preliminary version : Not for diffusion

SCUOLA NORMALE SUPERIORE
Classe di Scienze

AN INTRODUCTION TO DYNAMICAL SYSTEMS

by

Carlo Carminati and Stefano Marmi

PRELIMINARY VERSION : Nov 25, 2008

Preface

The material treated in this book was brought together for introductory courses taught at the Scuola Normale Superiore since 2001. It is intended to be an introduction to dynamical systems. There are so many excellent introductions to dynamical systems that one needs to justify the write–up of these notes. Our goal here is not at all to be complete but just to introduce some of the main ideas motivating the study of dynamical systems and their ubiquitous presence in contemporary mathematics as well as to lead the reader in understanding some of the main conjectures. These lectures contain many problems (some of which may challenge the reader) : they should be considered as an essential part of the text. The proof of many useful and important facts is left as an exercise.

Pisa, January 15, 2009

Carlo Carminati and Stefano Marmi

Table of Contents

Part I. Basic Concepts

Lecture 1. Dynamical systems and group actions. Orbits. Structural stability

The most general definition of *dynamical system* is the action of a group (or a semigroup) G on some space X, called the *phase space*. We denote $\mathrm{End}\,(X)$ the space of maps $f : X \to X$ which preserve the structure of X.

Definition 1.1 *A (semi–)group G acts on a space X if there exists a map $\varphi : G \times X \to X$ such that :*

(i) For all $g \in G$ the map $\varphi_g : X \to X$, $\varphi_g(x) = \varphi(g, x)$, where $x \in X$, belongs to $\mathrm{End}\,(X)$; if G is a Lie group we will also require φ to depend smoothly on g ;

(ii) If e denotes the neutral element of G, $\varphi_e = \mathrm{id}_X$;

(iii) For all $g_1, g_2 \in G$, $\varphi_{g_1 g_2} = \varphi_{g_1} \circ \varphi_{g_2}$.

This includes a huge number of possibilities. The two most important examples correspond to X being a Hausdorff topological space or X being a measure space. In the first case $\mathrm{End}\,(X)$ is the space of continous maps on X whereas in the second it is the space of measure–preserving maps on X. Other possibilities are not only conceivable but actually occur very frequently.

Roughly speaking, the goal of the theory of dynamical systems is to understand *most* of the dynamics of *most* systems. As we will see later understanding what is the meaning of "most" in the previous sentence is already doing some very important progress.

In looking for the simplest cases we are led to consider actions of $\mathbb{N}$, $\mathbb{Z}$ (discrete time systems) or $\mathbb{R}, \mathbb{C}$ (continuous time systems) and to ask for the lowest possible dimension of the ambient space together with the highest possible regularity of the action. In the first case one has a diffeomorphism f of a smooth compact manifold X whereas in the continuous time case one considers the flow associated to a vector field v, i.e. the one–parameter group $(\varphi_t)_{t\in\mathbb{R}} \subset \mathrm{Diff}\,(X)$ of solutions of $\dot{x} = v(x), x(0) = x_0 : x(t, x_0) = \varphi_t(x_0)$. We are assuming the vector field $v : X \to TX$ to be complete, which is automatically satisfied if X is compact.

An important example of higher rank actions is given by the actions of $G = \mathbb{Z}^2$ on X : to the canonical basis $\{(1,0),(0,1)\}$ of $\mathbb{Z}^2$ one can associate two generators $f, g \in \mathcal{G} = \mathrm{Diff}\,(X)$ so $f \circ g = g \circ f$. In this case the study of the action is equivalent to the study of a pair of commuting diffeomorphisms.

The action of a group on a space X determines an equivalence relation on the space : two points $x_1, x_2 \in X$ are *equivalent* if there exists an element $g \in G$ such that $x_2 = \varphi_g(x_1)$. Two equivalent points belong to the same *orbit* $Gx = \{\varphi_g(x) \mid g \in G\}$ and the orbits of the points of X are equivalence classes. Note that one can regard each orbit also as the image of the map $\psi_x : G \to X$, $\psi_x(g) = \varphi_g(x)$. A *fixed point* of the action is an orbit consisting of a single point.

The stabilizer H_x of the action at a point $x \in X$ is the set of elements $h \in G$ which leave x fixed $\varphi_h(x) = x$. H_x is a closed subgroup of G and ψ_x induces a 1–1 immersion of G/H_x into M. An action is *effective* if and only if the global stabilizer $H_X := \cap_{x \in X} H_x = \{e\}$.

The action of G on X is *free* if for $g \in G$ and $x \in X$, $g \neq e$ implies $\varphi_g(x) \neq x$. Equivalently $H_x = \{e\}$ for all x. The action is properly *discontinuous* if each $x \in X$ has a neighborhood U such that $U \cap \varphi_g(U) = \emptyset$ for all but finitely many $g \in G$, $g \neq e$. The action is *transitive* if given any two points $x, y \in X$ there is $g \in G$ such that $\varphi_g(x) = y$.

If X is a Hausdorff topological space then the *quotient space* X/G is a topological space w.r.t. the topology induced asking that the projection $p : X \to X/G$ is continuous and open (the open subsets of X/G are the projection of the open subsets of X).

Exercise 1.2 Let X be a manifold and assume that G acts effectively on X. Show that if the action of G on X is free and properly discontinuous X/G is a differentiable manifold and the projection is a local diffeomorphism. In this case X is an *(unramified) covering space* of X/G with the *covering transformation group* G. If the action is not free but it is dicontinuous then X is a *ramified covering* of X/G and the ramifying points (in X) are nothing but the points fixed by the action (i.e. whose stabilizer is not trivial). Show that properly discontinuous actions always have finite stabilizers at all points.

Exercise 1.3 Consider the actions of $\mathbb{Z}^k$ on $\mathbb{R}^l$ by translations and find necessary and sufficient conditions for the quotient being a smooth manifold. By definition, if $k = l$, when the quotient is smooth then it is a k–dimensional torus.

The strongest possible notion of equivalence between different actions on the same space is obtained asking that they are indeed the same action "apart from a global change of coordinates on the space".

Definition 1.4 *Two actions φ, φ' of the same group G on X are* conjugate *if there is an isomorphism $h : X \to X$ such that $\varphi_g h = h\varphi'_g$ for all $g \in G$. If all the actions sufficiently close to a given one (w.r.t. a sensible topology) are conjugate to it we say that the given action is* rigid.

In the above definition, if X is a topological space then its isomorphisms are just the homeomorphisms of X so h is required to be a homeomorphism. When X is a differentiable manifold then one can require h to be a diffeomorphism of X (same regularity as X). If X is a measure space then its isomorphisms are measure–preserving maps whose inverse is defined modulo zero–measure sets and it is also measure–preserving.

1.1 Discrete Dynamical Systems. Conjugation, Symmetries.

When $G = \mathbb{Z}$ one has a *discrete dynamical system* and the space of all possible dynamical systems on X is just the group $\mathcal{G} = \mathrm{Iso}\,(X)$ of isomorphisms of X. One can also consider actions of the semigroup $G = \mathbb{N}$, i.e. a homomorphism of G in the semigroup $\mathcal{G} = \mathrm{End}\,(X)$ of endomorphisms of X : in this case φ_g is not necessarily invertible.

Let f be a generator of the action (i.e. $f \in \mathcal{G}$ and $\varphi_n = f^n$ for all $n \in G$, where f^n denotes the composition of f with itself n times if $n > 0$, $f^0 = \mathrm{id}_X$ is the identity, $f^{-n} = (f^{-1})^n = f^{-1} \circ \ldots \circ f^{-1}$). *From now on we will omit the symbol $\circ$ for the composition of two transformations* (unless some confusion may be possible). The *orbit* $\mathcal{O}_f(x)$ of $x \in X$ relative to f is the subset $\{f^n(x) \mid n \in G\}$. The finite orbits are called *periodic orbits* or *cycles* and their points *periodic points*. The *period* m of a periodic point x is the smallest positive integer such that $f^m(x) = x$. If $m = 1$ x is a *fixed point*. Note that all periodic points of f are fixed points of some iterate of f. A subset Y of X is *invariant* if $f(Y) = Y$.

Definition 1.5 *Let $f \in \mathcal{G}$. We say that g is* equivalent *or* conjugate *to f if there is a diffeomorphism h of X such that*

$$f \sim g \iff \exists h \in \mathcal{G} \,:\, g = h^{-1} f h \,.$$

Remark 1.6 Sometimes one may compare also dynamical systems defined on two different phase spaces. If $f : X \to X$, $g : Y \to Y$ and $h : X \to Y$ is such that $gh = hf$ then conjugacy corresponds to h being a diffeomorphism. When h is an *embedding* we say that f is a *subsystem* of g. When h is a *submersion*, f is an *extension* of g and g is a *factor* of f.

If $g : Y \to Y$ is surjective but not injective, it is possible to define a *natural extension* of g : it is possible to find a space X and an isomorphism $f : X \to X$ such that $gh = hf$. One way to do this construction is setting $X = \{(x_k)_{k\in\mathbb{Z}} \in Y^{\mathbb{N}} : f(x_k) = x_{k+1}\}$ and defining h the projection on the 0-th coordinate : h is surjective because g is. There are often other (more direct) methods to define an extension : see for instance (xxx)

Let X be a compact manifold. If $0 \le r \le \infty$ the space $\mathcal{G}^r$ of $\mathcal{C}^r$ discrete dynamical systems on X with the $\mathcal{C}^r$ topology is a Baire space (any countable intersection of open dense sets is dense) and if $r < \infty$ it is even a Banach manifold. Similarly, if the *parameter space* P is also a compact manifold then the space of $\mathcal{C}^r$ families of discrete dynamical systems $F : P \times X \to X$ is also a Baire space. A *generic property* is a property that is true of dynamical systems belonging to some Baire set of $\mathcal{G}$ (a Baire set is the intersection of a countable number of open dense sets). When we deal with parametrized families we can also consider a different notion of typical : a property is *Lebesgue typical* if it is satisfied for maps corresponding to a full Lebesgue measure in the parameter space $P = \mathbb{R}^k$.

