\square$

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

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

□

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 (15). 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 (15) gives

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

Inverting x, y we also have

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

Combining these two bounds we get

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

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

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

□

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

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

□

8 Inducing

Most maps, even in one-dimension, are of course not full branch and/or do not satisfy the bounded distortion property. Thus the result we have proved above apples in principle only to a small class of systems. It turns out however that such full branch maps are often *embedded* in much more general. In this section we explain this idea and give an example of its application.

Definition 21. 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 $x \in \Delta$. Then the map $F : \Delta \to \Delta$ defined by

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

is called the *induced map* of f on Δ 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. For convenience, we introduce the notation

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

8.1 Spreading the measure

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. More specifically, if $f : X \to X$ is a map and $F := f^\tau : \Delta \to \Delta$ the induced map on some subset $\Delta \subseteq X$ corresponding to the return time function $\tau : \Delta \to \mathbb{N}$, and $\hat{\mu}$ is a probability measure on Δ, we can define a measure

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

By the measurability of τ each Δ_n is a measurable set. 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

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

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. Notice that ν is not a probability measure in general. Indeed, we have

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

If $\hat{\tau} := \int \tau d\hat{\mu} < \infty$, i.e. if the inducing time is integrable with respect to $\hat{\mu}$, then the total measure is finite and we can normalize it by defining

$$\mu = \frac{1}{\hat{\tau}} \sum_{n=1}^{\infty} \sum_{i=0}^{n-1} f_*^i(\hat{\mu}|_{\Delta_n}). \tag{21}$$

which is clearly a probability measure. The natural question then concerns the relation between the measure μ and the dynamics associated to the map f.

Proposition 8.1. *Let X be a measure space, $f : X \to X$ a measurable map, $f^\tau : \Delta \to \Delta$ an induced map, and $\hat\mu$ a probability measure on Δ with $\int \tau d\hat\mu < \infty$. The following holds.*

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[4] *with respect to m.*

3. *If $\hat\mu \ll m$ then $\mu \ll m$.*

As an immediate consequence we have the following

Theorem 11. *Suppose M is a Riemannina manifold with Lebesgue measure m and $f : M \to M$ 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 12. Notice that Proposition 13 and its Corollary are quite general and in particular apply to maps in arbitrary dimension.

Proof of Proposition 13. 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

$$\sum_{n=1}^{\infty} \hat\mu(B \cap \Delta_n) = \sum_{n=1}^{\infty} \hat\mu(f^{-n}(B) \cap \Delta_n)) \tag{22}$$

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

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

[4]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 $m(f^{-1}(A)) = 0$.

This proves (22) 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

$$F^{-1}(\hat{B})=\hat{B} \quad \text{and} \quad \hat{\mu}(\hat{B})=1. \tag{23}$$

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 (23), 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. □

8.2 Intermittency maps

We give a relatively simple but non-trivial application of the method of inducing. Let $\gamma\geq 0$ 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 $f(x) = 2x \mod 1$ (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 8.1. *F is uniformly expanding.*

Proof. Exercise. $\square$

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 8.2. *$x_n \approx 1/n^{\frac{1}{\gamma}}$ 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

$$|\Delta_n| \approx x_{(n)}^{1+\gamma} \quad \text{for } any\ x_{(n)} \in \Delta_n. \tag{24}$$

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

$$|J_k| \approx y_{(k)}^{1+\gamma} \quad \text{for } any\ y_{(k)} \in J_k \tag{25}$$

Combining (24) and (25) 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 $|\Delta_n| \approx x_n^{1+\gamma} = 1/n^{\frac{1}{\gamma}+1}$. $\square$

Lemma 8.3. *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

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

for some $\xi_i \in (f^i(x), f^i(y))$, where we have used here the Mean Value Theorem and the fact that $D(\log Df) = D^2f/Df$. 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

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

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

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

we get

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

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

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

□

Lemmas 8.1 and 8.3 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 8.4. *For $\gamma\in(0,1)$, the induced map F has integrable inducing times. Moreover, for every $n\ge 1$ we have*

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

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

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

□

8.3 Lyapunov exponents

Let $f : [0.1] \to [0,1]$ be a piecewise C^1 map and μ and f-invariant ergodic probability measure. Suppose that $\ln|f'| \in L^1_\mu$. Then, by Birkhoff's Ergodic Theorem, for μ a.e. x we have

$$\frac{1}{n}\sum_{i=0}^{n-1} \ln|f'(f^i(x))| \to \int \ln|f'|d\mu = \lambda_\mu.$$

