An introduction to Smooth Ergodic Theory for one-dimensional dynamical systems

Stefano Luzzatto
Abdus Salam International Centre for Theoretical Physics
Strada Costiera 11, Trieste, Italy
luzzatto@ictp.it

Draft version: August 31, 2010

Contents

Introduction and motivation

The material to be presented in this course is motivated by the general and classical problem of studying an autonomous Ordinary Differential Equation

$$\dot{x} = f(x).$$

This problem goes back centuries and through the years many different approaches and techniques have been developed. The most classical approach is that of finding explicit analytic solutions. This can provide a great deal of information but is essentially only applicable to an extremely restricted class of differential equations. From the beginning of the 20th century there has been great development on topological methods to obtain qualitative topological information such as the existence of periodic solutions. Again this can be a very successful approachin certain situation but there are many equations which have for example infinitely many periodic solutions possibly intertwined in very complicated ways to which these methods do not really apply. Finally there are numerical methods for approximating solutions. In the last few decades with the increasing computing power there has been hope that numerical methods could play an important role. Again, while this is true in many situations, there are also many equations for which the numerical methods have very limited applicability because the approximation errors grow exponentially and quickly become uncontrollable. Moreover this "sensitive dependence on initial conditions" is now understood to be an intrinsic feature of certain equations than cannot be resolved by increasing the computing power.

Example 1. The Lorenz equations were introduced by the metereologist E. Lorenz in 1963, as an extremely simplified model of the Navier-Stokes equations for fluid flow.

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

This is a very good example of a relatively simple ODE which is quite intractable from many angles. It does not admit any explicit analytic solutions; the topology is extremely complicated with infinitely many periodic solutions which are knotted in many different ways (there are studies from the point of view of knot theory of the structure of the periodic solutions in the Lorenz equations); numerical integration has very limited use since nearby solutions diverge very quickly.

One thing that can be proved about the Lorenz equations using classical methods is the fact that all solutions eventually end up in some bounded region $\mathcal{U} \subset \mathbb{R}^3$. This simplifies things significantly since it means that it is sufficient to concentrate on the solutions inside $\mathcal{U}$. A combination of results obtained over almost 40 years by several different people can be formulated in the following theorem which can be thought of essentially as a statement in ergodic theory.We give here a precise but slightly informal statement as some of the terms will be defined more precisely below.

Theorem 1 (1963-2000, combinations of several results by different people)**.** *For every ball $B \subset \mathbb{R}^3$, there exists a "probability" $p(B) \in [0,1]$ such that for "almost every" initial condition $x \in \mathbb{R}^3$ we have*

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T \mathbb{1}_B(x_t)dt = p(B).$$

A few remarks about this result. Recall first of all that $\mathbb{1}_B$ is the *characteristic function* of the set B defined by

$$\mathbb{1}_B(x) = \begin{cases} 0 & \text{if } x \notin B \\ 1 & \text{if } x \in B \end{cases}$$

Moreover, x_t denotes the solution with initial condition $x_0 = x$ and therefore the integral $\int_0^T \mathbb{1}_B(x_t)dt$ is simply the amount of time that the solution spends inside the ball B between time 0 and time T, and $T^{-1}\int_0^T \mathbb{1}_B(x_t)dt$ is simply the *proportion* of time that the solution spends in the ball B between time 0 and time T. The Theorem therefore makes two highly non trivial assertions:

1. that this proportion *converges*;
2. that the limit is *independent* of x.

The convergence itself is non trivial as there is no a priori reason why this should be true. But perhaps themost remarkable fact is that this limit is the same for almost all initial conditions (the notion of "almost all" will be made precise below). This says that the *asymptotic time averages* of the solution x_t with initial condition $x = x_0$ are actually independent of this initial condition. In this sense we can really talk about the ball B having a certain probability in the sense that there is a given probability that the solution through a random initial condition at some random time as that particular probability of belonging to the set B.

The moral of the story is that even though the Lorenz equations are difficult to describe from an analytic or topological point of view, and are essentially intractable form a numerical point of view, they are very well behaved from a *probabilistic* point of view. The tools and methods of probability theory are therefore very well suited to study and understand these equations and other similar dynamical systems. This is essentially the point of view on ergodic theory that we will take in these lectures. Since this is an introductory course we will focus on the simplest examples of dynamical systems for which there is already an extremely rich and interesting theory, which are one-dimensional maps of the interval or the circle. However, the ideas and methods which we will present often apply in much more generality and usually form at least the conceptual foundation for analogous results in higher dimensions. In fact in some situation results about interval maps are applied directly to higher dimensional situations. For example in the Lorenz equations it turns out that the result stated above does essentially reduce to an analogous result for one-dimensional maps by taking a cross section for the flow and the Poincaré first return map to this cross section.

Chapter 1

Invariant measures

Let $(X, \mathcal{B}, \mu)$ be a probability measure space. We say that a map $f : X \to X$ is *measurable* if $f^{-1}(A) \in \mathcal{B}$ for all $A \in \mathcal{B}$. We shall always assume that our maps are measurable.

Definition 1. A measure μ is *f-invariant* if, for every $A \in \mathcal{B}$ we have

$$\mu(f^{-1}(A)) = \mu(A).$$

We also say that f is *measure-preserving* with respect to μ.

Exercise 1. Show that if f is invertible then this condition is equivalent to $\mu(f(A)) = \mu(A)$. Find an example of a non-invertible map and a measure μ for which the two conditions are not equivalent.

Invariant measures play a fundamental role in dynamics. To motivate the definition we begin by stating and proving an abstract result about the dynamics of maps having an invariant measure. We then give several examples of invariant measures and conclude this chapter with another abstract result concerning the structure of the space of invariant measures.

1.1 Poincaré's Recurrence Theorem

Theorem (Poincaré Recurrence Theorem)**.** *Let $(X, \mathcal{B}, \mu)$ be a probability space and $f : X \to X$ a measure-preserving map. Let $A \in \mathcal{B}$ with $\mu(A) > 0$. Then μ almost every point $x \in A$ returns to A inifinitely often.*

Exercise 2. The finiteness of the measure μ plays a crucial role in this result. Find an example of an infinite measure space $(\hat{X}, \hat{\mathcal{B}}, \hat{\mu})$ and a measure-preserving map $f : \hat{X} \to \hat{X}$ for which the conclusions of Poincare's Recurrence Theorem do not hold.

Proof. We show first of all that μ almost every point in A returns to A at least once. Let

$$A'_0 = \{x \in A : f^n(x) \notin A \text{ for all } n \geq 1\}$$

denote the set of points of A that never return to A. Let

$$A'_n = f^{-n}(A'_0).$$

It is easy to see that $A'_n \cap A'_m = \emptyset$ for all $m, n \geq 0$ with $m \neq n$. Indeed, supppose by contradiction that there exists $n > m \geq 0$ such that there exists $x \in A'_n \cap A'_m$. This would imply that

$$f^n(x) \in f^n(A'_n \cap A'_m) = f^n(f^{-n}(A'_0) \cap f^{-m}(A'_0)) = A'_0 \cap f^{n-m}(A'_0)$$

But this implies $A'_0 \cap f^{n-m}(A'_0) \neq \emptyset$ which contradicts the definition of A'_0. By the invariance of the measure μ we have $\mu(A'_n) = \mu(A')$ for every $n \geq 1$ and therefore

$$1 = \mu(X) \geq \mu(\bigcup_{n=1}^{\infty} A'_n) = \sum_{n=1}^{\infty} \mu(A'_n) = \sum_{n=1}^{\infty} \mu(A').$$

Thus $\mu(A') = 0$ since otherwise the sum on the right hand side would be infinite. This completes the proof that almost every point of A returns to A at least once.

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

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

□

1.2 The space of invariant measures

We now prove our first general theorem about the existence of invariant measures. Let X be a metric space and let $\mathcal{M}$ denote the set of all Borel probability measures on X. Note that $\mathcal{M} \neq \emptyset$ since it always contains for examples the Dirac-delta measures δ_x on points of X. Now let $f : X \to X$ be a Borel measurable map and let $\mathcal{M}_f \subseteq \mathcal{M}$ denote the set of f-invariant Borel probability measures on I. We would like to study certain properties of the set $\mathcal{M}_f$. Recall that by definition $\mathcal{M}_f$ 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 $\mu_t \in \mathcal{M}_f$.

Exercise 3. Show that $\mathcal{M}_f$ is convex.

The convexity is of course trivial if $\mathcal{M}_f = \emptyset$ and this can indeed happen.