The set of diffeomorphisms equivalent to f obviously forms an equivalence class, the *orbit* of f under the adjoint action of Diff (X) :

$$[f] = \mathrm{Ad}_{\mathrm{Diff}\,(X)} f = \{g \in \mathcal{G}\,,\ \exists h \in \mathrm{Diff}\,(X)\ : g = \mathrm{Ad}_h f = h^{-1} f h\}\ .$$

In some cases one can achieve a *complete* classification of the conjugacy classes [PY] but this is in general a totally unrealistic goal, as the following example shows.

Example 1.7 Let $\mathbb{C}\{z\}$ denote the ring of convergent power series in one variable. The germs of analytic diffeomorphisms of $\mathbb{C}$ with a fixed point at 0 are the elements of the group $\{f \in z\mathbb{C}\{z\}\,, f'(0) \neq 0\}$. We say that two such germs f and g are conjugate if there is another germ h such that $fh = hg$. Note that if f and g are conjugate then they have the same derivative $\lambda = f'(0)$ at the fixed point 0. If $|\lambda| \neq 1$ then a classical result of Poincaré and Koenigs says that λ is the only conjugacy invariant, i.e. all germs with the same derivative at the fixed point are conjugate. When $|\lambda| = 1$ the situation changes dramatically. Even in this apparently very simple context (a local problem in the smallest possible dimension with the highest possible regularity of the dynamics) a complete classification of the conjugacy classes of germs is open and perhaps unreasonable as the following result of Yoccoz [Y1] shows :

Theorem 1.8 *There is a generic set of values of $\lambda \in \mathbb{S}^1$ for which there exists a set with the power of the continuum of different conjugacy classes of germs all with linear part λ and each of which class does not contain an entire function.*

The proof of the previous result goes much beyond the scope of this introduction. We refer to [Y1] for its complete proof.

Exercise 1.9 Prove the Poincaré–Koenigs Theorem. [Hint : note that it is sufficient to show that if $|\lambda| \neq 1$ then all f with $f'(0) = \lambda$ are conjugate to the linear map $R_\lambda(z) = \lambda z$. Then one can apply either the contraction principle to the functional equation $f(h(z)) = h(\lambda z)$ or to prove directly the convergence of the Taylor expansion of h obtained by recurrence matching powers on the two sides of the functional equation. The latter is the classical method exposed also in [Mar2].]

If we drop the requirement in Definition 1.5 that h is a diffeomorphism and we only ask h to be a homeomorphism then f and g are *topologically conjugate* : in Lecture 2 we will study various invariants by topological conjugacy. Usually differentiable conjugacy is too a fine classification : indeed in order to be differentiably conjugated two dynamical systems must have the same derivative at all corresponding cycles. The notion of stability associated to topological conjugacy is called *structural stability* and is due to Andronov and Pontrjagin [AP].

Definition 1.10 *The set $SS(X)$ of* structurally stable discrete dynamical systems on X *consists of all systems f in $\mathcal{G}$ such that there is an open neighborhood $\mathcal{N}$ of f such that if $g \in \mathcal{N}$ then f and g are topologically conjugate.*

Clearly structurally stable dynamical systems form an open subset of $\mathcal{G}$. Its complement sometimes is also called the bifurcation set.

Definition 1.11 *A diffeomorphism g is a* symmetry *of $f \in \mathcal{G}$ if $g \in Cent\,(f)$, i.e. if $Ad_g f = f$.*

Exercise 1.12 Assume $g \sim f$ (i.e. $f = h^{-1}gh$) and show that
(1) $\mathrm{Cent}\,(f)$ is conjugated to $\mathrm{Cent}\,(g)$, i.e. $\mathrm{Cent}\,(f) = h^{-1}\mathrm{Cent}\,(g)h$;
(2) if $f(x) = x$ then $g(h(x)) = h(x)$ and $f'(x) = \lambda$ is invariant under conjugation : $g'(h(x)) = \lambda$;
(3) $f^{\mathbb{Z}} = \{f^n\,,\ n \in \mathbb{Z}\} \subset \mathrm{Cent}\,(f)$.

1.2 Flows vs. discrete dynamical systems. Cross–sections and suspensions

The relationship between discrete time dynamical systems and continuous time dynamical systems is very rich. Usually in dynamical systems one prefers to consider the discrete time case not just because of its great natural appeal but also because "the same phenomena and problems of the qualitative theory of ordinary differential equations are present in their simplest form in the diffeomorphism problem. Having first found theorems in the diffeomorphism case, it is usually a secondary tas to translate the results back into the differential equations framework" (Smale, [Sm] p. 747). Moreover many differential equations have *cross–sections*, i.e. codimension 1 submanifolds Y of the phase space X which are transversal to the flow and on which the original problem of dynamics reduces to the study of the iteration of a diffeomorphism of Y. This method was first used by Poincaré more than a century ago in his study of the three body problem of celestial mechanics.

More recently another important fact has put emphasis on the study of discrete time dynamical systems : the massive use of computers to study numerically the solutions of differential equations. Virtually all methods of numerical integration of differential equations are based on a discretization of time and replace an o.d.e. (or a p.d.e.) with the iteration of a discrete time dynamical system. One should be however aware that this discretization is far from being an innocuous artifact ! Indeed even in the simplest cases the dynamics obtained can be on the long run dramatically different from the original one. For example consider the most classical mechanical example of a *simple pendulum*

$$\ddot{x} = \sin x\ .$$

A first order finite difference scheme of time step μ transforms it into

$$x(t+\mu) - 2x(t) + x(t-\mu) = \mu^2 \sin x(t) \; .$$

Setting $\varepsilon = \mu^2$, $x = x(t)$, $x' = x(t+\mu)$, $y = x(t) - x(t-\mu)$ and $y' = x(t+\mu) - x(t)$ one obtains the map of $\mathbb{T}^1_x \times \mathbb{R}_y$

$$x' = x + y' \; , \quad y' = \varepsilon \sin x + y$$

which is the celebrated *standard map* of the theory of twist maps (see Lecture ? ? ?). Whereas all the trajectories of the simple pendulum are either periodic or connect the unstable equilibrium position with itself in infininite time (separatrices) the trajectories of the standard map, even for very small values of ε, are extremely complicated, many aperiodic orbits exist and it is even conjectured that for all values of ε (different from zero) there exists a positive measure set of chaotic trajectories (in a very well precise sense, see Lecture 11).

Most of the concepts we define for discrete time dynamical systems have a natural reformulation for flows. In general we will leave to the reader the duty of translating them into the flow framework.

One exception is the notion of equivalence of two flows where various possibilities arise since one can choose to consider equivalent two flows with the same orbits (up to homeomorphisms) preserving or not the time evolution along them.

Definition 1.13 *A flow φ'_t on X is a* time reparametrization *of another flow φ_t on X if there exists a function $\chi : \mathbb{R} \times X \to \mathbb{R}$ (called an* (untwisted) one–cocycle *over φ_t) such that :*
(i) $\varphi'_t(x) = \varphi_{\chi(t,x)}(x)$ for all $x \in X$ and for all $t \in \mathbb{R}$;
(ii) $\chi(t,x) \geq 0$ if $t \geq 0$ and either $\chi(t,x) > 0$ if $t > 0$ or $\chi(t,x) \equiv 0$ (then x is a fixed point).

¿From the group properties of the flow one sees that χ must verify the equations

$$\chi(t+s,x) = \chi(t,x) + \chi(s, \varphi'_t(x)) \; ,$$
$$\chi(-t,x) = -\chi(t, \varphi'_{-t}(x)) \; .$$

One has now to possibilities : one can consider equivalence classes of flows up to time reparametrizations (*orbit equivalence*) or not (*flow equivalence*). In the former case one says that the flow φ_t is equivalent to ψ_t if there exists a diffeomorphism $h \in \mathrm{Diff}\,(X)$ such that $h\varphi_t h^{-1}$ is a time reparametrization of ψ_t. In the latter case one requires $\psi_t = h\varphi_t h^{-1}$, i.e. the two actions of $\mathbb{R}$ are conjugate. We will study time–reparametrizations of linear flows on tori in some detail in Lecture 3.

Exercise 1.14 Show that the vector fields of $\mathbb{R}^2$ $v(x,y) = (x,y)$ and $w(x,y) = (x+y, -x+y)$ give rise to topologically conjugate flows whereas they are not conjugate to $u(x,y) = (-y,x)$. Show that the vector field $v(x) = x^2$ of $\mathbb{R}$ is not structurally stable. Is u structrurally stable ?

Let X be a smooth manifold and let $\mathcal{X}(M)$ be the set of $\mathcal{C}^1$ (or $\mathcal{C}^r$) vector fields on X equipped with the $\mathcal{C}^1$ (resp. $\mathcal{C}^r$) topology.

Definition 1.15 *A vector field $v \in \mathcal{X}(M)$ is* structurally stable *if there is an open neighborhood $\mathcal{N}$ of v in $\mathcal{X}(M)$ such that all $w \in \mathcal{N}$ are $\mathcal{C}^0$–orbit equivalent to v (i.e. the flow $\varphi_{t,v}$ is topologically conjugate to a time reparametrization of $\varphi_{t,w}$).*

Once again structurally stable vector fields form an open subset of X. Peixoto [Pe] showed that the set of structurally stable vector fields (resp. diffeomorphisms) on a compact orientable 2–dimensional manifold X (resp. of the circle $\mathbb{T}^1$) is open and dense in $\mathcal{X}^r(M)$ (resp. $\mathrm{Diff}^r(\mathbb{T}^1)$) for $1 \le r \le \infty$. Indeed they do coincide with *Morse–Smale* systems (see [Sm3]) : these are the natural generalizations of gradient flows for which the Morse inequalities still hold [Sm1]. Unfortunately in dimension ≥ 3 (for flows, resp. ≥ 2 for diffeomorphisms) structural stability is not a dense property [Sm2, Ne1]. A nice introduction to structural stability is due to Arnold [Ar4, Chapter 3].

We conclude this short survey on the relationship between flows and diffeomorphisms making the notion of cross-section more precise and showing conversely how to associate to each diffeomorphism a flow.

Definition 1.16 *A compact codimension one submanifold Y of a compact manifold X is called* a cross–section *for a flow φ_t on X if Y intersects every orbit, has transversal intersection with the flow and whenever $x \in Y$, $\varphi_t(x) \in Y$ for some $t > 0$.*

A flow φ_t induces a diffeomorphism $f \in \mathrm{Diff}\,(Y)$ by setting $f(x) = \varphi_{t_0}(x)$ where t_0 is the first positive t such that $\varphi_t(x) \in Y$. Note that there cannot be any fixed points for the flow whenever a cross–section exists. But one can have periodic orbits and they will become cycles (or fixed points) of f.