We call $\lambda = \lambda_\mu$ the *Lyapunov exponent* associated to the measure μ. It gives the "asymptotic growth rate" of the derivative. Indeed, notice that by the chain rule, we have

$$\frac{1}{n}\sum_{i=0}^{n-1} \ln|f'(f^i(x))| = \frac{1}{n}\ln\prod_{i=0}^{n-1}|f'(f^i(x))| = \frac{1}{n}\ln|(f^n)'(x)|$$

and therefore, by the definition of limit, for any $\epsilon > 0$ there exists $N > 0$ such that for all $n \geq N$ we have

$$\lambda - \epsilon \leq \frac{1}{n}\ln|(f^n)'(x)| \leq \lambda + \epsilon$$

and so, taking exponentials,

$$e^{n(\lambda-\epsilon)} \leq |(f^n)'(x)| \leq e^{n(\lambda+\epsilon)}.$$

This means that that essentially the derivative has a well-defined asymptotic growth rate and that this growth rate is the same for μ almost every point and corresponds, as can be expected, to the average growth rate with respect to the measure μ.

Example 14. If f is a C^1 contraction with unique fixed point p, then the Lyapunov exponent with respect to the unique invariant ergodic measure δ_p is

$$\lambda := \int \ln|f'(x)|d\delta_p = \ln|f'(p)| < 0.$$

Moreover since $\ln|f'|$ is continuous and $f^n(x) \to p$ for every x, it is easy to verify that

$$\frac{1}{n}\ln|(f^n)'(x)| \to \lambda.$$

Example 15. If f is a piecewise C^2 full branch, uniformly expanding map with bounded distortion, and μ is its unique ergodic f-invariant absolutely continuous invariant probability measure, then

$$\lambda := \int \ln|f'(x)|d\mu > 0.$$

There is a natural relation between the Lyapunov exponent of a measure and the induced map from which such a measure was obtained. Indeed, suppose $F = F^\tau : \Delta \to \Delta$ is an induced map with an invariant ergodic probability measure $\hat{\mu}$ such that $\hat{\tau} := \int \tau\delta\hat{\mu} < \infty$ and such that $\ln|F'| \in L^1_{\hat{\mu}}$. Let $\hat{\lambda} = \int \ln|F'|d\hat{\mu}$ be the Lyapunov exponent associated to the measure $\hat{\mu}$ and let μ be the f-invariant ergodic probability measure corresponding to $\hat{\mu}$. Then we have the following.

Proposition 8.2. *The Lyapunov exponent associated to μ is $\lambda = \hat{\lambda}/\hat{\tau}$.*

Proof. Let $x \in \Delta$ and

$$R_n(x) := \tau(x) + \tau(F(x) + \cdots + \tau(F^{n-1}(x)).$$

Then, by the integrability of τ we have

$$\frac{1}{n} R_n(x) = \frac{1}{n} \sum_{i=0}^{n-1} \tau(F^i(x)) \to \int \tau d\hat{\mu} = \hat{\tau}.$$

Therefore

$$\frac{1}{n} \ln |(F^n)'(x)| = \frac{1}{n} \ln |(f^{R_n(x)})'(x)| = \frac{R_n}{n} \frac{1}{R_n} \ln |(f^{R_n(x)})'(x)|$$

Since the left hand side converges to $\hat{\lambda}$ and $R_n/n \to \hat{\tau}$ it follows that

$$\frac{1}{R_n} \ln |(f^{R_n(x)})'(x)| \to \frac{\hat{\lambda}}{\hat{\tau}}.$$

$\square$

Combining this with Theorem 13 we immediately get the following extension which we state in the one-dimensional setting, but also admits a higher-dimensional version.

Theorem 13. *Suppose $f : [0,1] \to [0,1]$ 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 $\hat{\mu}$ with positive Lyapunov exponent, for which the return time τ is integrable, then f admits an invariant ergodic absolutely continuous probability measure with a positive Lyapunov exponent.*

9 The quadratic family

9.1 The Ulam-Von Neumann map

In example 13 we studied the *Ulam-von Neumann* map $f : [-2,2] \to [-2,2]$ given by

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

We showed there the existence of an absolutely continuous invariant measure by using a special property of the map, namely the differentiable conjugacy with the tent map. Here we sketch the construction of an induced map as an alternative method for proving the same result. Of course this method does not give an explicit formula for the invariant measure but it is much more general and indeed can be applied also for maps fo the form $f_a(x) = x^2 - a$ for other values of the parameter a.