Exercise 4. Find a continuous map on the open interval $(0,1)$ such that $\mathcal{M}_f = \emptyset$.

Theorem 2 (Krylov-Boguliobov)**.** *If X is compact and f is continuous, then $\mathcal{M}_f \neq \emptyset$.*

We define a map

$$f_* : \mathcal{M} \to \mathcal{M}$$

by letting, for any measurable set A,

$$f_*\mu(A) = \mu(f^{-1}(A)). \tag{1.1}$$

We call $f_*\mu$ the *push-forward* of μ. Similarly we can define $f_*^i\mu(A) = \mu(f^{-i}(A))$.

Exercise 5. Show that the map f_* is well defined, i.e. that $f_*\mu$ is a probability measure.

It follows immediately from the definition that μ is f-invariant if and only if $f_*\mu = \mu$, i.e. if μ is a fixed point for the map f_*. It is therefore sufficient to show that f_* has a fixed point.

Lemma 1.2.1. *The map $f_* : \mathcal{M} \to \mathcal{M}$ is continuous.*

Proof. We show first of all that for all $\varphi \in L^1(\mu)$ we have

$$\int \varphi d(f_*\mu) = \int \varphi \circ f d\mu. \tag{1.2}$$

In particular, if μ is invariant, then

$$\int \varphi d\mu = \int \varphi \circ f d\mu.$$

First let $\varphi = \mathbb{1}_A$ be the characteristic function of some set $A \subseteq X$. In this case we have

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

The statement is therefore true for characteristic functions and thus follows for general measurable 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

$$\int \varphi_n d(f_*\mu) \to \int \varphi d(f_*\mu) \quad \text{and} \quad \int \varphi_n \circ f d\mu \to \int \varphi \circ f d\mu$$

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.

The continuity of f_* now follows easily from (1.2). Indeed, suppose $\mu_n \to \mu$ in $\mathcal{M}$. Then, by the definition of convergence in the weak star topology, for any continuous function $\varphi : X \to \mathbb{R}$ and the previous Lemma, 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. □

Proof of Theorem 2. Let $\mu_0 \in \mathcal{M}$ be an arbitrary measure, for example the Dirac-delta measure δ_x for some arbitrary point $x \in X$. Define the sequence of measures

$$\mu_n = \frac{1}{n} \sum_{i=0}^{n-1} f_*^i \mu_0. \tag{1.3}$$

Since each $f_*^i \mu_0$ is a probability measure, the same is also true for μ_n. We recall that $\mu_n \to \mu$ in the *weak-⋆* (weak-star) topology if

$$\int \varphi d\mu_n \to \int \varphi d\mu \quad \text{as } n \to \infty$$

for all continuous functions φ. By standard results of analysis, the space $\mathcal{M}$ of probability measures is compact in the weak star topology. In particular, there exists a measure $\mu \in \mathcal{M}$ and a subsequence $n_j \to \infty$ with

$$\mu = \lim_{j \to \infty} \mu_{n_j}.$$

We will show that any such limit point μ is invariant, i.e. $f_*\mu = \mu$. First of all, by the continuity of f_* and the fact that $\mu_{n_j} \to \mu$ we have $f_*\mu_{n_j} \to f_*\mu$. It is therefore sufficient to show that also $f_*\mu_n \to \mu$. 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 concludes the proof. □

1.3 Fixed and periodic orbits

We now begin a series of explicit examples of invariant measures. Let X be a metric space and $p \in X$ a point. The *Dirac* measure δ_p is

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

In this case the entire mass is concentrated on the single point p.

Proposition 1.3.1. *Let X be a metric space and $f : X \to X$ a measurable map. Suppose $f(p) = p$. Then the Dirac measure δ_p is invariant.*

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

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

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

Exercise 6. δ_P is invariant if and only if $\rho_i = 1/n$ for every $i = 1, \ldots, n$.

1.4 Circle rotations

Proposition 1.4.1. *Let $X = \mathbb{S}^1$ and $f(x) = x + \alpha$ for some $\alpha \in \mathbb{R}$. Then Lebesgue measure is invariant.*

Proof. f is just a translation and Lebesgue measure is invariant under translations. □

However depending on the value of α there may be other invariant measures as well.

Exercise 7. Show that for $\alpha = 0$ every $\mu \in \mathcal{M}$ is invariant.

1.5 Piecewise affine full branch maps

We now consider an important class of maps on an interval I. For any subinterval $J \subseteq I$ we shall write $int(J)$ the denote the interior of J.

Definition 2. We say that $f : I \to I$ is a *full branch* map if there exists a finite or countable partition $\mathcal{P}$ of $I(\bmod 0)$ into subintervals such that for each $\omega \in \mathcal{P}$ the map $f|_{int(\omega)} : int(\omega) \to int(I)$ is a C^2 diffeomorphism.

We recall that a partition $\mathcal{P}$ of $I(\bmod 0)$ into subintervals means that there exists a family $\mathcal{P} = \{\omega_i\}$ of subintervals of I with disjoint interiors and a subset $\tilde{I} \subset I$ of full Lebesgue measure such that $\tilde{I} \subset \cup_i \omega_i$. The cardinality of the family $\mathcal{P}$ is sometimes referred to as the number of *branches* of the map f. A full branch map $f : I \to I$ is *piecewise affine* if it's derivative is constant on the interior of each $\omega \in \mathcal{P}$. In this case we shall write f'_ω to denote the derivative of f on $int(\omega)$.

Proposition 1.5.1. *Let f be a piecewise affine full branch map. Then Lebesgue measure in invariant.*

Proof. The simplest examples of full branch maps are given by $f(x) = kx \mod 1$ for some $k \in \mathbb{Z}$ with $k \geq 2$. For these maps, it is easy to see that any subinterval J has exactly k preimages, each one of length $|J|/k$. Thus the total length of the preimage $f^{-1}(J)$ is the same length as J and this proves that Lebesgue measure is invariant. 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. □

Exercise 8. Show that the maps $f(x) = kx \mod 1$ for $k \geq 2$ have at least $k-1$ invariant measures besides Lebesgue. Show that for $k = 10$ there are at least countably many distinct invariant measures. Show that the same is true also for other values of k.

1.6 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 on $[0, 1]$ by defining, for every interval $A = (a, b)$,

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

Proposition 1.6.1. *Let f be the Gauss map. Then μ_G is invariant.*

Proof. 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

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

□

1.7 Ulam-von Neumann maps

Let $I = [-1, 1]$ and define the *Ulam-von Neumann* map $f : I \to I$ by

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

This is also a full branch map with partition $\mathcal{P} = \{(-1, 0), (0, 1)\}$. However we shall see below that the existence of a *critical point*, i.e. a point c for which $f'(c) = 0$, creates additional complications in general. Define a measure μ_{UN} on I by

$$\mu_{UN}(A) = \frac{2}{\pi}\int_a^b \frac{1}{\sqrt{4 - x^2}}$$

for any interval $A \subseteq I$.

Proposition 1.7.1. *Let f be the Ulam-von Neumann map. Then μ_{UN} is invariant.*

We will use here a very different argument from that used in the construction of the invariant measure for the Gauss map. For this we need to introduce a new and very important idea in dynamical systems.

Definition 3. Let X, Y be two 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 $h \circ f = g \circ h$..

Exercise 9. Show that a conjugacy h maps orbits of f to orbits of g in the sense that for every i and every $x \in X$ we have $h(f^i(x)) = g^i(h(x))$.

If two maps f and g are conjugate then they are "equivalent" in some sense.

Exercise 10. Show that conjugacy is an equivalence relation on the space of all maps.

If the sets X, Y have some additional structure which is preserved by the maps f, g then we can define stronger forms of conjugacy.

Definition 4. Suppose $f : X \to X$ and $g : Y \to Y$ are conjugate by a conjugacy $h : X \to Y$.

1. If X, Y are measurable spaces (i.e. equipped with σ-algebras) and f, g, h, h^{-1} are all measurable, then we say that f, g are *measurably conjugate* or that h is a *measurable conjugacy.*

2. If X, Y are topological spaces and f, g, h, h^{-1} are all continuous (in particular h is a homeomorphism), then we say that f, g are *topologically conjugate* or that h is a *topological conjugacy.*

3. If X, Y are differentiable manifolds and f, g, h, h^{-1} are all differentiable (in particular h is a diffeomorphism), then we say that f, g are *differentiably conjugate* or that h is a *differentiable conjugacy.*

It is quite possible for two differentiable maps to be measurably conjugate but not topologically conjugate or topologically conjugate but not differentiably conjugate. Different kinds of conjugacy preserve different aspects of the structure of the systems.