Exercise 1.17 Show that the topological equivalence of flows with cross–sections is determined by the topological equivalence of the corresponding diffeomorphisms.

Let us complete our discussion of cross–sections of flows showing how to associate to each diffeomorphism a flow.

Definition 1.18 *If f is a diffeomorphism of X the* suspension *of f is the flow φ_t on the manifold X_0 of dimension* $\dim X + 1$ *defined as follows :*

(i) X_0 is obtained as the orbit space of $X \times \mathbb{R}$ under the free action of $\mathbb{Z}$ generated by the diffeomorphism $\psi(x,u) = (f(x), u+1)$, i.e. is the quotient space $X \times \mathbb{R}/\psi^{\mathbb{Z}}$;

(ii) The flow φ_t on X_0 is given by $\psi_t([x,u]) = [x, u+t]$, where $[x,u]$ denotes the equivalence class of $(x,u) \in X \times \mathbb{R}$ w.r.t. the action of $\mathbb{Z}$ generated by ψ.

Exercise 1.19 Show that X_0 has a cross–section $Y_0 = p(X \times \{0\})$ where $p : X \times \mathbb{R} \to X_0$ is the projection on the quotient. Check that the diffeomorphism f_0 associated to the cross–section Y_0 is differentiably conjugate to f^{-1}.

Lecture 2. Topological and statistical properties of dynamics

In this lecture we will first introduce several properties of a dynamical system which are preserved by topological conjugacy. Then we will introduce the notion of invariant measure and ergodicity. When the phase space is at the same time equipped of topological and measurable structures the two approaches can be compared : this will be the object of the next Lecture.

2.1 Topological properties : Recurrence, transitivity and minimality

Let X be a compact metric space : $\mathrm{End}\,(X) = \mathcal{C}(X)$ and $\mathrm{Aut}\,(X) = \mathrm{Homeo}\,(X)$. Let $f \in \mathrm{Aut}\,(X)$ and consider the discrete time dynamical system generated by f.

Definition 2.1 *A point $x \in X$ is called a* wandering point *when there is a neighborhood U of x such that $\cup_{|n|>0} f^n(U) \cap U = \emptyset$. If $f \in End\,(X)$ we only consider $n > 0$. A point will be called* nonwandering *if it is not a wandering point.*

These nonwandering points are those with the mildest possible form of recurrence. They form a closed invariant set usually denoted as $\Omega(f)$.
Given any point x we define the α *and* ω *limit sets* of its orbit

Definition 2.2 *A point $y \in X$ is called an ω–*limit point *(respectively, an α–*limit point*) of $x \in X$ if there is a sequence $(n_j)_{j\in\mathbb{N}}$ going to $+\infty$ (respectively, to $-\infty$) such that $f^{n_j}(x) \to y$. The set of all ω–limit points of x is denoted $\omega_f(x)$ or simply $\omega(x)$ (respectively, $\alpha_f(x)$ or $\alpha(x)$).*

¿From the definition of $\alpha(x)$ and $\omega(x)$ it follows that

$$\omega(x) = \bigcap_{n>0} \left(\overline{\bigcup_{m\geq n} f^m(x)} \right) ,$$

$$\alpha(x) = \bigcap_{n<0} \left(\overline{\bigcup_{m\leq n} f^m(x)} \right) .$$

Recurrent points must verify a much stronger constraint than nonwandering points :

Definition 2.3 *A point $x \in X$ is* positively recurrent *if $x \in \omega_f(x)$,* negatively recurrent *if $x \in \alpha_f(x)$,* recurrent *if it is both positively and negatively recurrent. We will denote $Rec_f(X)$ (resp. $Rec_f^+(X)$, $Rec_f^-(X)$) the set of recurrent points (resp. positively recurrent, negatively recurrent).*

Exercise 2.4xxx Show that f is positively recurrent iff for all $\delta > 0$ $\cap_n \{f^k(x), k \geq n\} \cap B(x,\delta) \neq \emptyset$.

Exercise 2.4 Show that $\Omega(f)$, $\alpha(x)$ and $\omega(x)$ are closed and invariant and that $\Omega(f)$ contains the ω– and α–limit sets of all points. Show that since X is compact they cannot be empty.

Exercise 2.5 Give the definition of nonwandering point, α– and ω–limit sets and recurrent points for flows. Of course the properties of the above exercise are true also for flows.

Exercise 2.6 Consider the flows associated to the vector fields of $\mathbb{R}^2$ $v(x,y) = (x,-y)$ and $w(x,y) = (y + x(1-x^2-y^2), -x + y(1-x^2-y^2))$ and determine the α– and ω–limit sets of all orbits. Show that $[0,+\infty) \times \{0\}$ is a cross–section of the flow associated to w and compute the induced diffeomorphism f. [Answer : $f(x) = e^{2\pi}x/\sqrt{1-x^2+x^2e^{2\pi}}$.]

Periodic points are the simplest kind of orbits one can conceive and they give the most perfect case of recurrence. However they do not always exists. The next case of very strong and uniform recurrence is represented by *minimal sets.*

Definition 2.7 *A discrete dynamical system $f : X \to X$ is* topologically transitive *if there exists a point $x \in X$ such that its orbit $\mathcal{O}_f(x)$ is dense in X. If the orbit of every point is dense in X then f is called* minimal. *A closed non–empty f–invariant subset A of X is called a* minimal set *if $f|_A$ is minimal.*

Example 2.8 Let $X = [0,1]$, $f(x) = 2x \pmod 1$. This maps acts on the binary expansion of $x \in X$, $x = \sum_{j=1}^{\infty} a_j 2^{-j}$, $a_j \in \{0,1\}$, as a *shift map* $\sigma : f(x) = \sum_{j=1}^{\infty} a_{j+1} 2^{-j}$. Note that the space of sequences $\{0,1\}^{\mathbb{N}}$ with the topology induced by asking that $\pi : \{0,1\}^{\mathbb{N}} \to [0,1]$, $\pi(\{a_j\}) = \sum_{j=1}^{\infty} a_j 2^{-j}$, to be continuous is completely disconnected. The map π is 1 to 1 at all but countably many points (the dyadic rationals) where is 2 to 1. Thus f is a *factor* of the shift σ. The orbit of the point x whose binary expansion is 0100011011000001010011100101110111...

is dense in X, thus f is topologically transitive. On the other hand f cannot be minimal since it has infinitely many periodic orbits : e.g. 010101 . . .. The dynamics of f provides one of the simplest examples of hyperbolic (chaotic) dynamics (see Lecture 6).

Topological transitivity gives some sort of topological indecomposability. We will later compare it to ergodicity which gives measure indecomposability. We now see some criteria for topological transitivity (see [Wa], p. 127)

Proposition 2.9 *Let $f : X \to X$ be a homeomorphism of a compact metric space X. The following are equivalent :*

(i) f is topologically transitive ;
(ii) if U is open and f–invariant then either U is empty or U is dense ;
(iii) for any two non–empty open sets U, V there exists an integer $N = N(U, V)$ such that $f^N(U) \cap V \neq \emptyset$;
(iv) the set of points of X with a dense orbit is a dense G_δ.

Proof. (i)⇒(ii). Let the orbit of $x \in X$ be dense and let $U \neq \emptyset$, U open and $f(U) = U$. There exists an integer j such that $f^j(x) \in U$, thus by the f–invariance of U the orbit of x is contained in U and U is dense.
(ii)⇒(iii). Let U, V be two open non–empty sets. Then the invariant non–empty open set $\cup_{n\in\mathbb{Z}} f^n(U)$ is necessarily dense. Thus $\cup_{n\in\mathbb{Z}} f^n(U) \cap V \neq \emptyset$ from which (iii) follows.
(iii)⇒(iv). Let $(U_n)_{n\in\mathbb{Z}}$ be a countable open basis for X. Then $\{x \in X \mid \overline{\mathcal{O}_f(x)} = X\} = \cap_{n=1}^{\infty} \cup_{m\in\mathbb{Z}} f^m(U_n)$ and by (iii) $\cup_{m\in\mathbb{Z}} f^m(U_n)$ is dense.
(iv)⇒(i). Obvious. □

Exercise 2.10 Let X be compact and $f \in \mathrm{Aut}\,(X)$. Show that the following are equivalent :

(i) f is minimal ;
(ii) X has no proper closed non–empty f–invariant subset ;
(iii) for every non–empty open subset U of X one has $\cup_{-\infty}^{+\infty} f^n(U) = X$.

Remark 2.11 If A is a minimal set and γ is an orbit contained in A then γ is recurrent. This is because $\omega(\gamma)$ is closed, nonempty, invariant and $\omega(\gamma) \subset A$. Thus $\gamma \subset A = \omega(\gamma)$. Similarly $\gamma \subset \alpha(\gamma)$.

Topological transitivity (and a fortiori minimality) has an interesting consequence on the *observables*, i.e. on continuous functions $F : X \to \mathbb{C}$.

Proposition 2.12 *If $f : X \to X$ is topologically transitive then the only f–invariant continuous functions are the constants.*

Proof. Let $F \circ f = F$. Then the same holds iterating f n times $F \circ f^n = F$ $\forall n \in \mathbb{Z}$. Pick $x \in X$ whose orbit is dense : F must be constant on the orbit. By continuity F is constant on X. □

The converse of Proposition 2.12 is in general false : see [Wa, p. 131] for a counterexample.

Exercise 2.13 Let G be a topological group and consider the action of G on itself by left translations : $L : G \times G \to G$, $L_g(g') = L(g, g') = gg'$. Show that if the discrete dynamical system on G generated by the left translation L_g by a fixed element g is topologically transitive then it is also minimal (thus in this setting being topologically transitive is equivalent to being minimal).

The example of translations on abelian groups is the prototype of quasiperiodic dynamics (see Lecture 6). The next exercise gives the most important example of quasiperiodic dynamics.

Exercise 2.13-bis Find an example of a topologically transitive dynamical system which is not minimal.

