MAT733 - HS2018

Dynamical Systems and Ergodic Theory

Part II: Topological and Symbolic Dynamics

Prof. Dr. Corinna Ulcigrai

January 11, 2019

Chapter 2

Topological Dynamics and Symbolic Dynamics

2.1 Review of metric Spaces and Basic Topology notions

In this section we briefly overview some basic notions about metric spaces and topology.

A *metric space* (X, d) is a space X with a *distance function* $d : X \times X \to \mathbb{R}^+$ (also called *metric*, from which the name metric space), that is a function which assigns to each pair of points $x, y \in X$ a real number $d(x, y)$ (their *distance*) and has the following properties:

Definition 2.1.1. A distance d is a function $d : X \times X \to \mathbb{R}^+$ such that

1. If $d(x, y) = 0$ then $x = y$;
2. For each $x, y \in X$ we have $d(x, y) = d(y, x)$ (symmetry);
3. The triangle inequality holds, that is for all $x, y, z \in X$

$$d(x, z) \leq d(x, y) + d(y, z).$$

Examples of metric spaces and distances are the following. The first three are classical examples, while the following two are useful in dynamical systems.

Example 2.1.1. 1. $X = \mathbb{R}$ or $X = [0, 1]$ with

$$d(x, y) = |x - y|.$$

2. $X = \mathbb{R}^2$ or $X = [0, 1] \times [0, 1]$ with the Euclidean distance: if $\underline{x} = (x_1, x_2)$ and $\underline{y} = (y_1, y_2)$ are points in $\mathbb{R}^2$, their distance is

$$d(\underline{x}, \underline{y}) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

3. $X = S^1$ with the arc length distance $d(z_1, z_2)$ defined in §1.2.
4. $\Sigma^+ = \{0, 1\}^{\mathbb{N}}$, the shift space of one-sided sequences, is a metric space with the following distance:

$$d((a_i)_{i=1}^{\infty}, (b_i)_{i=1}^{\infty}) = \sum_{i=1}^{\infty} \frac{|a_i - b_i|}{2^i}.$$

In particular two points $(a_i)_{i=1}^{\infty}, (b_i)_{i=1}^{\infty} \in \Sigma^+$ are close if and only if the first block of digits agree: for example, if $a_k = b_k$ for $1 \leq k \leq n$, then the distance is less than $1/2^n$.

5. Let (X, d) be any metric space and $f : X \to X$. Then for each $n \in \mathbb{N}^+$ we can define a new distance, d_n, given by
$$d_n(x, y) = \max_{k=0,\dots,n-1} d(f^k(x), f^k(y)).$$
Two points x, y are close in the d_n metric if their orbits up to time n stay close. We will use this distance to defined topological entropy in §2.3.

Exercise 2.1.1. Check that the distances in the previous Examples satisfy the properties of distance in Definition 2.1.1. For each, describe the ball of radius ϵ at a point.

In a metric space one can talk about convergence and continuity as in $\mathbb{R}^n$. Let (X, d) be a metric space. Given $x \in X$ and $\epsilon > 0$, let $B_d(x, \epsilon)$ be the ball of radius ϵ around the point x defined using the distance d, that is
$$B_d(x, \epsilon) = \{y \in X \text{ such that } d(x, y) < \epsilon\}.$$
If there is no ambiguity about the distance, we will often write simply $B(x, \epsilon)$, dropping the subscript d. We can use balls to define convergence:

Definition 2.1.2. A sequence $(x_n)_{n\in\mathbb{N}} \subset X$ converges to $\overline{x}$ and we write $\lim_{n\to\infty} x_n = \overline{x}$ if for any $\epsilon > 0$ there exists $N > 0$ such that $x_n \in B_d(\overline{x}, \epsilon)$ for all $n \geq N$.

We can use the distance to define the notion of open and closed sets.

Definition 2.1.3. A set $U \subset X$ of a metric space (X, d) is *open* if for any $x \in U$ there exists an $\epsilon > 0$ such that
$$B_d(x, \epsilon) \subset U.$$
A set $C \subset X$ is *closed* if its complement $X \backslash C$ is open.