Exercise 11. Let $\omega(x) = \{x' \in X : f^{t_n}(x) \to x' \text{ for some sequence } t_n \to \infty\}$, also called the "omega-limit" of x, denote the set of all accumulation points of the forward orbit of the point x. Show that if h is a topological conjugacy, then $h(\omega(x)) = \omega(h(x))$.

Measurable conjugacies map sigma-algebras to sigma-algebras and therefore we can generalize the notion of a push-forward of measures as in (1.1) and define a map

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

by

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

The difference here is that this maps measures defined on the space X to measures defined on the different space Y. More importantly for your purposes is that this induces a well defined map on the corresponding spaces of invariant measures.

Lemma 1.7.1. *Suppose h is a measurable conjugacy. Then the push-forward map h_* maps invariant measures to invariant measures.*

Proof. Let $\mu_X \in \mathcal{M}_f(X)$, i.e. μ_X is invariant under f. We show that $\mu_Y := h_*\mu_X \in \mathcal{M}_g(Y)$, i.e. μ_Y is invariant under g. For any measurable set $A \subseteq Y$ we have

$$\begin{aligned}\mu_Y(g^{-1}(A)) &= \mu_X(h^{-1}(g^{-1}(A))) = \mu_X((h^{-1}\circ g^{-1})(A)) = \mu_X((g\circ h)^{-1}(A)) \\ &= \mu_X((h\circ f)^{-1}(A)) = \mu_X(f^{-1}(h^{-1}(A))) = \mu_X(h^{-1}(A) = \mu_Y(A).\end{aligned}$$

□

Proof of Proposition 1.7.1. Consider the *tent map* $T : [0,1] \to [0,1]$ is defined by

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

The map T is a piecewise affine full branch map and thus Lebesgue measure m on $[0,1]$ is an invariant probability measure for T. Now define the map $h : [0,1] \to [-2,2]$ by $h(z) = 2\cos \pi z$.

Exercise 12. Show that h is a conjugacy between the tent map T and the Ulam-von Neumann map f.

Notice that h is continuously differentiable and thus is in particular measurable. Therefore we can define a measure $\mu := h_* m$. By Lemma 1.7.1 the measure μ is an invariant probability measure which is invariant under f. It just remains to show that μ has the epxlicit form required. Notice first of all that we have

$$h^{-1}(x) = \frac{1}{\pi}\cos^{-1}\left(\frac{x}{2}\right).$$

and

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

For an interval $A = (a,b)$ we have, using the fundamental theorem of calculus, factor 2 ?

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

□

Chapter 2

Ergodic measures

Let X be a measure space and $f : X \to X$ a measurable transformation.

Definition 5. A measure $\mu \in \mathcal{M}$ (not necessarily invariant) is *ergodic* if every measurable set A such that $f^{-1}(A) = A$ satisfies either $\mu(A) = 0$ or $\mu(A) = 1$.

Sometimes we also say that the map f is ergodic with respect to the measure μ. If a set A satisfies $f^{-1}(A) = A$ we sometimes say that is is *fully invariant*. We remark that this condition is much stronger than the forward invariance condition $f(A) = A$.

Exercise 13. Show that $f^{-1}(A) = A$ implies both $f(A) = A$ and $f^{-1}(A^c) = A^c$ where $A^c = X \setminus A$ denotes the complement of A, and therefore also $f(A^c) = A^c$ Show that $f(A) = A$ by itself does not imply $f(A^c) = A^c$.

To motivate the concept of ergodicity we start by presenting an abstract result, Birkhoff's Ergodic Theorem which can be thought of as a qualitative strengthening of Poincare's recurrence theorem. We then talk about the space of ergodic measures and finally present a series of examples.

2.1 Birkhoff's Ergodic Theorem

Let $f : X \to X$ be a measurable map preserving an ergodic probability measure μ and let $\varphi : X \to \mathbb{R}$ be a μ integrable function. We sometimes say that φ is an *observable* since it can be thought of as giving the result of a "measurement" which depends on the point x of the phase space at which it is evaluated. The integral

$$\int \varphi d\mu$$

is sometimes referred to as the *space average* of φ (with respect to the measure μ) whereas, for a given point x, the averages

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

are often referred to as the *time averages* of φ along the orbit of x. There is no a priori reason for these two quantities to be related. However we have the following fundamental

Theorem. *If μ is an ergodic invariant probability measure, then*

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

as $n \to \infty$ for μ almost every x.

We can formulate this result informally by saying that when μ is ergodic the *the time averages converge to the space average.* We shall not prove this result here but we shall discuss various techniques for proving ergodicity of invariant measures and analyse certain consequences. The following Corollary justifies the idea of Birkhoff's Ergodic Theorem being a strong qualitative version of Poincare's Recurrence Theorem.

Corollary 2.1.1. *Let $f : X \to X$ and μ an ergodic invariant probability measure. Then, for any measurable set $A \subseteq X$ and μ almost every x we have*

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

as $n \to \infty$.

Proof. Letting $\varphi = \mathbb{1}_A$ be the characteristic function of the set A and applying Birkhoff's ergodic theorem we get that for μ almost every x we have every $x \in M$,

$$\frac{\#\{1 \leq j \leq n : f^j(x) \in A\}}{n} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_A(f^i(x)) \to \int \mathbb{1}_A d\mu = \mu(A). \tag{2.1}$$

□

2.2 The space of ergodic invariant measures

We have seen in the previous section that the space $\mathcal{M}_f$ of invariant probability measures is convex. We now state and proof a result about the structure of the subset of ergodic invariant probability measures.

Proposition 2.2.1. *$\mu \in \mathcal{M}_f$ is ergodic if and only if it is an extremal point of $\mathcal{M}_f$.*

We recall that an extremal point of a convex set A is a point x such that if $x = tx_0 + (1-t)x_1$ for $x_0, x_1 \in \mathcal{M}_f$ with $x_0 \neq x_1$ then $t = 0$ or $t = 1$. As an immediate corollary we get the following

Corollary 2.2.1. *Suppose X is a compact metric space and $f : X \to X$ is continuous. Then there exists at least one ergodic measure.*

Proof. By standard results of Analysis, $\mathcal{M}_f$ satisfies the assumptions of the the Krein-Milman Theorem which says that a convex set is the convex hull of its extremal points. In particular, if the set is non-empty then the set of extremal points is also non-empty. If X is compact and $f : X \to X$ is continuous, the set $\mathcal{M}_f$ is non-empty and the statement follows. □

To prove Proposition 2.2.1 we need to introduce some important notions which allow us to compare two measures in different ways.

Definition 6. Let μ_1, μ_2 be probability measures. We say that μ_1 is *absolutely continuous* with respect to μ_2, and write $\mu_1 \ll \mu_2$ if $\mu_2(A) = 0$ implies $\mu_1(A) = 0$ for every measurable set A. If $\mu_1 \ll \mu_2$ and $\mu_2 \ll \mu_1$ then we say that μ_1 and μ_2 are *equivalent*.

By the Radon-Nykodym Theorem $\mu_1 \ll \mu_2$ if and only if there exists a function $\varphi \in L^1(\mu_2)$ such that

$$\mu_1(A) = \int_A \varphi d\mu_2 \tag{2.2}$$

for any measurable set A. We say that φ is the *density* of μ_1 with respect to μ_2 and we sometimes write $\varphi = d\mu_1/d\mu_2$.

Exercise 14. Suppose $\mu_1 \ll \mu_2$. Show that for any measurable set A:

$$\mu_2(A) = 1 \Longrightarrow \mu_1(A) = 1 \tag{2.3}$$

and

$$\mu_1(A) > 0 \Longrightarrow \mu_2(A) > 0. \tag{2.4}$$

Definition 7. μ_1, μ_2 are *mutually singular* if there exists A such that $\mu_1(A) = 1$ and $\mu_2(A) = 0$.