Exercise 2.14 Let $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ denote the n–dimensional torus and consider the discrete time dynamical system generated by the *translation* $R_\alpha(x) = x + \alpha \,(\mathrm{mod}\, \mathbb{Z}^n)$, where $\alpha \in \mathbb{T}^n$. Show that R_α is minimal if and only if α is *rationally independent* (or *non–resonant*) : $\alpha \cdot k + p \neq 0$ for all $(k, p) \in \mathbb{Z}^n \times \mathbb{Z}$, $(k, p) \neq (0, 0)$. In this case R_α is called an *irrational translation*. [Hint : use the previous exercise, Fourier expansions and apply Proposition 2.9 (iii) to show that the condition is sufficient, Proposition 2.12 to show that it is also necessary.]

A dynamical system need not be minimal but it always has a minimal set.

Proposition 2.15 *Every continuous map f on a compact metric space X has a minimal set.*

Proof. Let $\mathcal{A}$ denote the collection of all closed f–invariant non–empty closed subsets of X. Inclusion gives a partial ordering on $\mathcal{A}$. Every ordered chain in $\mathcal{A}$ has a non–empty least element (the intersection of the elements of the chain, which is clearly closed, non–empty and invariant). Thus by Zorn's lemma $\mathcal{A}$ has a minimum element which is by construction a minimal set. □

Denjoy [De] constructed a $\mathcal{C}^1$ vector field of the two–dimensional torus $\mathbb{T}^2$ with a nontrivial minimal set (i.e. distinct from a fixed point or a periodic orbit) distinct from $\mathbb{T}^2$ (see also Appendix 1). On the other hand one can prove [Sch] that a minimal set A of a $\mathcal{C}^2$ vector field on a two–dimensional compact surface X is either trivial or the whole of X and, in this case, $X = \mathbb{T}^2$.

The *Cherry flow* [Ch] is an example of an analytic vector field on the two–torus with highly nontrivial recurrence. It has exactly two fixed points, no closed orbit and it has a circle cross–section and the non–wandering set of the induced diffeomorphism is the union of an attracting fixed point and an invariant Cantor set (i.e. a compact, totally disconnected, nonempty perfect set). We refer to [PDM] (pp. 181–188) for its construction.

2.2 Statistical properties : Poincaré recurrence theorem

Ergodic theory is an attempt to study the statistical behaviour of orbits of dynamical systems restricting the attention to their asymptotic distribution. One waits until all transients have been wiped off and looks for an invariant probability measure describing the distribution of typical orbits.

This approach is especially fruitful for systems with a very strong sensitivity to initial conditions.

Let $(X, \mathcal{A}, \mu)$ be a probability space, i.e. X is a set, $\mathcal{A}$ is a σ–algebra of subsets of X and μ is a measure on $(X, \mathcal{A})$ such that $\mu(X) = 1$.

Definition 2.16 *A map $f : X \to X$ is* measurable *if for all $A \in \mathcal{A}$, $f^{-1}(A) \in \mathcal{A}$. A measurable transformation is* non singular *if $\mu(f^{-1}(A)) = 0$ for all $A \in \mathcal{A}$ such that $\mu(A) = 0$. A non singular measurable transformation is* measure–preserving *if for all $A \in \mathcal{A}$, $\mu(f^{-1}(A)) = \mu(A)$.*

In the above definition $f^{-1}(A) = \{x \in X \mid f(x) \in A\}$ and f need not be invertible. If f is invertible with a non singular measurable inverse and it preserves the measure μ then $\mu(f^{-1}(A)) = \mu(A) = \mu(f(A))$, $\forall\ A \in \mathcal{A}$.

Exercise 2.17 Why should one use f^{-1} instead of f in Definition 2.16 ?

Example 2.18 The Hénon map on $\mathbb{R}^2$: $f_c(x,y) = (x^2 + c - y, x)$ is a polynomial automorphism of the plane and preserves the Lebesgue measure. The Jacobian determinant of any polynomial automorphism of the plane is a constant (when this constant equals one the automorphism preserves Lebsegue measure). The *Jacobian conjecture* asserts the converse : a polynomial endomorphism of the plane with constant Jacobian is invertible.

Definition 2.19 *A* measurable dynamical system $(X, \mathcal{A}, \mu, f)$ *is the datum of a probability space* $(X, \mathcal{A}, \mu)$ *and of a measure preserving (surjective ? ? ?) map* $f : X \to X$.

If $f : X \to X$ is surjective it is possible to perform the following *natural extension* construction : consider the subset of bi-infinite sequences

$$\hat{X} = \{(x_n)_{n\in\mathbb{Z}} \ : x_n \in X, f(x_n) = x_{n+1} \ \forall n \in \mathbb{Z}\},$$

consider the shift map $\hat{f} : \hat{X} \to \hat{X}$ defined as $(\hat{f}\hat{x})(k) = x_{k+1}$ and the projection $\pi : (x_n)_{n\in\mathbb{Z}} \mapsto x_0$. Consider also $\hat{\mathcal{A}}$, the smallest σ-algebra that ensures that $\pi \circ \hat{f}^k$ is measurable for all $k \in \mathbb{Z}$: the pullback $\hat{\pi} := \pi_{\#}\mu$ defines a $\hat{f}$-invariant probability measure on $(\hat{X}, \hat{\mathcal{A}})$ and π is a measurable semiconjugacy between the shift $(\hat{X}, \sigma)$ and (X, f).

Example 2.20 The dyadic map of Example 2.8 preserves the Lebsegue measure on $[0, 1]$.

Exercise 2.21 Let $X = [0, 1]$, $f : X \to X$ non singular. Assume that f is piecewise monotonic and $\mathcal{C}^1$, i.e. there exists a finite or countable decomposition of the interval $[0, 1]$ in subintervals $[a_i, a_{i+1}]$, $i \in \mathcal{I}$, on which f si monotonic and of class $\mathcal{C}^1$ in their interior. On each of these subintervals an inverse branch f_i^{-1} of f is well–defined. Show that a measure $\mu_\rho(x) = \rho(x)dx$ is f–invariant if and only if $\rho(x) = \sum_{i\in\mathcal{I}_x} \frac{\rho(f_i^{-1}(x))}{|f'(f_i^{-1}(x))|}$, where $\mathcal{I}_x$ denotes the subset of $\mathcal{I}$ corresponding to the indices i such that $f_i^{-1}(x) \neq \emptyset$. Use this result to check that the following probability measures are invariant :

(i) Ulam–Von Neumann's map $f(x) = 4x(1 - x)$, $\rho(x)dx = \frac{dx}{\pi\sqrt{x(1-x)}}$;

(ii) p–adic map $f(x) = px(\operatorname{mod} 1)$, $\rho(x)dx = dx$;

(iii) Gauss' map $f(x) = \{\frac{1}{x}\}$ if $x \neq 0$, $f(0) = 0$, where $\{x\}$ denotes the fractional part of x, $\rho(x)dx = \frac{dx}{(1+x)\ \log 2}$.

Example 2.22 Translations on compact topological groups X : they preserve the *Haar measure* μ defined on the Borelian subsets of X. The Haar measure is the unique probability measure on the Borel σ–algebra of X which is translation invariant and regular (i.e. for all $\varepsilon > 0$ and for all $A \in \mathcal{A}$ there exists a compact subset K and an open subset U such that $K \subset A \subset U$ and $\mu(U \setminus K) < \varepsilon$. When $X = \mathbb{T}^1$ the Haar measure is just the normalized circular Lebsegue measure. When $X = \mathbb{T}^n$ it is the direct product of the measure on $\mathbb{T}^1$.

Let X be a (nonempty) compact metric space, $\mathcal{A}$ be the σ–algebra of its Borel subsets and $f \in \mathrm{Homeo}\,(X)$. Let $\mathcal{M}(X)$ denote the topological space of all probability measures on X with the usual weak–$*$ topology : $\mu_n \to \mu$ in $\mathcal{M}(X)$ if and only if for all $\varphi \in \mathcal{C}(X,\mathbb{R})$ one has $\int_X \varphi d\mu_n \to \int_X \varphi d\mu$.

Exercise 2.23 Prove that $\mathcal{M}(X)$ is compact in the weak–$*$ topology. [Hint : use *Riesz representation theorem* : if J is a continuous linear functional on $\mathcal{C}(X,\mathbb{R})$ such that $J \geq 0$ (i.e. if $\varphi \geq 0$ then $J(\varphi) \geq 0$) and $J(1) = 1$, then there exists $\mu \in \mathcal{M}(X)$ such that $J(\varphi) = \int_X \varphi d\mu$ for all $\varphi \in \mathcal{C}(X,\mathbb{R})$.]

The compactness of $\mathcal{M}(X)$ implies that there is always at least an f–invariant probability measure on X [KB] :

Theorem 2.24 (Krylov–Bogolubov) *There exists at least an f–invariant probability measure on X.*

Proof. Consider the continuous map $f^* : \mathcal{M}(X) \to \mathcal{M}(X)$ defined by $(f^*\mu)(A) = \mu(f^{-1}(A))$ for all $A \in \mathcal{B}$. A probability measure is invariant if $f^*\mu = \mu$. Take any measure $\mu_0 \in \mathcal{M}(X)$ and consider the sequence $\mu_m = \frac{1}{m}\sum_{j=0}^{m-1} f^{*m}\mu_0$. By the compactness of $\mathcal{M}(X)$ any weak limit μ of the sequence $(\mu_m)_{m\in\mathbb{N}}$ is an invariant probability measure since for all $\varphi \in \mathcal{C}(X,\mathbb{R})$ one has $|\int_X (\varphi - \varphi \circ f) d\mu_m| \leq \frac{2}{m}\|\varphi\|_\infty$. □

Definition 2.1xxx *The support* $\mathrm{supp}(\mu)$ *of a borel mesure μ is the complement of the union of open sets U such that $\mu(U) = 0$. Since X is supposed to be a compact metric space it is easy to show that* $\mathrm{supp}\mu$ *is a closed set. Let us remark that* $\mathrm{supp}(\mu) = X$ *iff any nonempty open set has strictly positive measure.*

Exercise 2.xxx Show that if μ is f-invariant, $\mathrm{supp}(\mu)$ is a closed f-invariant set ; moreover the support of μ is a subset of the nonwandering set of f : $\mathrm{supp}(\mu) \subset \Omega(f)$.