Example 2.1.2. If $X = \mathbb{R}$ and $d(x, y) = |x - y|$, the intervals (a, b) are open sets and the intervals $[a, b]$ are closed sets. Also intervals of the form (a, ∞) or (∞, b) are open and intervals of the form $[a, \infty)$ or $(\infty, b]$ are closed. Intervals of the form $[a, b)$ or $(a, b]$ are neither open nor closed.

Exercise 2.1.2. Prove that a ball $B_d(x, \epsilon)$ is open (use the triangle inequality).

Open and closed sets in a metric space enjoy the following property:[1]

Lemma 2.1.1. *(1) Countable unions of open sets are open: if $U_1, U_2, \dots, U_n, \dots$ are open sets, than $\cup_{k\in\mathbb{N}} U_k$ is an open set;*

(2) Finite intersections of open sets are open: if $U_1, U_2, \dots, U_N$ are open sets, than $\cap_{k=1}^N U_k$ is an open set.

Exercise 2.1.3. Prove the lemma using the Definition 2.1.3 above.

Exercise 2.1.4. Given an example in $X = \mathbb{R}$ of a countable collection of open sets whose interesection is *not* open.

By using De Morgan's Laws, it follows that closed sets have the following properties (note that the role of intersections and unions is reversed):

Corollary 2.1.1. *(1) Countable* intersections *of closed sets are closed: if $C_1, C_2, \dots, C_n, \dots$ are closed sets, than $\cap_{k\in\mathbb{N}} C_k$ is a closed set;*

(2) Finite unions *of closed sets are closed: if $C_1, C_2, \dots, C_N$ are closed sets, than $\cup_{k=1}^N C_k$ is a closed set.*

Exercise 2.1.5. Prove the Corollary from Lemma 2.1.1

Exercise 2.1.6. Given an example in $X = \mathbb{R}$ of a countable collection of closed sets whose union is *not* closed.

[1] It is possible to define open sets as an abstract collection of the subsets of X which satisfy certain properties. In this case, Property (1) in the Lemma is taken as an axiom. A collection $\mathscr{U}$ of subsets of X such that $\emptyset, X \in \mathscr{U}$ and Property (1) is satisfied is called a topology. In this case, sets in $\mathscr{U}$ are called open sets and complement of sets in $\mathscr{U}$ are called closed sets. A topological space $(X, \mathscr{U})$ is a space X with a topology $\mathscr{U}$, see Extra.

Definition 2.1.4. A subset $Y \subset X$ is *dense* if for any non-empty open set $U \subset X$ there is a point $y \in Y$ such that $y \in U$.

One can check that this definition of dense set reduces to the usual definition of dense set for a subset $Y \subset \mathbb{R}$, that is, for each $y \in Y$ and $\epsilon > 0$ there exists $y \in Y$ such that $|x - y| < \epsilon$.

Definition 2.1.5. A metric space (X, d) is called *separable* if it contains a countable dense subset.

Example 2.1.3. If $X = \mathbb{R}^n$ with the Euclidean distance, X is separable since the set $\mathbb{Q}^n$ given by all points $(x_1, \ldots, x_n) \in \mathbb{R}^n$ whose coordinates x_i are rational numbers is dense and it is countable.

Let (X, d_X) and (Y, d_Y) be a metric space. We will now consider properties of functions $f : X \to Y$.

Definition 2.1.6. A function $f : X \to Y$ is an *isometry* if it preserves the distances, that is

$$d_Y(f(x), f(y)) = d_X(x, y) \quad \forall\, x, y \in X.$$

We already saw an example of isometry:

Example 2.1.4. If $X = Y = S^1$ is the circle with the arc length distance $d = d_X = d_Y$, then $f = R_\alpha$ the rotation by $2\pi\alpha$ is an isometry.