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. However, if μ_1, μ_2 are *ergodic* invariant measures then we have the following

Lemma 2.2.1. *Let μ_1, μ_2 be distinct ergodic invariant measures. Then μ_1 and μ_2 are mutually singular.*

Proof. We start by proving that μ_1 cannot be absolutely continuous with respect to μ_2. Suppose by contradiction that $\mu_1 \ll \mu_2$, we will show that $\mu_1 = \mu_2$ contradicting the assumption that they are distinct. Let φ be an arbitrary bounded measurable function (and thus in particular integrable with respect to any invariant probability measure). Then, by Birkhoff's ergodic theorem the time averages of φ converge to $\int \varphi d\mu_2$ on a set A with $\mu_2(A) = 1$. Since $\mu_1 \ll \mu_2$ we have from (2.3) that $\mu_1(A) = 1$. Thus the time averages of φ converge to $\int \varphi d\mu_2$ for μ_1 a.e. x. However, applying Birkhoff's ergodic theorem again to μ_1, the time averages of φ converge to $\int \varphi d\mu_1$ for μ_1 a.e. x. It follows that $\int \varphi d\mu_1 = \int \varphi d\mu_2$

for any bounded measurable function. This includes in particular characteristic functions and so $\mu_1(A) = \mu_2(A)$ for any measurable set and so $\mu_1 = \mu_2$.

We now prove that μ_1 and μ_2 are mutually singular. Since we have just shown that they are not absolutely continuous, there exists a measurable set E such that $\mu_1(E) > 0$ and $\mu_2(E) = 0$. Define the set

$$\hat{E} = \bigcap_{m=0}^{\infty} \bigcup_{j=m}^{\infty} f^{-j}(E).$$

We will show that $\mu_1(\hat{E}) = 1$ and $\mu_2(\hat{E}) = 0$ implying that μ_1 and μ_2 are mutually singular. Notice first of all that $f^{-1}(\hat{E}) = \hat{E}$ since $\hat{x} \in \hat{E}$ if and only if there exists some point $x \in E$ such that $\hat{x} \in f^{-j}(x)$ for infinitely many values of j. If $\hat{x}$ satisfies this property than so do all its preimages. Therefore it is sufficient to show that $\mu_1(\hat{E}) > 0$ to imply $\mu_1(\hat{E}) = 1$ using the ergodicity of μ_1. By the invariance of both measures we have

$$\mu_1\left(\bigcup_{j=0}^{\infty} f^{-j}(E)\right) \geq \mu_1(E) > 0 \quad \text{and } \mu_2\left(\bigcup_{j=0}^{\infty} f^{-j}(E)\right) = 0.$$

Moreover

$$\bigcup_{j=m}^{\infty} f^{-j}(E) = f^{-j}\left(\bigcup_{j=0}^{\infty} f^{-j}(E)\right)$$

and therefore, by the invariance of the measures we have

$$\mu_i\left(\bigcup_{j=m}^{\infty} f^{-j}(E)\right) = \mu_i\left(f^{-j}\left(\bigcup_{j=0}^{\infty} f^{-j}(E)\right)\right) = \mu_i\left(\bigcup_{j=0}^{\infty} f^{-j}(E)\right)$$

for $i = 1, 2$. In particular, the measure of each $\cup_{j=m}^{\infty} f^{-j}(E)$ is constant. Moreover the sets $\cup_{j=m}^{\infty} f^{-j}(E)$ are nested. Thus $\hat{E}$ is a countable intersection of a nested sequence of sets all of which have the same measure. It follows that $\mu_1(\hat{E}) > 0$ and $\mu_2(\hat{E}) = 0$ as required. □

Proof of Proposition 2.2.1. Suppose first that μ is ergodic. 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 $t \in (0, 1)$.

Exercise 15. Show that μ_1, μ_2 are both ergodic and that $\mu_1 \ll \mu$ and $\mu_2 \ll \mu$.

Then by Lemma 2.2.1 this implies $\mu_1 = \mu = \mu_2$ contradicting our assumptions. Thus if μ is ergodic it is an extremal point of $\mathcal{M}_f$.

Now suppose that μ is not ergodic, we will show that it cannot be an extremal point. Indeed, then there exists a set A with $f^{-1}(A) = A$, $f^{-1}(A^c) = A^c$ and $\mu(A) \in (0, 1)$. Define $\mu_1(B) = \mu(B \cap A)/\mu(A)$ and $\mu_2(B) = \mu(B \cap A^c)/\mu(A^c)$. Both μ_1 and μ_2 are clearly probability measures and $\mu = \mu(A)\mu_1 + \mu(A^c)\mu_2$. Therefore it just remains to show that they are invariant. For a measurable set B we have, using the invariance of A,

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

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

2.3 Non-ergodic measures

We start by giving some simple examples of non-ergodic measures.

2.3.1 Non-ergodic piecewise affine maps

Proposition 2.3.1. *Let $f : [0,1] \to [0,1]$ be given by*

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

Then Lebesgue measure is invariant bot not ergodic.

Proof. Exercise. □

2.3.2 Rational circle rotations

Proposition 2.3.2. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be the circle rotation $f(x) = x + \alpha$ with $\alpha = p/q$ rational. Then Lebesgue measure is invariant but not ergodic.*

Proof. The invariance of Lebesgue measure follows from Proposition 1.4.1. To show that Lebesgue is not ergodic, suppose first that $\alpha = 0$. Then $f(x) = x$ is just the identity map. It is clear then that any subset $A \subset \mathbb{S}^1$ satisfies $f^{-1}(A) = A$. Therefore it is sufficient to choose some A with $m(A) \in (0,1)$ to contradict the definition of ergodicity. The more general case of α rational is almost the same.

Exercise 16. Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be the circle rotation $f(x) = x + \alpha$ with $\alpha = p/q$ rational. Show that every point x is periodic of period q.

It is then sufficient to choose an arbitrary set A_0 of small measure $m(A_0) = \epsilon > 0$. We then let $A_j = f^j(A_0)$ and $A = A_0 \cup \cdots \cup A_{q-1}$. Then A is necessarily a union of periodic orbits and thus satisfies $f^{-1}(A) = A$. By choosing ϵ sufficiently small (for example $\epsilon < 1/q$) we can guarantee that $m(A) \in (0,1)$. This proves that Lebesgue is not ergodic. □

2.4 Fixed and periodic orbits

Proposition 2.4.1. *Let $f : X \to X$, $P = \{p_1, \ldots, p_n\}$ a periodic orbit, and δ_P the Dirac measure uniformly distributed on points of P. Then δ_P is ergodic.*

Proof. Exercise. □

2.5 Circle rotations

Proposition 2.5.1. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be the circle rotation $f(x) = x + \alpha$ with α irrational. Then Lebesgue measure is invariant and ergodic.*

We shall need a preliminary combinatorial result about irrational circle rotations. The proof is not difficult but it is quite involved and thus we shall not prove it here. A version can be found in [**?**].

Lemma 2.5.1. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be the circle rotation $f(x) = x+\alpha$ with α irrational. Then for every $x \in \mathbb{S}^1$ there exists a family of arc neighbourhoods J_n of x and with $m(J_n) \to 0$ as $n \to \infty$ and an integer q_n such that $\mathbb{S}^1 \subseteq \cup_{i=0}^{q_n} f^i(J_n)$, and each $x \in \mathbb{S}^1$ is contained in at most three intervals $f^i(J_n)$ with $i = 0, \dots, q_n$. In particular we have*

$$\sum_{i=0}^{q_n} m(f^i(J_n)) \leq 3. \tag{2.5}$$

Proof of Proposition 2.5.1. Let $A \subseteq \mathbb{S}^1$ satisfy $f^{-1}(A) = A$ and $m(A) > 0$. We want to show that $m(A) = 1$. By Lebesgue's density theorem, m almost every point of A is a Lebesgue density point of A. Thus let $x \in A$ be one such Lebesgue density point. Now fix arbitrary $\epsilon > 0$ and choose $n = n_\epsilon$ large enough so that

$$m(A \cap J_{n_\epsilon}) \geq (1-\epsilon) m(J_{n_\epsilon}) \tag{2.6}$$

where J_{n_ϵ} is a sufficiently small arc neighbourhood of x given by Lemma 2.5.1. We shall make three simple statements which combined will give us the desired result. First of all notice that (2.6) is equivalent to

$$\frac{m(J_{n_\epsilon} \setminus A)}{m(J_{n_\epsilon})} \leq \epsilon. \tag{2.7}$$

Secondly, since f is just a translation and Lebesgue measure is invariant by translation we have $m(f^i(J_n)) = m(J_n)$ for any i and, and using also the fact that A is invariant we have $m(f^i(J_{n_\epsilon} \setminus A)) = m(f^i(J_{n_\epsilon} \setminus A))$ for any i. In particular this gives

$$\frac{m(f^i(J_{n_\epsilon} \setminus A))}{m(f^i(J_{n_\epsilon}))} = \frac{m(J_{n_\epsilon} \setminus A)}{m(J_{n_\epsilon})} \tag{2.8}$$