The key idea is to induce on the interval $[-p, p]$ where $-p$ is the fixed point in $(-2, 2)$. To be continued....

We conclude this chapter by introducing an important family of maps $f_a : \mathbb{R} \to \mathbb{R}$, the so-called *quadratic family*, defined by

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

Notice that for large negative of the parameter a there exists an interval I and two closed disjoint subintervals $I_0, I_1 \subset I$ on which f is expanding and such that $f(I_0) = f(I_1) = 1$. Thus for these parameters the maps have invariant Cantor sets as described in the previous chapter. For parameters $a < -2$ but close to -2 we still have two closed disjoint subintervals but the map is no longer expanding. nevertheless it is possible to show that in this case the weaker expansivity condition (**??**) holds, and thus we continue to have invariant Cantor sets. In this section we will focus on the parameter $a = -2$ in which we have a full branch map with two intervals whose closures are not disjoint and which also clearly cannot satisfy the expansivity condition (**??**) since it has a point where $f'(x) = 0$. Notice in particular that the symbolic coding argument cannot be applied, at least not directly, in this case, since we used in an esential way the expansivity properties of the map to to show that the symbolic coding was injective.

Proposition 9.1. *The maps $f : [-2, 2] \to [-2, 2]$ and $g : [0, 1] \to [0, 1]$ defined by*

$$f(x) = x^2 - 2 \quad \text{and} \quad g(z) = \begin{cases} 2z, & 0 \le z < \frac{1}{2} \\ 2 - 2z, & \frac{1}{2} \le z \le 1. \end{cases}$$

are topologically conjugate.

The map $f(x) = x^2 - 2$ is sometimes called the *Ulam-von Neumann* map.

Proof. This is one of the very exceptional situations in which we can find a conjugacy completely explicitly. Define the map $h : [0, 1] \to [-2, 2]$ by

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

h is clearly a (orientation reversing) homeomorphism and so we just need to show that it is a conjugacy, i.e. that it satisfies the conjugacy equation $f \circ h = h \circ g$. On one hand we have

$$f(h(z)) = f(2 \cos \pi z) = (2 \cos \pi z)^2 - 2 = 4 \cos^2 \pi z - 2 = 2(2 \cos^2 \pi z - 1) = 2 \cos 2\pi z.$$

On the other hand we have, for $z \in [0, 1/2)$,

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

and, for $z \in [1/2, 1]$,

$$h(g(z)) = h(2 - 2z) = 2 \cos \pi(2 - 2z) = 2 \cos(2\pi - 2\pi z) = 2 \cos(-2\pi z) = 2 \cos 2\pi z$$

This proves the conjugacy. $\square$

The conjugacy gives that

$$f^n = h \circ g^n \circ h^{-1}$$

and therefore, by the chain rule, letting $z = h^{-1}(x)$, we have

$$(f^n)'(x) = h'(g^n(h^{-1}(x))) \cdot (g^n)'(h^{-1}(x)) \cdot (h^{-1})'(x) = \frac{h'(g^n(z))}{-h'(z)}(g^n)'(h^{-1}(x))$$

Since $h'(z) = -2\pi \sin \pi z$ and $|g'(z)| \equiv 2$ this gives

$$|(f^n)'(x)| = \frac{|\sin \pi(g^n(z))|}{|\sin \pi z|} 2^n$$

which shows that the derivative along every orbit grows at a rate 2^n with a constant that depend sonly on the initial and final point in the orbit, and which in particular can get arbitrarily small. In particular we have that

$$\frac{1}{n} \ln |(f^n)'(x)| = \frac{1}{n}\left(\ln \frac{|\sin \pi(g^n(z))|}{|\sin \pi z|}\right) + \ln 2.$$

and so

$$\limsup_{n\to\infty} \frac{1}{n} \ln |(f^n)'(x)| = \ln 2.$$

Proposition 9.2. *For every periodic point $p \in (-2, 2)$ of period n, $|(f^n)'(p)| = 2^n$.*

Remark 13. Notice that by the topological conjugacy with the tent map, f has a dense set of periodic points which means that there are periodic points arbitrarily close to the critical point. For a periodic point p of period n we have

$$|(f^n)'(p)| = |f'(p)f'(f(p)) \cdots f'(f^{n-1}(p))|$$

thus the result says that for any periodic orbit the derivatives compensate each other exactly. In particular, any orbit for which some point lies very close to the critical point must have a very high period in order to compensate the small derivative near the critical point.