Definition 2.1.7. A function $f : X \to Y$ is *continuous* if for any $x \in X$ and any $\epsilon > 0$ there exists a $\delta > 0$ such that

$$f(B_{d_X}(x, \delta)) \subset B_{d_Y}(f(x), \epsilon).$$

This definition generalizes the definition of continuity for a function $f : \mathbb{R} \to \mathbb{R}$ that you might have seen in calculus/analysis classes, that is f is continuous if for any $x \in \mathbb{R}$ and any $\epsilon > 0$ there exists $\delta > 0$ such that $|x - y| < \delta$ implies $|f(x) - f(y)| < \epsilon$.

Exercise 2.1.7. Check that if $X, Y \subset \mathbb{R}$ and $d_X(x_1, x_2) = |x_1 - x_2|$, $d_Y(y_1, y_2) = |y_1 - y_2|$ this gives the usual ϵ, δ definition of continuity of a real function.

Exercise 2.1.8. Let $X = Y = S^1$ and $d_X = d_Y = d$ be the arc length distance. Prove that

(a) The rotation $R_\alpha : S^1 \to S^1$ by $2\pi\alpha$ is continuous;

(b) The doubling map $f : S^1 \to S^1$ given in this coordinates by $f(e^{2\pi i\theta}) = (e^{2\pi i 2\theta})$ is continuous.

It is enough to know which are the open sets in X and Y to define the notion of continuity:[2]

Lemma 2.1.2. *A function $f : X \to Y$ is continuous if and only if for each open set $U \subset Y$ the preimage $f^{-1}(U)$ is an open set of X.*

Proof. Assume that f is continuous. Let $U \subset Y$ is open and let us show that $f^{-1}(U)$ is open. We have to show that for each $x \in f^{-1}(U)$ there is $\delta > 0$ such that $B_{d_X}(x, \delta) \subset f^{-1}(U)$. Let $y = f(x)$. Clearly $y \in U$ since $x \in f^{-1}(U)$. By definition of open set there exists $\epsilon > 0$ such that $B_{d_Y}(y, \epsilon) \subset Y$. By definition of continuity, there exists $\delta > 0$ such that

$$f(B_{d_X}(x, \delta)) \subset B_{d_Y}(y, \epsilon) \subset U,$$

thus $B_{d_X}(x, \delta) \subset f^{-1}(U)$. This shows that $f^{-1}(U)$ is open.

The other implication is left as an exercise. □

Exercise 2.1.9. Prove the other implication in Lemma 2.1.2, that is show that if a function $f : X \to Y$ between two metric spaces $(X, d_X), (Y, d_Y)$ is such that for each open set $U \subset Y$ the preimage $f^{-1}(U)$ is an open set of X, then f is continuous in the sense of Definition 2.1.7.

The last metric space notion that we will use is the notion of *compact* sets. Let (X, d) be a metric space.

[2] The following Lemma can be taken as definition of a continuous function when $(X, \mathscr{U})$ is a topological space, see Extra.

Definition 2.1.8. [Sequentially compact] A subset $K \subset X$ is (sequentially) *compact* if for any sequence $(x_n)_{n\in\mathbb{N}} \subset K$ there exists a convergent subsequence $(x_{n_k})_{k\in\mathbb{N}}$ and the limit $\lim_{k\to\infty} x_{n_k} = \overline{x}$ belong to K.

This property is called *sequentially compactness* since the definition involves sequences. There are other notion of compactness (see compactness by covers below) which are equivalent in a metric space, so we will simply say that a set is *compact* and use the term *sequentially* compact only when we specifically want to use the above property of compact sets.

Example 2.1.5. Closed bounded intervals $[a, b] \subset \mathbb{R}$ are sequentially compact (this is known as *Heine-Borel theorem*).

Conversely, in $\mathbb{R}$, if a set is not bounded or not closed, it is not compact. The following two are *non-examples*, that is examples of spaces that are *not* compact.

Example 2.1.6. The unbounded closed interval $[0, \infty)$ is *not* sequentially compact: consider for example the sequence $(x_n)_{n\in\mathbb{N}}$ given by $x_n = n$. The sequence has no convergent subsequence.

The open interval $(0, 1)$ is *not* sequentially compact: consider for example the sequence $(x_n)_{n\in\mathbb{N}}$ given by $x_n = 1/n$. We have $\lim_{n\to\infty} x_n = 0$, but $0 \notin (0, 1)$.