The existence of an invariant probability measure for a homeomorphism on a compact space is sufficient to show that "most" orbits are recurrent, i.e. the probability for an orbit to be recurrent is equal to one :

Theorem 2.25 (Poincaré recurrence theorem) *Let $f \in Homeo\,(X)$ preserve the probability measure μ. Then all points of X except a set E of measure zero are recurrent under f. If $X = \mathrm{supp}(\mu)$ the exceptional set E is of first category.*

We recall that a set is called first category if it can be represented as a countable union of nowhere dense sets (see [Ox], p. 2 and p. 67). Of course it can be a dense set (e.g. $\mathbb{Q}$ in $\mathbb{R}$).

Lemma 2.26 (Weak Poincaré recurrence theorem) *If f is a measure-preserving transformation of a probability space $(X, \mathcal{A}, \mu)$ then for all $A \in \mathcal{A}$ the subset A_{rec} of the points $x \in A$ such that $f^n(x) \in A$ for infinitely many $n \in \mathbb{N}$ belongs to $\mathcal{A}$ and $\mu(A) = \mu(A_{rec})$.*

Let us remark that this statement looks weaker than the previous one, but does not use neither invertibility nor the metric structure.

Proof.

$A_{rec} = A \cap [\cap_n \cup_{k \geq n} f^{-k}(A)]$, it is clear that $A_n := \cup_{k\geq n} f^{-k}(A)$ is a decreasing sequence of measurable sets such that $f^{-1}A_n = A_{n+1}$. Since $\mu(A_n) = \mu(f^{-1}A_n) = \mu(A_{n+1})$ we deduce that $Z_n := A_n \setminus A_{n+1}$ is a null set : $\mu(Z_n) = 0$. We can prove by induction that $A_{n+1} = A_0 \setminus \cup_{k=0}^{n} Z_k$ so, setting $Z = \cup_{k=0}^{+\infty} Z_k$, we get that $\cap_n A_n = A_0 \setminus Z$. On the oter hand $\mu(Z) = 0$ and $A_{rec} = A \cap (A_0 \setminus Z)$, thus

$$\mu(A_{rec}) = \mu(A \cap (A_0 \setminus Z)) = \mu(A \cap A_0) = \mu(A),$$

so $\mu(A \setminus A_{rec}) = 0$. □

Exercise 2.27 Show that if $X = \mathrm{supp}(\mu)$ then A_{rec} is a $\mathcal{G}_\delta$–set and is dense in A ; hence $A \setminus A_{rec}$ is a set of first category.

Show that if A is open then A_{rec} is a G_δ set. [Hint : the sets B_n are relatively closed in A.]

Proof of Theorem 2.25 Let $\{V_n\}_{n\in\mathbb{N}}$ be a countable basis for the topology of X such that diam $(V_n) \to 0$ as $n \to \infty$ and, for all $m \geq 0$, $\cup_{n\geq m} V_n = X$. By Lemma 2.26 one has $\mu(V_n \setminus V_{n,rec}) = 0$ for all n.
Claim : $\text{Rec}\,(X) = \cap_{m\geq 0} \cup_{n\geq m} V_{n,rec}$.

If one accepts the claim then it is immediate to conclude that $\mu(\text{Rec}\,(X)) = 1$ since

$$\begin{aligned}\mu(X \setminus \text{Rec}\,(X)) &= \mu\left(\cap_{m=0}^{\infty} \cup_{n\geq m} V_n \setminus \cap_{m=0}^{\infty} \cup_{n\geq m} V_{n,rec}\right) \\ &\leq \mu\left(\cap_{m=0}^{\infty} \cup_{n\geq m} V_n \setminus V_{n,rec}\right) = 0\ .\end{aligned}$$

For all $m \geq 0$ one has $\text{Rec}\,(X) \subset \cap_{n\geq m} V_{n,rec}$ as one can easily check. To see that $\cap_{m=0}^{\infty} \cup_{n\geq m} V_{n,rec} \subset \text{Rec}\,(X)$ we let $\delta > 0$ such that diam $(V_n) < \delta/3$ for all n large enough.

Pick a point $x \in V_{n,rec}$ for some $n \geq m$. Since m can be chosen large one will have diam $(V_n) < \delta/3$ thus

- $V_n \subset B(x, \delta)$;
- $f^j(x) \in B(x, \delta)$ if $f^j(x) \in V_n$;

which implies $d(f^j(x), x) < \delta$ for infinitely many j. Since δ is arbitrary this concludes the proof of the claim.

To see that $\text{Rec}\,(X)$ is the complement of a first category set we use Exercise 2.27 which shows that $V_n \setminus V_{n,rec}$ is a nullset of first category. Hence also $\cup_{n=1}^{\infty} V_n \setminus V_{n,rec}$ is a first category nullset. It is then immediate to check that its complement belongs to $\text{Rec}\,(X)$. □

Poincaré's Recurrence Theorem leads to (apparent) paradoxes when applied to systems relevant in statistical mechanics. If we insert a partition in a box and pump out all the air on one side of the partition then our system is in an initial state for which all the molecules are in half of the box. Remove the partition and wait long enough : almost surely all the molecules will concentrate once again in their original half of the box. But . . . the expected return time is so long to be physically irrelevant : see [Pet, pp. 35 ff]

2.3 More statistical properties : frequency of visit and ergodicity

To understand better the statistical distribution of orbits it is useful to introduce the *frequency of visit* of a (measurable) set. Among all systems, those for which the interpretation of the invariant probability measure as the frequency of visit of measurable sets by typical orbits is correct, will have a distinguished role.

Let $(X, \mathcal{A}, \mu, f)$ be a measurable dynamical system. Given a set $A \in \mathcal{A}$ we denote χ_A its characteristic function.

Definition 2.28 *The* frequency of visit $\nu(x, A)$ *of the set* A *by the orbit of* x *is the limit (when it exists)*

$$\nu(x, A) = \lim_{n\to+\infty} \frac{1}{n} \sum_{j=1}^{n-1} \chi_A(f^j(x)).$$

If $\varphi \in L^1(X, d\mu)$ *we call n-th Birkhoff sum the quantity* $S_n\varphi(x) := \sum_{j=0}^{n-1} \varphi(f^j(x))$ *: thus the frequency of visits is obtained taking the limit of means of the Birkhoff sums of* χ_A *(the characteristic function of* A*).*

Exercise 2.29 Give examples of systems where the frequency of visit is independent on the choice of the initial point x and examples where the frquency depends on the initial point x.

Definition 2.30 *A measurable dynamical system* $(X, \mathcal{A}, \mu, f)$ *is* ergodic *if for all* $A \in \mathcal{A}$ *one has* $\nu(x, A) = \mu(A)$ *for* μ*–a.e.* $x \in X$.

Exercise 2.31 Try to prove directly the following “primitive form of Birkhoff ergodic theorem : For all $A \in \mathcal{A}$ and for μ–almost every point $x \in X$ the frequency of visit $\nu(x, A)$ exists.(Hint : see [BKS])

The preceding exercise can be solved using quite elementary tools. We shall prove a slightly more general statement : we shall consider the limit of means of Birkhoff sums of a general L^1 function (not just χ_A). Before stating the result we need some theoretical tools which will be useful to characterize the limit of Birkhoff sums.

If $\varphi \in L^1(X, \mathcal{A}, \mu)$ and $\mathcal{B}$ is a subσalgebra then then $\nu_\varphi(A) := \int_A \varphi d\mu$ defines a (signed) measure on $\mathcal{B}$ which is absolutely continuous with respect to $\mu_{|\mathcal{B}}$ and, by Radon-Nikodym theorem, admits a unique density $\varphi_\mathcal{B} \in L^1(X, \mathcal{B}, \mu)$ namely the function uniquely characterized by the twoproperties

[(i)] $\varphi_\mathcal{B} \in L^1(X, \mathcal{B}, \mu)$;

[(ii)] $\nu_\varphi(A) = \int_A \varphi_\mathcal{B} d\mu \quad \forall A \in \mathcal{B}$.

This density is usually called *conditional expectation of* φ *given* B and is denoted $\varphi_\mathcal{B} = E(\varphi|\mathcal{B})$. This shows that it is well defined a linear operator

$E(\cdot|\mathcal{B}) : L^1(X, \mathcal{A}, \mu) \to L^1(X, \mathcal{B}, \mu)$ that maps $\varphi \mapsto E(\varphi, \mathcal{B})$. It is not hard to prove that the operator $E(\cdot|\mathcal{B})$ is in fact a projector and has the following properties :

(a) $E(\cdot, \mathcal{B})$ is a linear continuous operator ;

(b) If $\mathcal{B}_0 \subset \mathcal{B}_1 \subset \mathcal{A}$ then $E(\cdot, \mathcal{B}_0)E(\cdot, \mathcal{B}_1) = E(\cdot, \mathcal{B}_0)$; in particular if $\mathcal{B}_0 = \mathcal{B}_1 = \mathcal{B}$ then $[E(\cdot, \mathcal{B})]^2 = E(\cdot, \mathcal{B})$ i.e. $E(\cdot, \mathcal{B})$ is a linear projector ;

(c) $E(|\varphi|, \mathcal{B}) = |E(\cdot, \mathcal{B})|$;

(d) If Φ is a convex function then$E(\Phi \circ \varphi|\mathcal{B}) \geq \Phi \circ E(\varphi, \mathcal{B})$; in particular $E(\cdot, \mathcal{B}) : L^p(X, \mathcal{A}, \mu) \to L^p(X, \mathcal{B}, \mu)$ for all $p \in [1, +\infty]$.