Thirdly, using the invariance of A and the fact that $\cup_{i=0}^{q_n} f^i(J_{n_\epsilon})$ covers $\mathbb{S}^1$ we have

$$m(\mathbb{S}^1 \setminus A) \leq \sum_{i=0}^{q_n} m(f^i(J_n \setminus A)). \tag{2.9}$$

Now, from (2.7), (2.8), (2.9) and (2.5) we get

$$\begin{aligned} m(\mathbb{S}^1 \setminus A) &\leq \sum_{i=0}^{q_n} m(f^i(J_n \setminus A)) \leq \sum_{i=0}^{q_n} \frac{m(J_{n_\epsilon} \setminus A)}{m(J_{n_\epsilon})} m(f^i(J_{n_\epsilon})) \\ &= \frac{m(J_{n_\epsilon} \setminus A)}{m(J_{n_\epsilon})} \sum_{i=0}^{q_n} m(f^i(J_{n_\epsilon})) \leq 3 \frac{m(J_{n_\epsilon} \setminus A)}{m(J_{n_\epsilon})} \leq 3\epsilon. \end{aligned}$$

Since ϵ is arbitrary this means that $m(\mathbb{S}^1 \setminus A) = 0$ and thus $m(A) = 1$. □

2.6 Piecewise affine full branch maps

Proposition 2.6.1. *Let $f : I \to I$ be a piecewise affine full branch map. Then Lebesgue measure is invariant and ergodic.*

We first need to make to show some simple properties which apply to more general full branch maps.

Lemma 2.6.1. *Let $f : I \to I$ be a full branch map. Then there exists a family of partitions $\mathcal{P}^{(n)}$ of $I(\mod 0)$ into subintervals, such that for each $\omega^{(n)} \in \mathcal{P}^{(n)}$ the map $f^n : |_{int(\omega^{(n)})} : int(\omega^{(n)}) \to int(I)$ is a C^2 diffeomorphism.*

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. Then the required property follows immediately by the definition. 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} to the entire interval I and therefore $\omega^{(n-1)}$ can be subdivided into disjoint subintervals each of which maps bijectively to 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 $\mathcal{P}^{(n)}$. □

Lemma 2.6.2. *Let $f : I \to I$ be a full branch map. Suppose that there exists a constant $\lambda > 0$ such that for all $\omega \in \mathcal{P}$ and all $x \in \omega$ we have $|f'(x)| \geq e^\lambda > 1$. Then $\max\{|\omega^{(n)}| : \omega^{(n)} \in \mathcal{P}^{(n)}\} \to 0$ as $n \to \infty$.*

Proof. The proof uses a simple but very important argument in one-dimensional dynamics which relies on two basic results of calculus. The first one is the *chain rule* for differentiation. Recall that the chain rule for differentiating the composition of two functions is $(f \circ g)'(x) = f'(g(x)) \cdot g'(x)$. If we are composing a function f with itself, i.e. if $f = g$, this gives $((f^2)'(x) = f'(f(x)) \cdot f'(x)$. Composing several times then gives

$$(f^n)'(x) = f'(f^{n-1}(x))f'(f^{n-2}(x)) \cdots f'(x). \tag{2.10}$$

Thus *the derivative of f^n at a point x is the product of the derivatives of f along the orbit of the point x.* Applying this result to the maps under consideration, we have that for any x whose orbit is well defined for the first n iterations, i.e. such that for each $i = 0, ..., n-1$ the iterate $f^i(x)$ belongs to some element of the partition $\mathcal{P}$, we have

$$|(f^n)'(x)| \geq e^{\lambda n}. \tag{2.11}$$

The second important result of calculus we need is the *mean value theorem.* Recall that the mean value theorem says that if I, J are two intervals and $F : J \to I$ is differentiable and $F(J) = I$, then there exists $\xi \in J$ such that $|I| = |F'(\xi)| \cdot |J|$. Applying this to the maps under consideration we then have that for every $n \geq 1$ and every $\omega^{(n)} \in \mathcal{P}^{(n)}$ we get $|I| = |(f^n)'(x)| \cdot |\omega^{(n)}|$. Since $|I|$ is the length of the original domain of definition of the full branch map and is therefore fixed, rearranging and using (2.11) we get

$$|\omega^{(n)}| \leq e^{-\lambda n}|I|.$$

This clearly proves the Lemma. □

Proof of Proposition 2.6.1. 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 slope stricly larger than 1 uniformly in ω. Thus it satisfies the expansivity assumptions of Lemma 2.6.2. Therefore, by Lemma 2.6.1 we have a family of partition $\mathcal{P}^{(n)}$ each of which covers Lebesgue almost every point of I, and by Lemma 2.6.2 the diameter of the elements of $\mathcal{P}^{(n)}$ are decreasing uniformly in n. Now, since A has positive Lebesgue measure, 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 $m(\omega_n \cap A) \geq (1 - \epsilon)m(\omega_n)$ or, equivalently,

$$\frac{|\omega_n \setminus A|}{|\omega_n|} \leq \epsilon$$

Since the map $f^n : \omega_n \to I$ is an bijection and since $f^{-1}(A) = A$ implies $f^{-n}(A) = A$ we have $f^n(A \cap \omega_n) = f^n(f^{-n}(A) \cap \omega_n) = A \cap f^n(\omega_n) = A \cap I = A$ and therefore

$$f^n(\omega_n \setminus A) = I \setminus A.$$

Moreover, since $f^n : \omega_n \to I$ is an *affine* bijection it preserves ratios of measures of sets and therefore we have

$$\frac{|I \setminus A|}{|I|} = \frac{|f^n(\omega_n \setminus A)|}{|f^n(\omega_n)|} = \frac{|\omega_n \setminus A|}{|\omega_n|} \leq \epsilon. \tag{2.12}$$

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

Remark 1. 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 2.6.2, and in the last equality of (2.14). 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.

2.6.1 Application: Normal numbers

The relatively simple result on the invariance and ergodicity of Lebesgue measure for piecewise affine full branch maps has a remarkable application on the theory of numbers.

For any number $x \in [0,1]$ and any integer $k \geq 2$ we can write

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

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

$$x = 0.x_1 x_2 x_3 \dots$$

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 8. A number $x \in [0,1]$ is called *normal (in base k)* if its expansion $x = 0.x_1 x_2 x_3 \dots$ in base k contains asymptotically equal proportions of all digits, i.e. if for every $j = 0, \dots, k-1$ we have that

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

as $n \to \infty$.

Exercise 17. 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 3. *There exists set $\mathcal{N} \subset [0,1]$ with $m(\mathcal{N}) = 1$ such that every $x \in \mathcal{N}$ is normal in every base $k \geq 2$.*

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

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

We therefore fix some $k \geq 2$ and consider the set $\mathcal{N}_k$ of points which are normal in base k. The crucial observation is that the base k expansion of the number x is closely related to its orbit under the map f_k. Indeed, consider the intervals $A_j = [j/k, (j+1)/k)$ for $j = 0, \dots, k-1$. Then, the base k expansion $x = 0.x_1 x_2 x_3 \dots$ 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. □

2.7 Full branch maps with bounded distortion

Definition 9. We say that a full branch map has the bounded distortion property if there exists a constant $\mathcal{D}$ such that

$$Dist(f^n, \omega^n) := \sup_{x,y\in\omega^n} \left\{ \log \left| \frac{Df^n(x)}{Df^n(y)} \right| \right\} \leq \mathcal{D}$$

for every $n \geq 1$ and every $\omega^{(n)} \in \mathcal{P}^{(n)}$.

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

Proposition 2.7.1. *Let $f : I \to I$ be a full branch map satisfying the bounded distortion property. Then Lebesgue measure is ergodic.*

We remark that Lebesgue measure is not generally invariant if f is not piecewise affine. However the notion of ergodicity still holds and the ergodicity of Lebesgue measure implies the ergodicity of some natural invariant measures also in these general settings.