2.1.1 Compactness by covers

In addition to the definition of sequencial compact, there is an other definitions of compactness, *compactness by open covers*, which turn out to be equivalent in a metric space. Compactness by open covers is a more general definition of compactness and can be used as a defintion of compactness in any topological space (see Extra on topological spaces if interested).

Definition 2.1.9. An open cover of $K \subset X$ is a collection $\{U_\alpha\}_\alpha$ of open sets of X such that

$$K \subset \bigcup_\alpha U_\alpha$$

(this is why we say that they *cover* K). A *finite subcover* is a finite subset $\{U_{\alpha_1}, U_{\alpha_1}, \ldots, U_{\alpha_N}\} \subset \{U_\alpha\}_\alpha$ which still covers, that is such that $K \subset \cup_{i=1}^N U_{\alpha_i}$.

Definition 2.1.10. [Compact by covers] A subset $K \subset X$ is *compact by covers* if for any open cover $\{U_\alpha\}_\alpha$ there exists a finite subcover $\{U_{\alpha_1}, U_{\alpha_1}, \ldots, U_{\alpha_N}\} \subset \{U_\alpha\}_\alpha$ such that $X \subset \cup_\alpha U_\alpha$.

Example 2.1.7. The open interval $(0, 1)$ is *not* compact by covers: consider for example the collection

$$\mathscr{U} = \left\{ \left(\frac{1}{n+2}, \frac{1}{n} \right), \qquad n \in \mathbb{N} \right\}$$

is an open cover, but $\mathscr{U}$ does not admit a finite subcover. Indeed, a finite subset of intervals in $\mathscr{U}$ is of the form

$$\left\{ \left(\frac{1}{n_1+2}, \frac{1}{n_1} \right), \left(\frac{1}{n_2+2}, \frac{1}{n_2} \right), \qquad , \left(\frac{1}{n_k+2}, \frac{1}{n_k} \right) \right\}$$

so that if $\overline{n} = \max_{i=1,\ldots,k} n_i$, no point in $\left(0, \frac{1}{\overline{n}+2}\right)$ is covered by the finite collection.

Theorem 2.1.1. *In a metric space (X, d), a subset $K \subset X$ is sequentially compact if and only if it is compact by covers.*

Since we will work only with metric spaces, we will simply say that a set is *compact* and use equivalenty either Definition 2.1.8 or 2.1.10.

Remark 2.1.1. In $\mathbb{R}^n$, any subset $C \subset X$ which is *closed* and *bounded*, that is such that $\sup_{x,y\in C} d(x, y) < +\infty$, is compact.

Extras on Topological Dynamics: Topological Spaces

We will consider only metric spaces and define all notions of topological dynamics (see next section) in the context of metric spaces. More in general, all notions of topological dynamics that we will see can be applyed in the more general setting of *topological spaces*. Metric spaces are a special example of topological spaces. In a metric space, we defined the notion of open and closed sets (see Definition 1.2.3.The collection of open sets determines what is called a *topology* on the metric space. We also saw, from the definition of open set in a metric space, that countable unions and finite intersections of open sets are again open sets (see Lemma 2.1.1) and that countable intersections and finite unions of closed sets are closed (Corollary 2.1.1).These properties of open and closed sets can be taken as axioms to characterize open and closed sets in spaces where a distance is not necessarily given. This leads to the following definitions:

Definition 2.1.11. A *topology* $\mathscr{T}$ on X is a collection $\mathscr{T} \subset \mathscr{P}(X)^3$ of subsets of X, which are known as the *open sets* of X, which satisfy the following properties:

(T1) The empty set and the whole space X belong to $\mathscr{T}$;

(T2) Countable unions of open sets are open: if $U_1, U_2, \ldots, U_n, \ldots$ are open sets, then $\cup_{k\in\mathbb{N}} U_k$ is an open set;

(T3) Finite intersections of open sets are open: if $U_1, U_2, \ldots, U_n$ are open sets, then $\cup_{k=1}^{n} U_k$ is an open set.