(e) If f is measure preserving then $E(\varphi, \mathcal{B}) \circ f = E(\varphi \circ f, f^{-1}\mathcal{B})$ [see Billingsley xxx]

Let $\mathcal{J}$ the family of $\mathcal{A}$-measurable f-invariant sets ; it is not difficult to check that $\mathcal{J} := \{A \in \mathcal{A} \ : \ \mu(A \Delta f^{-1} A) = 0\}$ is a sub-σalgebra of $\mathcal{A}$ and $E(\varphi|\mathcal{J})$ is a well defined $\mathcal{J}$-measurable function which, by (e), is f-invariant as well. Nowwe are ready to prove

Theorem 2.32 (Birkhoff Ergodic Theorem) *Let $(X, \mathcal{A}, \mu, f)$ be a measurable dynamical system and let $\varphi \in L^1(X, \mathcal{A}, \mu)$. For μ–a.e. $x \in X$*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)) = \varphi_{\mathcal{J}}(x)$$

This limit is called the time average *of φ along the orbit of $x \in X$. If the system is ergodic then $\varphi_{\mathcal{J}}(x) = \int_X \varphi d\mu$ a.e.*

To prove the theorem we shall need the following

Lemma 2.33 *In the same hypotheses of Theorem 2.29, if $\varphi_{\mathcal{J}}(x) < 0$ for μ–a.e. $x \in X$ then $\limsup \frac{1}{n} S_n(\varphi) \leq 0$ for almost every $x \in X$.*

Proof. (of Lemma 2.33) Let $\varphi \in L^1(X, d\mu)$ be fixed, $S_0 = 0$ and $S_n\varphi(x) := \sum_{j=0}^{n-1} \varphi(f^j(x))$ $(n \geq 1)$ be the Birkhoff sums. Let us define a monotone sequence of positive functions $\Phi_n := \max\{S_k \ : \ 0 \leq k \leq n\}$ and call $A := \{x \in X \ : \ \Phi_n(x) \to +\infty\}$. Clearly

$$\forall x \in A^c \quad \limsup \frac{1}{n} S_n(x) \leq \limsup \frac{1}{n} \Phi_n(x) \leq 0.$$

Since

$$\Phi_n \circ f(x) = \max\{S_k(x) \ : \ 1 \leq k \leq n+1\} - \varphi(x)$$

we deduce easily that A is an f-invariant set. Moreover it is easy to check that

$$\Phi_{n+1}(x)-\Phi_n\circ f(x) = \{ \begin{matrix} \varphi(x) \text{if } \Phi_n\circ f(x)+\varphi(x)\geq 0 \\ \varphi(x) - \max\{S_k \;:\; 1\leq k\leq n+1\} \text{otherwise} \end{matrix},$$

and thus we deduce that

$$\Phi_{n+1}(x)-\Phi_n\circ f(x) = -\min\{0, \Phi_n\circ f(x)+\varphi(x)\}.$$

This shows that $\Phi_{n+1}(x)-\Phi_n\circ f(x)$ is a decreasing sequence which, for a.e. $x\in A$, converges to $\varphi(x)$. By f-invariance of μ and Fatou's lemme we get

$$0\leq \int_A [\Phi_{n+1}-\Phi_n]d\mu = \int_A [\Phi_{n+1}-\Phi_n\circ f] \to \int_A \varphi d\mu = \int_A \varphi_{\mathcal{J}} d\mu$$

Since $\varphi_{\mathcal{J}}<0$ μ–a.e. the condition $\int_A \varphi_{\mathcal{J}} d\mu \geq 0$ can hold only if $\mu(A)=0$: our claim is proved. □

Proof. (of Theorem 2.32) Let a be any f-invariant function such that $a>\varphi_{\mathcal{J}}$ (for instance $a=\varphi_{\mathcal{J}}+\varepsilon$, $\varepsilon>0$ is fine), we set $\varphi_a := \varphi - a$ and we have $\int_X \varphi_a d\mu < 0$; therefore the previous lemma we get $\limsup \frac{1}{n}S_n(\varphi_a) = \limsup \frac{1}{n}[S_n(\varphi)-na]\leq 0$ which implies $\limsup \frac{1}{n}S_n(\varphi)\leq a$. On the other hand, if b is an invariant function such that $b<\varphi_{\mathcal{J}}$ we set $\psi_b = b-\varphi$ and the same argument as above leads to $\liminf \frac{1}{n}S_n(\varphi)\geq b$. Since $a>\varphi_{\mathcal{J}}>b$ are arbitrary we get that $\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}\varphi(f^j(x)) = \varphi_{\mathcal{J}}(x)$ □

Exercise xxx : Prove that

$$\frac{1}{n}\sum_{j=0}^{n-1}\varphi\circ f^j \to \varphi_{\mathcal{J}}(x)$$

where the convergence is in L^1 as well.

It is interesting to list some equivalent formulations of ergodicity

Theorem 2.34 *Let $(X,\mathcal{A},\mu,f)$ a measurable dynamical system. The following properties are equivalent :*

1) *ergodicity ;*
2) measurable (or metric) indecomposabily *: for all f–invariant $A\in\mathcal{A}$ one has $\mu(A)=0$ or $\mu(A)=1$;*

3) if $\varphi \in L^1(X, \mathcal{A}, \mu)$ is invariant then φ is constant μ–a.e. ;
4) if $\varphi \in L^1(X, \mathcal{A}, \mu)$ then $\int_X \varphi d\mu = \lim_{n\to+\infty} \frac{1}{n} S_n\varphi(x)$ for μ–a.e. $x \in X$;
5) $\forall\ A, B \in \mathcal{A}$ one has $\lim_{n\to+\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mu(f^{-j}(A) \cap B) = \mu(A)\mu(B)$.

Proof. 1) ⇒ 2) Assume that there exists a measurable f–invariant set $A \in \mathcal{A}$ with $\mu(A) > 0$. Since A is invariant, for all $x \in A$ the frequqncy of visit of A is $\nu(x, A) = 1$. By ergodicity for μ–a.e. x one has $\nu(x, A) = \mu(A)$, thus $\mu(A) = 1$.
2) ⇒ 3) If $\varphi \in L^1(X, \mathcal{A}, \mu)$ is invariant, for all $\gamma \in \mathbb{R}$ the set $A_\gamma = \{x \in X\,,\ \varphi(x) \le \gamma\}$ is invariant. By 2) one gets $\mu(A_\gamma) = 0$ or $= 1$. On the other hand if $\gamma_1 < \gamma_2$ one obviously has $A_{\gamma_1} \subset A_{\gamma_2}$, thus setting $\gamma_f = \inf\{\gamma \in \mathbb{R}\,,\ \mu(A_\gamma) = 1\}$ one has $\varphi(x) = \gamma_f$ for μ–a.e. x.
3) ⇒ 4) The f–invariance of the time average $\hat\varphi$ implies that $\hat\varphi$ is μ–a.e. constant. Thus $\hat\varphi(x) = \int_X \varphi d\mu$ for μ–a.e. $x \in X$.
4) ⇒ 1) Just apply 4) to the characteristic function χ_A.
4) ⇒ 5) Let $\varphi = \chi_A$. For μ–a.e. $x \in X$ one has

$$\mu(A) = \int_X \chi_A\, d\mu = \hat\chi_A(x) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \chi_A(f^j(x))\ .$$

Thus by dominated convergence one has

$$\mu(A)\mu(B) = \int_X \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \chi_A(f^j(x))\chi_B(x)\, d\mu =$$

$$\lim_{n\to\infty} \frac{1}{n}\int_X \sum_{j=0}^{n-1} \chi_A(f^j(x))\chi_B(x)\, d\mu = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \mu(f^{-j}(A) \cap B)$$

5) ⇒ 2) Let A be invariant. Applying 5) to $B = X \setminus A$

$$\mu(A)\mu(X \setminus A) = \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n} \mu(f^{-j}(A) \cap (X \setminus A)) = 0$$

because A is f–invariant. Thus either $\mu(A) = 0$ or $\mu(X \setminus A) = 0$. □

Example 2.35 Ergodicity of irrational translations on the torus $\mathbb{T}^n$. Let φ be an invariant function. Then if one writes its Fourier series expansion $\varphi(x) = \sum_{k\in\mathbb{Z}^n} \hat\varphi(k)e^{2\pi i k\cdot x}$ the invariance condition implies

$$\hat\varphi(k)(e^{2\pi i k\cdot\alpha} - 1) = 0 \quad \text{for all}\, k \in \mathbb{Z}^n\ ,$$

and by the non resonance condition one obtains $\hat{\varphi}(k) = 0$ for all $k \neq 0$, i.e. φ is constant a.e. .

Example 2.36 Ergodicity of the p–adic maps. Let $p \in \mathbb{N}$, $p \geq 2$, and consider the map $f(x) = px \,(\mathrm{mod}\, 1)$. Clearly it preserves Lebesgue measure. If $\varphi \in L^1$ is f–invariant then using its Fourier expansion $\varphi(x) = \sum_{k \in \mathbb{Z}} \hat{\varphi}(k) e^{2\pi i k x}$ we obtain $\hat{\varphi}(k) = \hat{\varphi}(pk)$ for all $k \in \mathbb{Z}$. By the Riemann–Lebesgue lemma we have $|\hat{\varphi}(k)| \to 0$ as $k \to \infty$, thus $\hat{\varphi}(k) = 0$ for all $k \neq 0$ and φ is constant a.e. .

Exercise 2.37 Prove Borel's theorem on normal numbers : almost all numbers in $[0, 1)$ are normal to base 2, i.e. for a.e. $x \in [0, 1)$ the frequency of 1's in the binary expansion of x is 1/2. [Hint : use ergodicity of the dyadic map.] perche' non in base p con la mappa piadica e quindi in ogni base ? ? ?

Exercise 2.38 Ergodicity of hyperbolic automorphisms of $\mathbb{T}^2$. Here the two–dimensional torus $\mathbb{T}^2$ is considered multiplicatively, thus $\mathrm{Aut}\,(\mathbb{T}^2) \approx \mathrm{SL}\,(2, \mathbb{Z})$. A matrix $A \in \mathrm{SL}\,(2, \mathbb{Z})$ is *hyperbolic* if it has no eigenvalue of unit modulus. Equivalently if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $|\mathrm{Tr}\,(A)| = |a + d| > 2$. Hyperbolic automorphism of $\mathbb{T}^2$ are ergodic. (Hint : proceed as for the dyadic map.)

Exercise 2.39 Assume that the measurable dynamical system $(X, \mathcal{A}, \mu, f)$ is ergodic and let $\mu_1 : \mathcal{A} \to [0, 1]$ be another f–invariant probability measure. Show that if $\mu_1 \neq \mu$ then μ_1 is *not* absolutely continuous w.r.t. μ. [Hint : By contradiction, if it were absolutely continuous then the Radon-Nykodim derivative $\frac{d\mu_1}{d\mu}$ would be f–invariant.]

2.4 Characterization of ergodic measures. Unique ergodicity

Let X be a compact metric space, f a homeomorphism of X, and consider the compact space $\mathcal{M}(X)$ of the probability measures on the borelian subsets of X. It is immediate to check that it is a *closed subset* of the unit ball in $\mathcal{C}(X, \mathbb{R})^*$. The f–invariant probability measures on X form a *closed convex* subset of $\mathcal{M}(X)$ which we denote $\mathcal{M}_f(X)$.