Proof. The assumption that $p \in (-2, 2)$ implies that the entire orbit lies in $(-2, 0)\cup(0, 2)$. Indeed, if some iterate of p falls on the critical point at 0 or on one of the endpoints ± 2, it would then fall onto the fixed point at 2 contradicting the assumption that p is a periodic orbit and that $p \in (-2, 2)$. Since the entire orbit lies in $(-2, 2)$ we can use the fact that the conjugacy h is C^1 and that p is a fixed point for f^n and is therefore mapped to a fixed point q for g^n, to get that the derivative of f^n at p is the same as the derivative of g^n at q which is necessarily 2^n. $\square$

9.2 The Quadratic Family

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 14. *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 Jakobson 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).

10 Mixing Measures

Our emphasis so far has been on the existence of physical measures for which we obtain a description of the statistical distribution of typical orbits in phase space. However many dynamical systems may share a same physical measures, for example irrational circle rotations and the family of maps of $f(x) = \kappa x \bmod 1$ for $\kappa \geq 2$ an integer all admit Lebesgue measure as an invariant and ergodic measure. Thus almost all orbits are uniformly distributed with respect to Lebesgue. Nevertheless these are all different maps and the dynamics of each one contains some characteristic features. We have already seen that indeed irrational circle rotations exhibit some extra rigidity in that *every* orbit is uniformly distributed. Moreover, intuitively the dynamics of circle rotations is more "regular" than that of expanding maps which is quite "chaotic". In this section we introduce the notion of *mixing* which is a formal way to make a finer distinction between different kinds of dynamical behaviour, and in particular will enable us to formally distinguish between irrational circle rotations and expanding maps. A further, in some sense even finer, distinction can be achieved through the concept of *entropy* which distinguishes for example expanding maps with different number of branches. However the treatment of this concept is beyond the scope of these notes.

10.1 Mixing

Let M be a measure space and $f : M \to M$ a measurable map. Suppose that μ is an f-invariant probability measure.

Definition 22. μ is *mixing* if, for all measurable sets $A, B \subseteq M$,

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

as $n \to \infty$.

A good way to understand this definition is by dividing through by $\mu(B)$ and writing

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

or, using the fact that μ is f-invariant,

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

What this says therefore is that the *proportion* of $f^{-n}(B)$ which intersects A converges simply to the measure of A, i.e. $f^{-n}(B)$ is increasingly uniformly distributed over the whole space (as seen through the measure μ).

Lemma 10.1. *Suppose μ is mixing. Then it is ergodic.*

Proof. Exercise. □

As we shall see, however, the opposite is false and there are lots of systems which are ergodic but not mixing.

Example 16. Let $\{p_1, \dots, p_k\}$ denote the periodic orbit with $k \geq 2$ and consider the Dirac measure uniformly distributed on the points of the orbit. Let $A = p_j$ and $B = p_{j'}$ for some $j \neq j'$. Then $f^{-n}(A) \cap B \neq \emptyset$ for an infinite number of iterates and thus the measure cannot be mixing.

Remark 14. Notice that this argument does not work if $k = 1$, i.e. for a fixed point. Indeed, if p is a fixed point, and $\mu = \delta_p$ the Dirac measure on p, then for any two sets A, B we one of three possibilities. If they both contain p then $f^{-n}(A) \cap P$ also contains p and therefore $\mu(f^{-n}(A) \cap P) \equiv 1$ and $\mu(A)\mu(B) = 1$ as required. If neither contains p then we just have zero measure all round which also works. If only one contains p then we also have zero measure all round which also formally satisfies the definition. This is a kind of "anomaly" in the sense that it formally satisfies the definition of mixing even though it does not really satisfy the "spirit" of the definition. However this case essentially means that we are living in a one-point space and is therefore naturally degenerate.

Example 17. For an irrational circle rotation, consider two small intervals A, B. Then $f^{-n}(A) \cap B = \emptyset$ for some infinite number of times and therefore again clearly cannot be mixing.