Exercise 2.1.10. If (X, d) is a metric space, the collection $\mathscr{T}$ of all sets which are open in the metric space according to Definition 2.1.3, that is all the sets $U \subset X$ such that for each $x \in U$ there exists $\epsilon > 0$ such that $B_d(x, \epsilon) \subset U$, form a topology. Indeed, X and $\emptyset$ satisfy Definition 2.1.3 trivially and hence belong to $\mathscr{T}$, proving $(T1)$. The second property $(T2)$ follows from Lemma 2.1.1.The collection of open sets in a metric space give a topology to the metric space X

Definition 2.1.12. A *topological space* $(X, \mathscr{T})$ is a space X together with a topology $\mathscr{T}$.

Example 2.1.8. [Metric space topology] A metric space (X, d) with the topology given in the example 2.1.10 is a topological space.

The following two are examples of trivial topologies that exist on any set X.

Example 2.1.9. [Trivial topology] Consider a space X and let $\mathscr{T}_{tr} = \{\emptyset, X\}$. One can check that $\mathscr{T}_{tr}$ satisfies $(T1), (T2), (T3)$. This topology is known as *trivial topology.* Thus, $(X, \mathscr{T}_{tr})$ is a topological space.

Example 2.1.10. [Point topology] Consider a space X and let $\mathscr{T}_{pt} = \mathscr{P}(X)$ be the collection of all subsets of X. One can check that also $\mathscr{T}_{pt}$ satisfies $(T1), (T2), (T3)$. This topology is known as *point topology.* Thus, $(X, \mathscr{T}_{pt})$ is a topological space.

In a topological space one can define the notion of convergence or density in the same way we did with metric spaces, just using open sets instead than balls:

Definition 2.1.13. A sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ converges to $\overline{x}$ and we write $\lim_{n\to\infty} x_n = \overline{x}$ if for any open set U containing $\overline{x}$ there exists $N > 0$ such that $x_n \in U$ for all $n \geq N$.

Similarly, one can define what it means for a function to be continuous, taking as definition of continuity the equivalent characterization given by Lemma 2.1.2.

Definition 2.1.14. Let $(X, \mathscr{T})$ be a topological space. A sequence $\{x_n\}_{n\in\mathbb{N}} \subset X$ converges to $\overline{x}$ and we write $\lim_{n\to\infty} x_n = \overline{x}$ if for any open set $U \in \mathscr{T}$ that contains x, there exists $N > 0$ such that $x_n \in U$ for all $n \geq N$.

Lemma 2.1.3. *A function $f : X \to Y$ between two topological spaces $(X, \mathscr{T}_X)$ and $(Y, \mathscr{T}_Y)$ is* continuous *if and only if for each open set $V \in \mathscr{T}_Y$ the preimage $f^{-1}(V)$ is an open set of X, that is $f^{-1}(V) \in \mathscr{T}_X$.*

Finally, the notion of compactness via covers can be defined in any topological space:

Definition 2.1.15. Let $(X, \mathscr{T}_X)$ be a topological space. A subset $K \subset X$ is *compact by covers* if for any open cover $\{U_\alpha\}_\alpha$ there exists a finite subcover $\{U_{\alpha_1}, U_{\alpha_1}, \ldots, U_{\alpha_N}\} \subset \{U_\alpha\}_\alpha$ such that $X \subset \cup_\alpha U_\alpha$.

[3] The notation $\mathscr{P}(X)$ denotes the parts of X, that is the collection of all subsets of X.

In the next sections we will define, in the context of metric spaces, dynamical properties as topological transitivity, topological minimality and topological mixing. All this properties can be defined more in general for topological spaces. This is why they are called *topological* properties and why we talk of *topological* dynamics.

2.2 First topological dynamical properties

Topological dynamics is the branch of dynamical systems which studies topological dynamical systems and their topological dynamical properties. We will not define here topological dynamical systems in general, but we will work with metric spaces[4].

Let (X, d) be a metric space.

Definition 2.2.1. A (discrete-time) *topological dynamical system* is a map $f : X \to X$ where (X, d) is a metric space (or more in general, a topological space) and f is continuous.