Proposition 2.40 *A probability measure $\mu \in \mathcal{M}_f(X)$ is ergodic if and only if it is an extreme point of $\mathcal{M}_f(X)$.*

Proof. Let us first show by contradiction that being an extreme point is a sufficient condition for ergodicity. Indeed, if μ is not ergodic and $A \in \mathcal{A}$ is a Borel f–invariant set such that $0 < \mu(A) < 1$ then both $\mu_A = \frac{1}{\mu(A)}\mu|_A$ and $\mu_{X\setminus A} = \frac{1}{1-\mu(A)}\mu|_{X\setminus A}$ are f–invariant and $\mu = \mu(A)\mu_A + (1-\mu(A))\mu_{X\setminus A}$, i.e. μ is not an extreme point. Again by contradiction we can show that the condition is also necessary. For, if $\mu = t_1\mu_1 + t_2\mu_2$ with $\mu_1, \mu_2 \in \mathcal{M}_f(X)$, $t_1, t_2 \in (0,1)$, $t_1 + t_2 = 1$, then μ_1 is absolutely continuous w.r.t. μ and its Radon–Nikodym derivative is a non–constant f–invariant function in $L^1(X, \mathcal{A}, \mu)$, thus μ cannot be ergodic. □

Corollary 2.41 *Every homeomorphism of a compact metric space preserves at least an ergodic probability measure.*

Proof. It is an immediate consequence of Proposition 2.40, Krylov–Bogolubov's theorem and Krein–Milman's theorem (every compact convex set in a locally convex topological vector space is the closure of the convex envelope of its extreme points). □

Sometimes a dynamical system has a *unique invariant measure* μ : in this case it is called *uniquely ergodic*, since it is obviously authomatically ergodic. Moreover, if the support of μ is the whole space X then the system is called *strictly ergodic*. For uniquely ergodic systems one has this stronger formulation of Birkhoff's theorem :

Theorem 2.42 *Let $(X, \mathcal{A}, \mu, f)$ be uniquely ergodic. Then for all continuous function $\varphi : X \to \mathbb{R}$ and for all initial condition $x \in X$ the sequence of Birkhoff's averages $\frac{1}{n}\sum_{j=0}^{n-1}\varphi(f^j(x))$ converges* uniformly *to a constant independent of x. Thus the time average exists* for all initial points *$x \in X$ and equals $\int_X \varphi\, d\mu$.*

Proof. By contradiction : assume that for a continuous $\varphi : X \to \mathbb{R}$ the sequence of functions $\left(\frac{1}{n}\sum_{j=0}^{n-1}\varphi\circ f^j\right)_{n\in\mathbb{N}}$ is not uniformly convergent to $\int_X \varphi\, d\mu$ Then there exists $\varepsilon > 0$ and two sequences $(n_i)_{i\in\mathbb{N}} \subset \mathbb{N}$, $n_i \to \infty$ and $(x_i)_{i\in\mathbb{N}} \subset X$ such that for all $i \in \mathbb{N}$

$$\left|\frac{1}{n_i}\sum_{j=0}^{n-1}\varphi(f^j(x_i)) - \int_X \varphi\, d\mu\right| \geq \varepsilon\ .$$

By the compactness of the space of probability measures on X the sequence of probability measures

$$\nu_i := \frac{1}{n_i}\sum_{j=0}^{n_i-1}\delta_{f^j(x_i)}$$

converges to some measure ν. One easily checks that it is invariant : if $\psi : X \to \mathbb{R}$ is continuous then

$$\int_X \psi(f(x))\,d\nu = \lim_{i\to\infty} \int_X \psi(f(x))\,d\nu_i = \lim_{i\to\infty} \frac{1}{n_i} \sum_{j=0}^{n_i-1} \psi(f^{j+1}(x_i))$$
$$= \lim_{i\to\infty} \left[\int_X \psi(x)\,d\nu_i - \frac{1}{n_i}\psi(x_i) + \frac{1}{n_i}\psi(f^{n_i+1}(x_i)) \right].$$

The second and the third term have limit zero thus ν is invariant. Then one gets

$$\left| \int_X \varphi\,d\nu - \int_X \varphi\,d\mu \right| = \lim_{i\to\infty} \left| \int_X \varphi\,d\nu_i - \int_X \varphi\,d\mu \right| =$$
$$\lim_{i\to\infty} \left| \frac{1}{n_i} \sum_{j=0}^{n_i-1} \varphi(f^j(x_i)) - \int_X \varphi\,d\mu \right| \geq \varepsilon$$

which shows that $\nu \neq \mu$, a contradiction. □

Exercise 2.43 Show that if for all continuous functions $\varphi : X \to \mathbb{R}$ the time average exists for all initial points x and is independent of x then the system is uniquely ergodic.

Exercise 2.44 Prove that the irrational translations on tori are strictly ergodic. Use this fact to solve a famous exercise proposed by V.I. Arnol'd : does 8 appear more frequently than 7 in the sequence of the most significant digit in the powers of 2 (i.e. 1, 2, 4, 8, 1(6), 3(2), 6(4), etc.) ?

2.5 The spectral viewpoint.

Birkhoff theorem admits some variants which are obtained changing the space of "observables". The L^2-case is particularly interesting, not only because it admits a much simpler proof but also because in this setting one can easily establish close connections between dynamical properties and spectral theory. In the following $(X, \mathcal{A}, \mu, f)$ will be a measurable dynamical system and $L^2(X, d\mu) = \{\phi : X \to \mathbb{C} \ : \ \phi \text{ measurable } \int_X |\phi|^2 d\mu < +\infty\}$. It is then natural to consider the linear operator U_f defined on $L^2(X, d\mu)$ as $U_f\varphi = \phi \circ f$. It is strightforward to check that U_f is an isometry and, if f is invertible with nonsingular measurable inverse then U_f is an unitary operator and $U_f^* = U_f^{-1} = U_{f^{-1}}$.

Exercise 2.45 Prove that if H is an Hilbert space and $U : H \to H$ is an isometry then the spectrum of U (i.e. the set $\sigma(U) := \{\lambda \in \mathbb{C} \ : \ U - \lambda I \text{does not admit a bounded inverse}\}$) is contained in $\mathbb{S}^1 \cup \{0\}$.

Theorem 2.46 (Von Neumann mean ergodic theorem) *If U is an isometry on the Hilbert space H then*

$$\frac{1}{n}\sum_{k=0}^{n-1} U^k\varphi \to P\varphi \qquad \forall \varphi \in H$$

where P is the orthogonal projection on the space of U–invariant vectors.

Let us recall that, if $H = L^2(X, d\mu)$ and $U = U_f$ is the induced operator this shows that Birkhoff sums converge in L^2-norm

$$\frac{1}{n}\sum_{k=0}^{n-1} \varphi \circ f^k \to \int_{\mathcal{J}} \varphi d\mu \qquad \text{in} \;\; L^2.$$

Of course, if μ is ergodic then $\mathcal{J} = \{\emptyset, X\}$ and $\varphi_{\mathcal{J}} = \int_X \varphi d\mu$.
Proof. Let us first deal with the case U unitary operator. Let $Z := \ker(U - I)$ be the space of fixed points of U and $B := \text{range}(U - I)$. If $y \in \bar{B}^{\perp}$ then $(y, Ux - x) = 0$ for all $x \in H$. Hence $U^{-1}y - y \perp x$ for all $x \in H$ and y is a fixed point for U^{-1}. This shows that $\bar{B}^{\perp} \subset Z$. In fact the other inclusion is also true since, again using the invertibility of U, $Ux = x$ implies that $x \in \bar{B}^{\perp}$. Calling P the orthogonal projection on Z we get that $x = Px + y$ whith $y \in \bar{B}$. It is easy to check that, since $y \in \bar{B}$, $\frac{1}{n}\sum_0^{n-1} U^n y \to 0$ as $n \to \infty$ (this is obvious for $y \in B$ and must be true on $\bar{B}$ as well since the U^n are equi-lipschitz). Thus $\frac{1}{n}\sum_{k=0}^{n-1} U^k\varphi \to P\varphi \quad \forall \varphi \in H$ as claimed.

To deal with the noninvertible case let $u \in L(H)$ be an isometry, it is convenient to see u as a *factor* of a unitary operator $U \in L(H \times H)$ defined as follows

$$U := \begin{pmatrix} u & 1 - uu^* \\ 0 & u^* \end{pmatrix}.$$

It is a straightforward calculation to check that U is unitary ; moreover, if we define $j : H \to H \times H$ and $p : H \times H \to H$ as $j(x) = (x, 0)$ and $p(x, y) = x$, we see immediately that $ux = (pUj)x$ and (by induction) $u^k x = (pU^k j)x$. Thus $\frac{1}{n+1}\sum_0^n u^k x = p\left(\frac{1}{n+1}\sum_0^n U^k jx\right) \to Px$ with P is the orthogonal projection on the vector space $\ker(u - 1)$. □

Exercise xxx : Show directly (i.e. without using neither Radon-Nykodim nor any of the results of section 2.3) that, if $H = L^2(X, d\mu)$, $f : X \to X$ and $U = U_f$ is the induced operator, then the projector P of Von Neumann Theorem can be extended by continuity to a linear operator defined on $L^1(X, \mathcal{A}, \mu)$.

From now on we will denote with $L^2_0(X, d\mu)$ the space of L^2 functions with zero mean ; moreover in what follows we shall always use the symbol U instead of U_f since we always deal with a single map f.

Proposition 2.47 *The following conditions are equivalent :*

(i) μ *is ergodic ;*

(ii) 1 **is not** *an eigenvalue of* U_f *restricted to* $L^2_0(X, d\mu)$ *;*

(iii) $\frac{1}{n}\sum_{k=0}^{n-1} U^k\varphi \to 0 \ \ \forall \varphi \in L^2_0(X, d\mu)$ *;*

(iv) $\frac{1}{n}\sum_{k=0}^{n-1} (U^k\varphi, \psi) \to 0 \ \ \forall \varphi, \psi \in L^2_0(X, d\mu)$ *;*

(v) $\frac{1}{n}\sum_{k=0}^{n-1} (U^k\varphi, \varphi) \to 0 \ \ \forall \varphi \in L^2_0(X, d\mu)$ *;*

Proof.