Example 18. Consider the map $f(x) = 10x \bmod 1$. We will not give a formal argument here, but notice that given any set B of positive Lebesgue measure, the pre-images $f^{-n}(B)$ consist of exactly 10^n scaled down "copies" of B uniformly distributed in the unit interval. Thus it is intuitively clear that any given set A will intersect a proportion of these pre-images which converges to the measure of A, i.e. the proportion of A in the unit interval. The same argument works for $f(x) = \kappa x$ for any other integer $\kappa \geq 2$. Indeed, we will not prove it here but the unique ergodic invariant absolutely continuous probability measure for any full branch uniformly expanding maps with bounded distortion is also always mixing.

Mixing is also easily seen to be preserved under conjugacies. More precisely, let $f : X \to X$ and $g : Y \to Y$ be two measurable maps and let $h : X \to Y$ be a measurable conjugacy. Suppose that ν is a measure on X and define $\mu = h_*\nu$ so that $\mu(A) = \nu(h^{-1}(A))$. Then we have shown above μ is invariant and ergodic for g as long as ν is invariant and ergodic for f.

Lemma 10.2. *Suppose ν is mixing for f. Then μ is mixing for g.*

Proof. By the definition of conjugacy we have

$$h^{-1}(g^{-n}(A)) = \{x : g^n(h(x)) \in A\} = \{x : h(f^n(x)) \in A\} = f^{-n}(h^{-1}(A)).$$

Therefore, using the mixing of ν we have, for any measurable sets $A, B \subseteq X$,

$$\begin{aligned} \mu(g^{-n}(A) \cap B) &= \nu(h^{-1}(g^{-n}(A) \cap B)) = \nu(h^{-1}(g^{-n}(A))) \cap h^{-1}(B) \\ &= \nu(f^{-n}(h^{-1}(A)) \cap h^{-1}(B)) \to \nu(h^{-1}(A))\nu(h^{-1}(B)) = \mu(A)\mu(B). \end{aligned}$$

□

10.2 Decay of correlations

A natural question concerns the *speed of mixing*. If a measure μ is mixing for some map f is there a particular speed at which the mixing occurs, e.g. is it exponential? A priori there is no reason that such a rate should exist *independently* of the choice of sets A, B and indeed it is known that in general, it is possible to find subsets A, B such that the convergence in the definition of mixing is arbitrarily slow. However it turns out that we can still talk meaningfully about rates of mixing by simultaneously generalising and restricting the notion of mixing through the notion of *correlations*.

Definition 23. 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 *decays* if $\mathcal{C}_n(\varphi,\psi) \to 0$ as $n \to \infty$.

If φ, ψ are characteristic functions of sets A, B we recover the expression used in the definition of mixing:

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

It turns out that it is sometimes possible to give precise estimates for the rate of decay of the correlations function as long as we restrict our attention to specific classes of functions.

Definition 24. 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 polynomially. The key point here is that the *rate*, i.e. the sequence $\{\gamma_n\}$ is not allowed to depend on the observables but only on the class of observables. Thus the rate of decay becomes in some sense an intrinsic property of the system (and of the class of observables). As mentioned above, we cannot hope to obtain decay of correlations for classes of functions which are too big, e.g. that contain all characteristic functions. Most results that obtain rates of decay of correlations do so for Hölder continuous function or functions of bounded variation, but other results also exist for functions with weaker continuity properties or, in higher dimensions, even with quite strange regularity conditions.

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 seen 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, we recall that

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

Theorem 15. *The rate of decay of correlation with respect to $\mu$is determined by the regularity of the observables and therate of decay of $|\Delta_n|$. For Hölder continuous observables, if $|\Delta_n| \to 0$ exponentially, then the rate of decay is exponential, if $|\Delta_n| \to 0$ polynomially, then the rate of decay is polynomial. For non-Hölder continuous observables the rate slows down by factors related to the modulus of continuity.*

The precise formulation of this statement is contained in several papers.

References

[1] Lynch, Vincent, Decay of correlations for non-Hölder observables. Discrete Contin. Dyn. Syst. 16 (2006), no. 1, 19–46.

[2] Young, Lai-Sang Statistical properties of dynamical systems with some hyperbolicity. Ann. of Math. (2) 147 (1998), no. 3, 585–650.

[3] Young, Lai-Sang Recurrence times and rates of mixing. Israel J. Math. 110 (1999), 153–188.

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.