We have already seen many examples of topological dynamical systems:

1. The rotation $R_\alpha : S^1 \to S^1$;

2. The doubling map $f : S^1 \to S^1$ (we remarked in §2.1 that f is continuous, as it can be seen easily using the coordinates by $f(e^{2\pi i\theta}) = (e^{2\pi i 2\theta})$);

3. The cat map $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ (look at the cat map action in Figure 1.1 and, recalling that the sides of the parallelogram image of $[0, 1]^2$ by A are glued together, try to convince yourself that f_A is continuous) and more in general any toral automorphisms.

2.2.1 Topological dynamical properties

Let us now define some dynamical properties investigated in topological dynamics.

Definition 2.2.2. A topological dynamical system is called *topologically transitive* if there exists a dense orbit, that is there exists $x_0 \in X$ such that the set $\mathcal{O}_f^+(x_0)$ is dense[5].

Example 2.2.1. For example, the doubling map is topologically transitive, since we constructed a dense orbit (see Theorem 1.3.2 in §1.3.). Another example is given by the rotation R_α by an irrational number α: as we proved in Theorem 1.2.1 in §1.2, for example the orbit of $x = 0$ is dense.

Definition 2.2.3. A topological dynamical system is called *minimal* if all orbits are dense, that is *for all* $x \in X$ the set $\mathcal{O}_f^+(x)$ is dense.

Example 2.2.2. For example, the rotation R_α by an irrational number α is minimal, as we proved in Theorem 1.2.1 in § 1.2. On the other hand, the doubling map is *not* minimal, since there are periodic points which lead to non-dense orbits.

Remark 2.2.1. Minimality implies topological transitivity, since if all orbits are dense, there is in particular one dense orbit. On the other hand, we have just seen that the converse is not true, since there are systems that are topologically transitive but not minimal, as the doubling map.

A useful alternative characterisation of topological transitivity is the following. We say a point $x \in X$ is *isolated* if the singleton $\{x\}$ is an open set in X, or, equivalently, if there is $\epsilon > 0$ such that $B_d(x, \epsilon) = \{x\}$.

[4] More generally, it is enough to use a topological space, for which the notion of continuous map and of homeomorphism is well defined.

[5] If f is invertible, and f^{-1} is continuous (that is, f is a *homeomorphism*), one can require that there exists $x_0 \in X$ such that the full orbit $\mathcal{O}_f(x_0)$ is dense. In some books, this stronger definition of topologically transitive is used for homeomorphisms. Definition 2.2.2 is sometimes referred to as *forward* topologically transitive. If there exists $x_0 \in X$ such that the full orbit $\mathcal{O}_f(x_0)$ is dense, one can prove that there exists x such that $\mathcal{O}_f^+(x)$ is dense, but x could be different than x_0.

Proposition 1. *Let X be compact. A topological dynamical system $f : X \to X$ is topologically transitive if for each pair U, V of* non-empty *open sets there exists $n \in \mathbb{N}$ such that*

$$f^n(U) \cap V \neq \emptyset.$$

The reverse implication holds under the additional assumption that X has no isolated points.

The reverse implication requires the following lemma:

Lemma 2.2.1. *Assume X has no isolated points. If $\mathcal{O}_f^+(x_0)$ is dense, then for any $n \in \mathbb{N}$ also $\mathcal{O}_f^+(f^n(x_0))$ is dense.*

Proof of Lemma. Since X has no isolated points, every open non-empty set $U \subset X$ contains infinitely many distinct open subsets U_k. [For $x \in U$ and K sufficiently large, take for instance $U_k = B(x, \frac{1}{k}) \setminus \overline{B(x, \frac{1}{k+1})}$, $k \geq K$. Here $\overline{B(x, \epsilon)} = \{x \in X : d(x, \epsilon) \leq \epsilon)\}$ is the closed ball at x.] This implies that, since $\mathcal{O}_f^+(x_0)$ is dense, there are infinitely many integers m_k such that $f^{m_k}(x_0) \in U_k \subset U$. Given $n \in \mathbb{N}$, choose k such that $m_k \geq n$. Then $U \ni f^{m_k}(x_0) = f^{m_k - n}(f^n(x_0))$ and thus there is an integer m (namely $m = m_k - n$) such that $f^m(f^n(x_0)) \in U$. □