Lemma 2.7.1. *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. □

To prove Proposition 2.7.1, we start by showing that bounded distortion is sufficient to recover essentially the same properties as in the piecewise affine case. 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 2.7.2. *Let $\mathcal{D} = \mathcal{D}(f,^n J)$ be the distortion of f 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

$$\frac{|f^n(J')|}{|f^n(J)|}\frac{|J'|}{|J|} = \frac{|f^n(J')|/|J'|}{|f^n(J)|/|J|} = \frac{|Df^n(x)|}{|Df^n(y)|} \tag{2.13}$$

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 (2.13) gives

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

and rearranging gives the result. □

Lemma 2.7.3. *Let $f : I \to I$ be a full branch map satisfying 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 2.6.1 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 2.7.2 we have

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

Then

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

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

Proof of Proposition 2.7.1. The proof is almost identical to the piecewise affine case. In fact, the only difference is when we get to equation (2.14) where we now use the bounded distortion to get

$$\frac{|I \setminus A|}{|I|} = \frac{|f^n(\omega_n \setminus A)|}{|f^n(\omega_n)|} \leq e^{\mathcal{D}}\frac{|\omega_n \setminus A|}{|\omega_n|} \leq e^{\mathcal{D}}\epsilon. \tag{2.14}$$

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

2.8 Uniformly expanding full branch maps

The bounded distortion condition is at first sight a very strong conditions and not immediately verifiable in specific systems. We show here that in fact it follows from a very straightforward expansivity condition.

Definition 10. Let f be a full branch map. We say that 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

$$|(f^n)'(x)| \geq Ce^{\lambda n}.$$

Proposition 2.8.1. *Let f be a full branch map. Suppose that f is uniformly expanding and that there exists a constant $\mathcal{K} > 0$ such that*

$$\sup_{\omega\in\mathcal{P}} \sup_{x,y\in\omega} \frac{|f''(x)|}{|f'(y)|^2} \leq \mathcal{K}. \tag{2.15}$$

Then for every $n \geq 1$ and every $\omega^{(n)} \in \mathcal{P}^{(n)}$ and every $x, y \in \omega^{(n)}$ we have

$$\log \frac{|Df^n(x)|}{|Df^n(y)|} \leq \mathcal{K}|f^n(x) - f^n(y)|. \tag{2.16}$$

Thus f satisfies the bounded distortion property and thus Lebesgue measure is ergodic.

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

Lemma 2.8.1. *Let f be a full branch map satisfying* (2.15)*. Then, for all $\omega \in \mathcal{P}$, $x, y \in \omega$ we have*

$$\left|\frac{f'(x)}{f'(y)} - 1\right| \leq \mathcal{K}|f(x) - f(y)|. \tag{2.17}$$

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

$$|f'(x) - f'(y)| = \frac{|f''(\xi_2)|}{|f'(\xi_1)|}|f(x) - f(y)|. \tag{2.18}$$

Assumption (2.15) 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 (2.18) therefore gives $|f'(x) - f'(y)| = \mathcal{K}|f'(y)||f(x) - f(y)|$ and dividing through by $|f'(y)|$ gives the result. □

Lemma 2.8.2. *Let f be a full branch map satisfying* (2.17)*. Then, for any $n \geq 1$ and $\omega^{(n)} \in \mathcal{P}_n$ we have*

$$Dist(f^n, \omega^{(n)}) \leq \mathcal{K} \sum_{i=1}^{n} |f^i(x) - f^i(y)| \tag{2.19}$$

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

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

□

Lemma 2.8.3. *Let f be a uniformly expanding full branch map. Then there exists a constant $\tilde{\mathcal{K}}$ depending only on C, λ, such that for all $n \ge 1$, $\omega^{(n)} \in \mathcal{P}_n$ and $x, y \in \omega^{(n)}$ we have*

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

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

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

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

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

□

Proof of Proposition 2.8.1. Combining the above Lemmas we get (2.16) which clearly implies the bounded distortion property and thus, by Proposition 2.7.1, Lebesgue measure is ergodic. □

2.9 Gauss map

We now return to the Gauss map $f(x) = 1/x \mod 1$ and its invariant Gauss measure μ_G defined above.

Proposition 2.9.1. *Let $f : I \to I$ be the Gauss map. Then the measure μ_G is invariant and ergodic.*

Lemma 2.9.1. *The Gauss map is uniformly expanding*

Proof. Exercise □

Lemma 2.9.2. *Let $f : I \to I$ be the Gauss map. Then*

$$\sup_{\omega\in\mathcal{P}} \sup_{x,y\in\omega} \frac{|f''(x)|}{|f'(y)|^2} \leq 16.$$

Proof. Since $f(x) = x^{-1}$ we have $f'(x) = -x^{-2}$ and $f''(x) = 2x^{-3}$. Notice that both first and second derivatives are monotone decreasing, i.e. take on larger values close to 0. Thus, for a generic interval $\omega = (1/(n+1), 1/n)$ of the partition $\mathcal{P}$ we have $|f''(x)| \leq f''(1/(n+1)) = 2(n+1)^3$ 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. □

Proof of Proposition 2.9.1. From Lemmas 2.9.1 and 2.9.2 we have that f satisfies the assumptions of Proposition 2.8.1 and thus Lebesgue measure is ergodic for the Gauss map f. Since the Gauss measure $\mu_G \ll m$ ergodicity of μ_G then follows from Lemma 2.7.1. □

2.10 Maps with critical points

We restrict our attention here to the Ulam-von Neumann map $f(x) = x^2 - 2$ defined above and μ_{UN} the invariant measure defined in (**??**).

Proposition 2.10.1. *The measure μ_{UN} is ergodic.*

Lemma 2.10.1. *Let $(X$ and $(Y$ be two measure spaces, $f : X \to X$ and $g : Y \to Y$ two measurable maps, and $h : X \to Y$ a measurable conjugacy between f and g. Suppose that μ is invariant and ergodic for f. Then its pushforward measure $h_*\mu$ is invariant and ergodic for g.*

Proof. Exercise. □

Proof of Proposition 2.10.1. Recall from the of the invariance of μ that μ is defined as the push forward of Lebesgue measure by the conjugacy h between f and the tent map. Since Lebesgue measure is ergodic for the tent map (since the tent map is an example of a full branch piecewise affine map) it follows from lemma 2.10.1 that μ is also ergodic. □

2.11 Uncountably many non-atomic ergodic measures

We have shown above that under mild assumptions, such as the compactness of the space X and the continuity of the map $f : X \to X$ there are guaranteed to exist ergodic measures and that these are characterized as extremal points of the set of invariant measures. We have also seen that the identity map has an uncountable number of ergodic measures, namely the Dirac measure on each (fixed) point x. We now show a class of examples which also have an uncountable number of ergodic measures which however themselves live on uncountable sets.

Definition 11. A measure μ is called *atomic* if there exist a point p with $\mu(p) > 0$. Otherwise it is called *non-atomic*.

Example 2. Dirac measures and any combinations of Dirac measures are of course atomic. Lebesgue measure and continuous measures of the form μ_φ are non-atomic. Combinations of Dirac measures and continuous measures such as $\mu = (\delta_p + m)/2$ are of course atomic since they do give positive measure to some point.

Proposition 2.11.1. *The interval map $f(x) = 2x \mod 1$ admits an uncountable family of non-atomic, mutually singular, ergodic measures.*

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

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

Lemma 2.11.1. *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 18. Show that h_p is a conjugacy between f and f_p.

Exercise 19. 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 2.6.1 Lebesgue measure is ergodic and invariant for f_p and so it follows from Lemma 2.10.1 that μ_p is ergodic and invariant for f.

Exercise 20. Show that μ_p is non-atomic.

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

Lemma 2.11.2. *The measures in the family $\{\mu_p\}_{p\in(0,1)}$ are all mutually singular.*

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

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

and

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

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

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

Moreover, since the conjugacy preserves the symbolic coding we have

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

Thus, by the definition of the pushforward measure

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

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

□

Remark 2. 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.

Chapter 3

Physical measures

In this section we shall always suppose that the ambient space is a Riemannian manifold and we generally call the normalized Rimennain volume by Lebesgue measure and denote it by m. Birkhoff's ergodic theorem provides a very powerful tool for describing the dynamics of system although it's conclusions clearly depend in a fundamental way on the specific invariant ergodic measure.

Definition 12. For a map $f : X \to X$ we define the *basin of attraction* of a measure μ by

$$\mathscr{B}(\mu) := \left\{ x : \frac{1}{n} \sum_{i=0}^{n-1} \delta_{f^i(x)} \to \mu \text{ in the weak-star topology.} \right\}.$$

We say that μ is a *physical measure* if

$$m(\mathscr{B}(\mu)) > 0.$$

Remark 3. From the definition of weak-star convergence of measures, an equivalent definition of the basin is

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

So the basin is the set of points whose orbits is asymptotically sufficiently uniformly distributed in relation to μ. A priori of course there is no reason why this should should be even non-empty. However Birkhoff's ergodic Theorem implies that if μ is an ergodic invariant measure it is always true that

$$\mu(\mathscr{B}(\mu)) = 1$$

and thus in particular $B(\mu) \neq \emptyset$. As we have seen, however, a full μ-measure set of points may not be a very big set at all if, for example, μ is a Dirac measure on some fixed point. In this case, Birkhoff's ergodic theorem may not give very much useful information. In principle, we think of Lebesgue measure m as the given reference measure with respect to which we want to describe the dynamics, i.e. we would like to have information about sets of initial conditions which are large with respect to Lebesgue measure. The notion of a physical measure formalizes this idea.