A Review of measure theory

In this section we introduce only the very minimal requirements of Measure Theory which will be needed later. For a more extensive introduction see any introductory book on Measure Theory or Ergodic Theory, for example [?Kol33, ?Bil79, ?KF60, ?Fal97]. For simplicity, we shall restrict ourselves to measures on the unit interval $I = [0,1]$ although most of the definitions apply in much more general situations.

A.1 Definitions

A.2 Basic motivation: Positive measure Cantor sets

The notion of measure is, in the first instance, a generalization of the standard idea of *length*. Indeed, while we know how to define the length of an interval, we do not apriori know how to measure the size of sets which contain no intervals but which, logically, have positive "measure" Let $\{r_i\}_{i=0}^{\infty}$ be a sequence of positive numbers with $\sum r_i < 1$. We define a set $\mathcal{C} \subset [0,1]$ by recursively removing open subintervals from $[0,1]$ in the following way. Start by removing an open subinterval I_0 of length r_0 from the interior of $[0,1]$. Then $[0,1] \setminus I_0$ has two connected components. Remove intervals I_1, I_2 of lengths r_1, r_2 respectively from the interior of these components. Then $[0,1] \setminus (I_0 \cup I_1 \cup I_2)$ has 4 connected components. Now remove intervals $I_3, \ldots, I_7$ from each of the interiors of these components and continue in this way. Let

$$\mathcal{C} = [0,1] \setminus \bigcup_{i=0}^{\infty} I_i$$

Then $\mathcal{C}$ does not contain any intervals since every interval is eventually subdivided by the removal of one of the subintervals I_k from its interior, and therefore it does not make sense to talk about $\mathcal{C}$ as having any *length*. However the total length of the intervals removed is $\sum r_i < 1$ and therefore it would make sense to say that the size of $\mathcal{C}$ is $1 - \sum r_i$. The Theory of Measures formalizes this notion in a rigorous way and makes it possible to of assign a size to sets such as $\mathcal{C}$.

A.3 Non-measurable sets

The example above shows that it is desitable to generalize the notion of "length" to a notion of "measure" which can apply to more complicated subsets which are not intervals and which can formalize what we mean by saying for example that the Cantor set defined above has positive measure. It turns out that however that in general it is not possible to define a measure in a consistent way on all possible subsets. In 1924 Banach and Tarski showed that it is possible to divide the unit ball in 3-dimensional space into 5 parts and re-assemble these parts to form two unit balls, thus apparently doubling the volume of the original set. This implies that it is impossible to consistently assign a well defined volume in an additive way to *every* subset. See a very interesting discussion on wikipedia on this point.

A simpler example is the following. Consider the unit circle $\mathbb{S}^1$ and an irrational circle rotation $f_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$. Then every orbit is dense in $\mathbb{S}^1$. Let $A \subset \mathbb{S}^1$ be a set containing exactly one point from each orbit. Assuming that we have defined a general notion of a measure for which the measure $m(A)$ has meaning and that generalizes the length of intervals so that the measure of any interval coincides wth its length. In particular such a measure will be translation invariant in the sense that the measure of a set cannot be changed by simply translating this set. Therefore, since a circle rotation f is just a translation we have $m(f^n(A) = m(A)$ for every $n \in \mathbb{Z}$. Morover, since A contains only one single point from each orbit and all points on a given orbit are distinct we have $f^n(A) \cap f^m(A) = \emptyset$ for all $m, n \in \mathbb{Z}$ with $m \neq n$ and therefore we have

$$1 = m(\mathbb{S}^1) = m\left(\bigcup_{i=-\infty}^{+\infty} f^n(A)\right) = \sum_{i=-\infty}^{+\infty} m(f^n(A)) = \sum_{i=-\infty}^{+\infty} m(A)$$

This is clearly impossible as the right hand side is zero if $|A| = 0$ or infinity if $|A| > 0$.

Remark 15. This counterexample depends on the *Axiom of Choice* to ensure that it is psossible to define such a set constructed by choosing a single point from each of an uncountable family of subsets.

A.4 Algebras and sigma-algebras

Let X be a set and $\mathcal{A}$ a collection of (not necessarily disjoint) subsets of X.