The fact that (i) and (ii) are equivalent has already been proved in Theorem 2.34. The equivalence between (ii) and (iii) follows from Von Neumann Mean Ergodic Theorem together with Theorem 2.34. The implications (iii) $\Rightarrow$ (iv) and (iii) $\Rightarrow$ (v) are strightforward while the converse (iv) $\Rightarrow$ (iii) follows again from Theorem 2.34-(5) just setting $\varphi(x) = \chi_A(x)$, $\psi(x) = \chi_B(x)$. We leave to the reader to prove the implication (v) $\Rightarrow$ (iv) which closes the graph of arrows of this proof. □

Exercise 2.48. Prove that, if f is an ergodic transformation and $U = U_f$ is the induced operator, then

(a) each eigenfunction of U is μ-a.e. constant ;

(b) each eigenvector of U is simple ;

(c) the set of eigenvectors $\{\lambda \in \mathbb{C} \ : \ U\varphi$ for some $\varphi \in L^2\}$ is a multiplicative subgroup of $\mathbb{S}^1$.

The set of eigenvalues of a linear operator U is usually called *point spectrum.*

Another spectral property, stronger than ergodicity, is weak-mixing.

Definition 2.49 *The measurable dynamical system* $(X, \mathcal{A}, \mu, f)$ *is said to be* weak mixing *if the induced operator* $U : L^2_0(x, d\mu) \to L^2_0(x, d\mu)$ *has no (unitary) eigenvalue i.e. if* $U\varphi = \lambda\varphi$ *(and* $|\lambda| = 1$*) implies* $\varphi = 0$.

Proposition 2.50 *The following conditions are equivalent :*

(i') f *is weak-mixing ;*

(ii') $\frac{1}{n}\sum_{k=0}^{n-1} |(U^k\varphi, \psi)| \to 0 \ \ \forall \varphi, \psi \in L^2_0(X, d\mu)$ *;*

(iii') $\frac{1}{n}\sum_{k=0}^{n-1} |(U^k\varphi, \varphi)| \to 0 \ \ \forall \varphi \in L^2_0(X, d\mu)$.

Proof. See Parry. □

Definition 2.51 *The measurable dynamical system* $(X, \mathcal{A}, \mu, f)$ *is said to be* strong mixing *if* $\mu(f^{-n}(A) \cap B) \to \mu(A)\mu(B)$ *as* $n \to +\infty$ *for all* $A, B \in \mathcal{A}$.

Strong mixing has the following phisical interpretation : the portion of the set B which, after n iterates, is occupied by elements that were initially in A tends to the constant value $\mu(A)$ as n goes to ∞. This means that iterating f has the effect of spreading the set A uniformly all over the space X. On the other hand strong mixing has a spectral counterpart as well (although not as immediate as ergodicity or weak mixing). Using the language of probability we could express this sayng that the event B and $f^{-n}A$ tend to become *independent* as $n \to +\infty$. Indeed an analogue of Proposition 2.49 is still true :

Proposition 2.51 *The following conditions are equivalent :*
(i") f *is strong mixing ;*
(ii") $(U^k\varphi, \psi) \to 0 \ \ \forall \varphi, \psi \in L^2_0(X, d\mu)$ *;*
(iii") $(U^k\varphi, \varphi) \to 0 \ \ \forall \varphi \in L^2_0(X, d\mu)$.

Proposition 2.52 *Hyperbolic automorphism of* $\mathbb{T}^2$ *are mixing and hence ergodic.*

Lemma 2.53 *Let* H *be a Hilbert space and* $U \in \mathcal{L}(H)$ *be a unitary operator such that*
(a) $H = \text{span} < \phi_{\xi,j} \ : \ \xi \in \Xi, \ \ j \in \mathbb{Z} >$, *where* Ξ *is some set of indices ;*
(b) $U\phi_{\xi,j} = \phi_{\xi,j+1}$. *Then for all* $\varphi, \psi \in H$ *we have*

$$(U^n\varphi, \psi) \to 0 \text{ as } n \to +\infty.$$

Proof. The claim is strightforward for $\varphi = \phi_{\xi,j}$ *and* $\psi = \phi_{\xi',j'}$ *since the scalar product is zero only if* $\xi = \xi'$ *and, in any case, it vanishes as soon as* $n > h - j$. *Therefore, setting* $D := \{\varphi \in H \ : \ \varphi = \sum_{(\xi,j)\in F} c_{\xi,j}\phi_{\xi,j}, \ F \subset \Xi \times \mathbb{Z}, \ \text{card}(F) < +\infty, c_{\xi,j} \in \mathbb{C}\}$, *arguing by linearity we conclude that the claim is true also if* $\varphi, \psi \in D$. *To deal with the general case proceed as follows : for any* $\epsilon > 0$ *write* $\varphi = \varphi_0 + \varphi_1, \psi = \psi_0 + \psi_1$ *where* $\varphi_0, \psi_0 \in D$ *while both* φ_1, ψ_1 *have norm smaller than* ϵ.

$$\begin{aligned}\limsup_{n\to+\infty} |(U^n\varphi, \psi)| &= \limsup_{n\to+\infty} |(U^n\varphi_0, \psi_0) + (U^n\varphi_1, \psi_0) + (U^n\varphi_0, \psi_1) + (U^n\varphi_1, \psi_1)| \\ &= \limsup_{n\to+\infty} |(U^n\varphi_0, \psi_0)| + \epsilon(\|\varphi\| + \|\psi\|) + \epsilon^2 \leq \epsilon(\|\varphi\| + \|\psi\| + \epsilon)\end{aligned}$$

Since ϵ can be chosen arbitrarily small this implies that the limit must be zero. $\square$

Definition 2.54 *An invertible automorphism of a probability space $(X, \mathcal{A}, \mu)$ has* Lebesgue spectrum L^{Ξ} *if, setting $H := L^2_0(X, d\mu)$, the induced unitary operator $U_f \in \mathcal{L}(H)$ satisfies properties (a) and (b) of the above lemma.* So we have proved that Lebesgue automorphism are mixing. We shall now prove that hyperbolic automorphisms have Lebesgue spectrum.

Proof. The set $\{\Phi_h(x) := exp(2\pi i h \cdot x) \ : h \in \mathbb{Z}^2_*\}$ is an orthonormal base of $L^2_0(\mathbb{T}^2)$; we easily check that $U_A \Phi_h = \Phi_{A^* h}$. It is easily checked that the action of A^* on $\mathbb{Z}^2$ splits the frequencies in a disjoint union of $A*$–orbits and each orbit is unbounded. The claim follows just choosing a representative set $\Xi \subset \mathbb{Z}^2$ of orbits and setting $\phi_{\xi,j} := \Phi_{(A*)^j \xi}$ $(\xi \in \Xi,\ j \in \mathbb{Z})$. $\square$

Exercise 2.55 Prove directly, using only the definition, that hyperbolic automorphisms of $\mathbb{T}^2$ are mixing (see [KH] ?).

Lecture 3. Topological vs. statistical properties of dynamics. Observables and cohomology

In this lecture we will consider a dynamical system on a compact metric space equipped with an invariant probability measure and we will compare the topological notions of topological transitivity and minimality with ergodicity and unique ergodicity. Then we will proceed at introducing the fundamental tool for the study of the observables, and in doing so we will introduce some group cohomology. Finally we will describe an example due to Furstenberg of a minimal non–ergodic map on the two–dimensional torus and a remarkable theorem of Oxtoby–Ulam on the transitivity of the action of the group of homeomorphisms on the space of non–atomic probability measures with full support.

3.1 Topological transitivity and minimality vs. ergodicity and unique ergodicity

Let X be a compact metric space, $f \in \mathrm{Homeo}\,(X)$, μ an f–invariant probability measure defined on the Borelian subsets of X. In the previous lecture we have introduced several topological and statistical notions to study the asymptotic behaviour of the orbits of the dynamical system generated by f. Here we will clarify the relationship between the two approaches show how *ergodicity* relates to *topological transitivity* and *strict ergodicity* to *minimality.*

Definition 3.1 *A probability measure μ on borelian subsets of a compact metric space X has* full support *if $\mu(U) > 0$ for all nonempty open subset U of X. More generally the* support $\mathrm{supp}\,\mu$ *of a probability measure μ is the set of points $x \in X$ such that for all $U \subset X$ open containing x one has $\mu(U) > 0$.*

Exercise 3.2 Show that $\mathrm{supp}\,\mu$ is closed f–invariant and that $\mathrm{supp}\,\mu \subset \mathrm{Rec}_f\,(X)$.

Theorem 3.3 *If f preserves an ergodic probability measure μ with full support then $\mu(\{x \in X\,,\,\overline{\mathcal{O}_f(x)} = X\}) = 1$. In particular f is topologically transitive.*

Proof. Let $\{U_n\}_{n\in\mathbb{N}}$ be a countable basis for the topology of X. Then $\{x \in X\,,\,\overline{\mathcal{O}_f(x)} = X\} = \cap_{n\in N} \cup_{m\in Z} f^m(U_n)$ and $\cup_{m\in Z} f^m(U_n)$ is open nonempty and f–invariant. Since μ is ergodic and has full support one must have $\mu(\cup_{m\in Z} f^m(U_n)) = 1$, which implies the desired result. □

The converse of Theorem 3.3 is not true : a topologically transitive map may fail to have an ergodic measure with full support. An example can be obtained making a suitable time–reparametrization of an irrational linear flow on the two–torus so as to create a fixed point in such a way that the unique invariant probability measure is the Dirac delta at the fixed point. However typical orbits will be dense.

Theorem 3.4 *If f is uniquely ergodic and μ is the f–invariant probability measure then f is minimal if and only if μ has full support.*

Proof. Suppose f is minimal. If U is a nonempty open set then $\cup_{n\in\mathbb{Z}} f^n(U) = X$ by Exercise 2.10 (iii). Thus one must have $\mu(U) > 0$ otherwise $\mu(X)$ would vanish, a contradiction.

Conversely, suppose that μ has full support. If f is not minimal then there exists a nontrivial closed f–invariant set K (Exercise 2.10 (ii)). By Krylov–Bogolubov's theorem the restriction of f to K has an invariant probability measure ν_K. Then setting $\nu(A) = \nu_K(K \cap A)$ for all Borel subsets A of X one obtains an f–invariant probability measure on X which differs from μ since $\mu(X \setminus K) > 0 = \nu(X \setminus K)$. □