3.1 Basic physical measures

There are two basic examples of physical measures: Dirac measures and absolutely continuous measures. Let $f : M \to M$ be a measurable map.

Proposition 3.1.1. *Suppose p is an asymptotically stable fixed point. Then the Dirac measure δ_p on p is a physical measure.*

Proof. Exercise. Recall that an asymptotically stable fixed point is a point p which has a neighbourhood $\mathcal{U}$ such that $f^n(x) \to p$ as $nt \to \infty$ for all $x \in \mathcal{U}$. □

Proposition 3.1.2. *Suppose $\mu \ll m$ is an ergodic invariant probability measure. Then μ is a physical measure.*

Proof. Birkhoff's ergodic Theorem implies $\mu(\mathscr{B}(\mu)) = 1$ and therefore the absolute continuity of μ with respect to m implies that $m(\mathscr{B}(\mu)) > 0$ (for $m(\mathscr{B}(\mu)) = 0$ would imply $\mu(\mathscr{B}(\mu)) = 0$). Therefore μ is a physical measure. □

3.2 Strange physical measures

Physical measures can be somewhat counterintuitive. Consider $f : I \to I$ where

$$f(x) = x + x^2 \mod 1.$$

Proposition 3.2.1 (Pianigiani, '80)**.** *Lebesgue almost every x in I has an orbit which is dense in I and the time averages converge to δ_0. In particular δ_0 is a physical measure.*

We will not prove this proposition here. However notice that f is actually a full branch map with two branches. It is very similar to the map $g(x) = 2x \mod 1$ and in fact it can be easily proved, using the arguments used in Section 2.11, that it is topologically conjugate to g. This immediately implies for example that almost all points do not converge asymptotically to the origin, since such points would necessarily have a symbolic sequence ending in 0's and this is just a countable set. Nevertheless, the proposition states that the *time averages* do converge to the Dirac measure at 0. This means that the set of points whose symbolic sequence has asymptotically a proportion of 0's tending to 1 does have full measure. The crucial factor here is that f is not uniformly expanding. Indeed, the derivative is given by $f'(x) = 1 + 2x$ and so in particular the fixed point at 0 satisfies $f'(0) = 1$. This is sometimes called a *neutral* fixed point. Points close to such a neutral fixed point are repelled but only very slowly and thus they end up spending a long proportion of time near the neutral fixed point.

However this is not the only situation in which this can occur. There are also parameters in the quadratic family where the physical measure is a Dirac measure on a *hyperbolic repelling* fixed or periodic point. In this case, nearby points still move away exponentially fast but it just so happens that the map has a combinatorial structure that re-injects points very quickly very close to the point, and this most orbits end up spending most time near such fixed or periodic points.

3.3 Non-existence of physical measures

As we have seen above, any continuous map on a compact space admits at least one ergodic invariant measure. However the existence of physical measures is not at all guaranteed.

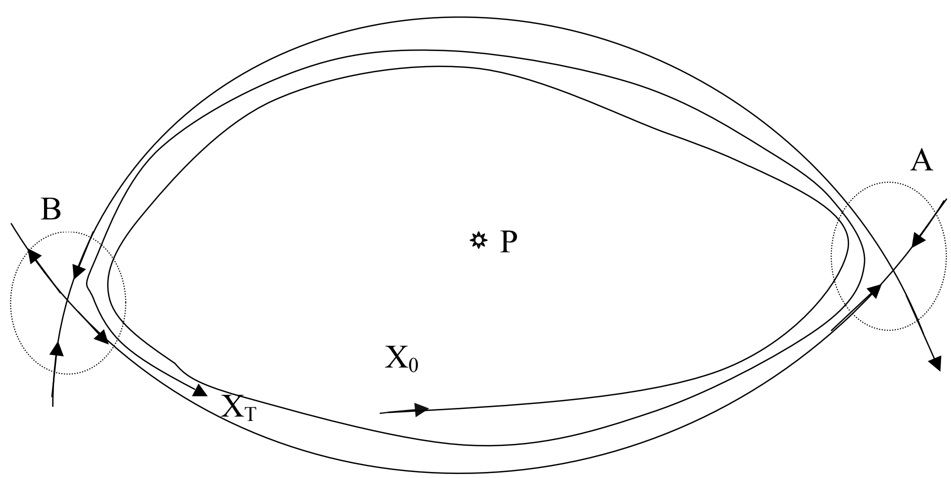
3.3.1 Circle rotations

Lemma 3.3.1. *Let $f : \mathbb{S}^1 \to \mathbb{S}^1$ be the circle rotation $f(x) = x + \alpha$ with $\alpha = p/q$ rational. Then f admits no physical measures.*

Proof. Exercise. □

3.3.2 Heteroclinic cycles

Consider a diffeomorphism $f : \mathbb{R}^2 \to \mathbb{R}^2$. Suppose that f has two hyperbolic fixed points p_A, p_B whose separatrices connect the two points defining a closed topological disk as in the following picture.



Suppose that area enclosed by the fixed points p_A, p_B and the separatrices contains a fixed point P which is repelling and that all trajectories spiral away from P and accumulate on the boundary of the disk. Under suitable conditions on the eigenvalues of the points p_A and p_B the situation is the following.

Consider some initial condition x_0 sufficiently close to the boundary of the disk. Then after some time it will enter a small neighbourhood of one of the fixed points, let's say p_A. Since p_A is a fixed point the orbit of x_0 sill spend many iterates in the neighbourhood A of p_A. If x_0 was sufficiently close to the boundary of the disk then this time will be large compared to the time it took for x_0 to reach A. The time that the orbit spends in A can be roughly estimated using the eigenvalues of the derivative at the fixed point A. Indeed let Df_{p_A} have eigenvalues $0 < e^{-\lambda} < 1 < e^{\sigma}$ and let us assume for simplicity that the dynamics is linear in the neighbourhood A of p_A. Then, in the appropriate coordinates, the dynamics is given by

$$f^n(x, y) = (e^{-\lambda n}x, e^{\sigma n}y)$$

Thus, supposing that this neighbourhood has radius ≈ 1 and that when it enters the neighbourhood A the orbit of A lies at a distance ε from the boundary we have that $(x, y) \approx (1, \varepsilon)$. Therefore the time it takes to leave the neighbourhood is a solution to the equation $\varepsilon e^{\sigma N} \approx 1$ which gives

$$N = \frac{1}{\sigma} \log \frac{1}{\varepsilon}.$$

At this moment, the new distance of the point from the separatrix is given by

$$e^{\lambda N} = e^{\frac{\lambda}{\sigma} \log \varepsilon^{-1}} = \varepsilon^{-\frac{\lambda}{\sigma}}$$

Using similar calculations it is possible to calculate the time that the orbit spends in the neighbourhood B of p_B and show that this is strictly large than the time spent in p_A and even of the whole time spent from the beginning. The points then comes back to A and spends even longer there, longer than the whole time spent so far from the beginning. Continuing in this way it is possible to show that the Dirac averages *do not converge* and thus the system *has no physical measure.*

Exercise 21. Complete the calculations to show that under appropriate conditions on the eigenvalues of p_A and p_B the systems described above has no physical measures.

3.4 The Palis conjecture

Notwithstanding the existence of examples of systems with no physical measures there is an expectation that these are very exceptional cases and that "most" systems should admit some physical measures. At the other extreme there are also systems with an infinite number of physical measures but these are also expected to be in some sense exceptional cases. This expectation has been formalized in the following conjecture by Pails.

Conjecture 1. *Most systems have a finite number of physical measures whose basins have full Lebesgue measure.*

Of course, the actual definition of what is meant by "most" systems is somewhat flexible. At the moment the conjecture is completely open and can be said to have been resolved only for an extremely limited class of families of one dimensional maps.