Definition 25. We say that $\mathcal{A}$ is an *algebra* (of subsets of X if

1. $\emptyset \in \mathcal{A}$ and $X \in \mathcal{A}$.
2. $A \in \mathcal{A}$ implies $A^c \in \mathcal{A}$
3. for any finite collection $A_1, \ldots, A_n$ of subsets in $\mathcal{A}$ we have
$$\bigcup_{i=1}^{n} A_i \in \mathcal{A}$$

We say that $\mathcal{A}$ is a σ-*algebra* (*sigma-algebra*) if moreover

(3') for any countable collection $A_1, A_2, \ldots$ of subsets in $\mathcal{A}$, we have
$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}.$$

Given an algebra $\mathcal{A}$ of subsets of a set X we define the *sigma*-algebra $\sigma(\mathcal{A})$ as the smallest σ-algebra containing $\mathcal{A}$. This is always well defined and is in general smaller than the *sigma*-algebra of all subsets of X.

A.5 Measures

Let X be a set and $\mathcal{A}$ be a σ-algebra of subsets.

Definition 26. A *measure* is a function

$$\mu : \mathcal{A} \to [0, \infty]$$

which is *countably additive*, i.e.

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

for any countable collection $\{A_i\}_{i=1}^{\infty}$ of *disjoint* sets in $\mathcal{A}$.

This definition shows that the σ-algebra is aas intrinsic to the definition of a measure as the space itself. In general therefore we talk of a *Measure Space* as a triple $(X, \mathcal{A}, \mu)$ although the space and the σ-algebra are often omitted if they are given as fixed.

Remark 16. We say that μ is a *finite measure* if $\mu(X) < \infty$ and that it is a *probability measure* if $\mu(X) = 1$. Notice that if $\hat{\mu}$ is a finite measure we can easily define a probability measure μ by simply letting

$$\mu = \frac{\hat{\mu}}{\hat{\mu}(X)}.$$

The fact that such a countably additive function exists is non-trivial. It is usually easier to find finitely additive functions on algebras; for example the standard *length* is a *finitely* additive function on the *algebra* of finite unions of intervals. The fact that this extends to a countably additive function on the corresponding σ-algebra is guaranteed by the following fundamental

Theorem (Extension Theorem)**.** *Let $\tilde{\mu}$ be a finitely additive function defined on an algebra $\tilde{\mathcal{A}}$ of subsets. Then $\tilde{\mu}$ can be extended* in a unique way *to a countably additive function μ on the σ-algebra $\mathcal{A} = \sigma(\tilde{\mathcal{A}})$.*

In the case in which X is an interval $I \subseteq \mathbb{R}$ (or the unit circle $\mathbb{S}^1$ which we think of as just the unit interval with its endpoints identified) there is a vary natural *sigma*-algebra.

Definition 27. Let $\tilde{\mathcal{B}}$ denote the algebra of all finite unions of subintervals of I. Then, the generated σ-algebra $\mathcal{B} = \sigma(\tilde{\mathcal{B}})$ is called the *Borel* σ-algebra. Any measure defined on $\mathcal{B}$ is called a *Borel* measure.

Remark 17. Notice that a Cantor set $\mathcal{C} \subset I$ is the complement of a countable union of open intervals and therefore belongs to Borel σ-algebra $\mathcal{B}$.

A.6 Integration

The abstract notion of *measure* leads to a powerful generalization of the standard definition of Riemann integral. For $A \in \mathcal{B}$ we define the characteristic function

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

A simple function is one which can be written in the form

$$\zeta = \sum_{i=1}^{N} c_i \chi_{A_i}$$

where $c_i \in \mathbb{R}^+$ are constants and the A_i are disjoint Borel measurable sets. These are functions which are "piecewise constant" on a finite partition $\{A_i\}$ of X.

Definition 28 (Integrals of nonnegative functions)**.** For simple functions let

$$\int_X \zeta d\mu = \sum_{i=1}^{N} c_i \mu(A_i).$$

Then, for general, measurable, *non-negative* f we can define

$$\int_X f d\mu = \sup \left\{ \int_X \zeta d\mu : \zeta \text{ simple, and } \zeta \leq f \right\}.$$

The integral is called the *Lebesgue integral* of the function f with respect to the measure μ (even if μ is not Lebesgue measure).

Remark 18. Notice that, in contrast to the case of Riemann integration in which the integral is given by a limiting process which may or may not converge, this supremum is always well defined, though it may not always be finite.

More generally, for any measurable f we can write $f = f^+(-f^-)$ where $f^+(x) = max\{f(x), 0\}$ and $f^-(x) = -min\{0, f(x)\}$ both of which are clearly non-negative.

Definition 29 (Integral of any measurable function)**.** Let f be a μ measurable function. If

$$\int f^+ d\mu < \infty \quad \text{and} \quad \int f^- d\mu < \infty$$

then we say that f is μ-integrable and let

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

We let $\mathcal{L}^1(\mu)$ denote the set of all μ-integrable functions.

Example 19. Let $f : [0,1] \to \mathbb{R}$ be given by

$$f(x) = \begin{cases} 0 & \text{if } x \in \mathbb{Q} \\ 1 & \text{otherwise.} \end{cases}$$

Notice that this function is not Riemann integrable in the sense that the required limit does not converge. From the point of view discussed above, however, it is just a simple function which takes the value 0 on the measurable set $\mathbb{Q}$ and the value 1on the measurable set $\mathbb{R} \setminus \mathbb{Q}$. For $m =$ Lebesgue measure we have $m(\mathbb{Q}) = 0$ since $\mathbb{Q}$ is countable, and therefore $m((\mathbb{R} \setminus \mathbb{Q}) \cap [0,1]) = 1$ and so

$$\int_{[0,1]} f dm = m((\mathbb{R} \setminus \mathbb{Q}) \cap [0,1]) = 1.$$

A.7 Lebesgue density theorem

Theorem 16 (Lebesgue Density Theorem)**.** *Let μ be a probability measure on I and let A be a measurable set with $\mu(A) > 0$. Then for μ almost every point $x \in A$ we have*

$$\frac{m(x-\epsilon, x+\epsilon)}{2\epsilon} \to 1 \tag{26}$$

as $\epsilon \to 0$.

Points x satisfying (26) are called (Lebesgue) density points of A. This result says that in some very subtle way, the measure of the set A is "bunched up". A priori one could expect that if $\mu(A) = 1/2$ then for any subinterval J the ratio between $A \cap J$ and J might be 1/2, i.e. that the ratio between the measure of the whole interval and the measure of the set A is constant at every scale. This theorem shows that this is not the case. We shall not prove this result here.

A.8 Absolutely continuous and singular measures

Definition 30. Let μ_1, μ_2 be probability measures.

1. μ_1 is *absolutely continuous* wrt μ_2 if $\mu_2(A) = 0 \Rightarrow \mu_1(A) = 0$ for every A.

2. μ_1, μ_2 are *mutually singular* if there exists A such that $\mu_1(A) = 1$ and $\mu_2(A) = 0$.

If μ_1 is absolutely continuous with respect to μ_2 we write $\mu_1 \ll \mu_2$. If $\mu_1 \ll \mu_2$ and $\mu_2 \ll \mu_1$ then we say that μ_1 and μ_2 are *equivalent*.

Example 20. Let $X = [0,1]$ and m denote Lebesgue measure. Let $\varphi \in L^1(m)$ be a non-negative function with $\int \varphi dm = 1$, and define the measure μ_φ by

$$\mu_\varphi(A) = \int_A \varphi dm.$$

Then it is easy to see that $\mu_\varphi \ll m$. In fact the Radon-Nykodim Theorem says that all absolutely continuous invariant measures are of this form: if $\mu_1 \ll \mu_2$ then there exists a non-negative function $\varphi \in L^1(\mu_2)$ with $\int \varphi d\mu_2$ such that

$$\mu_1(A) = \int_A \varphi d\mu_2 \tag{27}$$

for any measurable set A. The function φ is called the *density* or the *Radon-Nykodim derivative* of μ_1 with respect to μ_2, and is is sometimes written as $\varphi = d\mu_1/d\mu_2$

Example 21. Let $X = [0,1]$ and m denote Lebesgue measure. Then, for any $x \in [0,1]$ the Dirac delta measure δ_x and Lebesgue measure are mutually singular.

Exercise 33. Suppose $\mu_1 \ll \mu_2$. Show that for any measurable set A, $\mu_1(A) > 0 \Rightarrow \mu_2(A) > 0$ and $\mu_2(A) = 1 \Rightarrow \mu_1(A) = 1$

It is not the case that any two distinct measures need to be either absolutely continuous or mutually singular. For example if we let $\mu_1 = (\delta_p + \mu_2)/2$ where δ_p is a Dirac measure on some point p. Then μ_1 and μ_2 are neither absolutely continuous nor mutually singular.