Theorem 4. *Consider the family $f_a(x) = x^\ell - a$ for ℓ even and $a \in [0, 2]$. Then for Lebesgue almost every parameter, f_a has a unique physical measure [Bruin-Shen-van Strien]. For the case $\ell = 2$ this measure is either a Dirac measure on a periodic orbit of an absolutely continuous measure with respect to Lebesgue [Lyubich 2002].*

3.5 Physical measures for full branch maps

Theorem 5. *Let $f : I \to I$ be a uniformly expanding full branch map satisfying the bounded distortion property* (2.16)*. 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 22. For each $n \geq 1$ we have $\mu_n \ll m$. *Hint:* by definition f is a C^2 diffeomorphism on (the interior of) each element of the partition $\mathcal{P}$ and thus in particular it is non-singular in the sense that $m(A) = 0$ implies $m(f^{-1}(A) = 0$ for any measurable set A.

Since $\mu_n \ll m$ we can let

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

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

Proposition 3.5.1. *There exists a constant $K > 0$ such that*

$$0 < \inf_{n,x} H_n(x) \leq \sup_{n,x} H_n(x) \leq K \tag{3.1}$$

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

$$|H_n(x) - H_n(y)| \leq K|H_n(x)|d(x, y) \leq K^2 d(x, y). \tag{3.2}$$

Proof of Theorem assuming Proposition 3.5.1. 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 (3.1) and (3.2). We define the measure μ by defining, for every measurable set A,

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

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

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

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

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

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

Remark 4. The fact that $\mu_n \ll m$ for every n does not imply that $\mu \ll m$. Indeed, consider the following example. Suppose $f : [0,1] \to [0,1]$ is given by $f(x) = x/2$. We alreeady know that in this case the only physical measure is the Dirac measure at the unique attracting fixed point at 0. In this simple setting we can see directly that $\mu_n \to \delta_0$ where μ_n are the averages defined above. In fact we shall show that stronger statement that $f_*^n m \to \delta_p$ as $n \to \infty$. Indeed, let $\mu_0 = m$. And consider the measure $\mu_1 = f_* m$ which is give by definition by $\mu_1(A) = \mu_0(f^{-1}(A))$. Then it is easy to see that $\mu_1([0,1/2]) = \mu_0(f^{-1}([0.1/2])) = \mu_0([0,1]) = 1$. Thus the measure μ_1 is completely concentrated on the interval $[0,1/2]$. Similarly, it is easy to see that $\mu_n([0,1/2^n]) = \mu_0([0,1]) = 1$ and thus the measure μ_n is completely concetrated on the interval $[0.1/2^n]$. Thus the measures μ_n are concentrated on increasingly smaller neighbourhood of the origin 0. This clearly implies that they are converging in the weak star topology to the Dirac measure at 0.

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

It just remains to prove Proposition 3.5.1. We start by finding an explicit formula for the functions H_n.

Lemma 3.5.1. *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

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

Therefore

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

□

Lemma 3.5.2. *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 (2.16). 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 (2.16) gives

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

Inverting x, y we also have

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

Combining these two bounds we get

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

For $x, y \in I$ we have

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

□

Proof of Proposition 3.5.1. This Lemma clearly implies the Proposition since

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

□

3.6 Induced full branch maps

Full branch maps satisfying the bounded distortion property are an important class of maps but also relatively restricted class. Nevertheless it turns out that they can be obtained in many very general classes of maps by the procedure of *inducing*.

Definition 13. $f : I \to I$ admits an *induced full branch map with bounded distortion and integrable return times* if there exists a subinterval $\Delta \subset I$, a partition $\mathcal{P}$ (mod 0) of Δ into subintervals, a return time function $R : \Delta \to \mathbb{N}$ piecewise constant on each element of $\mathcal{P}$, such that the induced map $F : \Delta \to \Delta$ defined by $F(x) = f^{R(x)}$ is a full branch map satisfying the bounded distortion property, and $\int R dm < \infty$.

Theorem 6. *Suppose that $f : I \to I$ admits an induced full branch map with the bounded distortion property and integrable return times. Then f admits an ergodic invariant absolutely continuous probability measure.*

By the assumption on f there exists an induced map $F : \Delta \to \Delta$ which is full branch and has the bounded distortion property. It therefore admits a unique ergodic invariant absolutely continuous probability measure $\hat{\mu}$. We use this measure to define a new measure by defining, for any measurable set $A \subseteq I$,

$$\mu(A) = \sum_{\omega \in \mathcal{P}} \sum_{j=0}^{R(\omega)-1} \hat{\mu}(f^{-j}(A) \cap \omega).$$

Lemma 3.6.1. *μ is absolutely continuous.*

Proof. We have $m(A) = 0 \Rightarrow \hat{\mu}(A) = 0$ by the absolute continuity of $\hat{\mu}$, then $\hat{\mu}(A) = 0 \Rightarrow \hat{\mu}(f^{-j}A) = 0$ by the invariance of $\hat{\mu}$. Therefore, clealry $\hat{\mu}(f^{-j}(A) \cap \omega) = 0$ for every ω and so $\mu(A) = 0$. □

Lemma 3.6.2. *μ is ergodic.*

Proof. Suppose that $f^{-1}(A) = A$ and $\mu(A) > 0$. We will show that $\mu(A^c) = 0$. The we must have $\hat{\mu}(f^{-j}(A) \cap \omega) > 0$ for some $j \geq 0$ and some $\omega \in \mathcal{P}$. But then, by the backward invariance of A this means that $\hat{\mu}(A \cap \omega) > 0$ and therefore, by the ergodicity of $\hat{\mu}$ this implies that $\hat{\mu}(A) = 1$. In particular $\hat{\mu}(A^c) = 0$ and therefore $\mu(A^c) = 0$. □

Lemma 3.6.3. *μ is f-invariant.*

Proof. By definition $f^{\tau(\omega)}(\omega) = \Delta$ for any $\omega \in \mathcal{P}$. In particular, using the invariance of $\hat{\mu}$ under F, this gives

$$\sum_{\omega \in \mathcal{P}} \hat{\mu}(f^{-\tau(\omega)}(A) \cap \omega) = \sum_{\omega \in \mathcal{P}} \hat{\mu}(F^{-1}(A) \cap \omega) = \hat{\mu}(F^{-1}(A)) = \hat{\mu}(A).$$

Using this equality we get, for any measurable set $A \subseteq I$,

$$\begin{aligned}
\mu(f^{-1}(A)) &= \sum_{\omega\in\mathcal{P}} \sum_{j=0}^{\tau(\omega)-1} \hat{\mu}_\omega(f^{-(j+1)}(A)) \\
&= \sum_{\omega\in\mathcal{P}} \sum_{j=0}^{\tau(\omega)-1} \hat{\mu}(f^{-(j+1)}(A) \cap \omega) \\
&= \sum_{\omega\in\mathcal{P}} \hat{\mu}\left[(f^{-1}(A) \cap \omega) + \cdots + (f^{-\tau(\omega)}(A) \cap \omega)\right] \\
&= \sum_{\omega\in\mathcal{P}} \sum_{j=1}^{\tau(\omega)-1} \hat{\mu}(f^{-j}(A) \cap \omega) + \sum_{\omega\in\mathcal{P}} (f^{-R(\omega)}(A) \cap \omega) \\
&= \sum_{\omega\in\mathcal{P}} \sum_{j=1}^{\tau(\omega)-1} \hat{\mu}(f^{-j}(A) \cap \omega) + \hat{\mu}(A) \\
&= \sum_{\omega\in\mathcal{P}} \sum_{j=0}^{R(\omega)-1} \hat{\mu}(f^{-j}(A) \cap \omega) \\
&= \mu(A).
\end{aligned}$$

□

Lemma 3.6.4. *μ is a finite measure.*

Proof.

$$\mu(I) := \sum_{\omega\in\mathcal{P}} \sum_{j=0}^{R(\omega)-1} \hat{\mu}(f^{-j}(I) \cap \omega) = \sum_{\omega\in\mathcal{P}} \sum_{j=0}^{\tau(\omega)-1} \hat{\mu}(I \cap \omega) = \sum_{\omega\in\mathcal{P}} R(\omega)\hat{\mu}(\omega) = \int R d\hat{\mu} < \infty.$$

□

Proof of Theorem 6. By the finiteness of the measure μ we can define a new measure $\nu = \mu/\mu(I)$. Then ν is clearly a probability measure and inherits the properties of absolute continuity, invariance and ergodiciyt from μ. □

Appendix 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 [?, ?, ?, ?]. 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.1.1 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.1.2 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 5. 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.1.3 Algebras and sigma-algebras

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

Definition 14. 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.1.4 Measures

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

Definition 15. 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 6. 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 16. 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 7. 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.2 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 17 (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 8. 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 18 (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 3. 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.3 Lebesgue density theorem

Theorem 7 (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*

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

as $\epsilon \to 0$.

Points x satisfying (A.1) 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.