Proof of Proposition 1. We first prove the reverse implication. Assume that f is topologically transitive and X has no isoluated points. Let x_0 be such that $\mathcal{O}_f^+(x_0)$ is dense. Given U, V open sets, by density there exists n such that $f^n(x_0) \in U$. Since by Lemma 2.2.1 also $\mathcal{O}_f^+(f^n(x_0))$ is dense, there exists m such that $f^m(f^n(x_0)) \in V$. So

$$f^{m+n}(x_0) \in f^m(U) \cap V,$$

which shows that $f^m(U) \cap V \neq \emptyset$.

[Extra:] Let us now prove the first implication. Assume that for each pair U, V of non-empty open sets there exists $n \in \mathbb{N}$ such that $f^n(U) \cap V \neq \emptyset$. Since X is compact, one can prove that X has a countable dense subset, see the Exercise in §2.1 (for example, in the unit square $[0,1]^2$, which is a compact set in $\mathbb{R}^2$, the set $\mathbb{Q}^2 \cap [0,1]^2$ of points whose coordinates are rational is a dense subset and is countable).

Let $\{x_n\}_{n \in \mathbb{N}}$ be the points of thie countable dense subset. To show that the orbit of a point $x \in X$ is dense, it is enough to show that for each $k \in \mathbb{N}$ and $n \in \mathbb{N}$ there exists a m such that $f^m(x) \in B(x_n, 1/k)$. This is because any open set U contains a point x_n for some n (by density) and hence, since it is open, it contains the ball $B(x_n, 1/k)$ for some $k \in \mathbb{N}$, so if $f^m(x) \in B(x_n, 1/k) \subset U$, then $\mathcal{O}_f^+(x) \cap U \neq \emptyset$.

Since the balls $B(x_n, 1/k)$ for $n \in \mathbb{N}$ and $k \in \mathbb{N}$ are countable, we can relabel them and enumerate them as

$$\{B(x_n, 1/k), \quad n \in \mathbb{N},\ k \in \mathbb{N}\} = \{U_1, U_2, \ldots, U_n, \ldots\}.$$

Let us know use transitivity to construct an orbit which visits all these balls in the order in which we listed them[6]. Let $B_0 = B(x, \epsilon)$ be any ball and let $\overline{B_0}$ be the closed ball $\overline{B_0} = \{y \quad d(x, y) \leq \epsilon\}$. By assumption, there exists N_1 such that $f^{N_1}(B_0) \cap U_1 \neq \emptyset$. Thus, we can pick a ball inside the non-empty open set $B_0 \cap f^{-N_1}(U_1)$. Up to reducing the radius, we can assume that there is a smaller ball, that we call B_1, such that the closed ball

$$\overline{B_1} \subset B_0 \cap f^{-N_1}(U_1).$$

By assumption there exists N_2 such that $f^{N_2}(B_1) \cap U_2 \neq \emptyset$. Thus, as before we can find a smaller ball B_2 such that

$$\overline{B_2} \subset B_1 \cap f^{-N_2}(U_2).$$

Repeating by induction, we can construct a sequence balls B_n such that

$$\overline{B_{n+1}} \subset B_n \cap f^{-N_{n+1}}(U_{n+1}). \tag{2.1}$$

Since the closed balls are nested, that is $\overline{B}_{n+1} \subset B_n$ and we are in a compact space, their intersection is non-empty[7]. If $x \in \cap_n \overline{B}_n$, then $f^{N_n}(x) \in U_n$ for any n by (2.1), thus the orbit of x is dense. □

[6] Compare this with the proof that we gave that the doubling map has a dense orbit. For the doubling map, we used the collection of binary intervals of the form $I(a_0, \ldots, a_n)$, that is countable and plays the same role that the U_i in the proof and we then constructed an orbit which visits them all.

[7] This is a consequence of compactness that you might have seen in Metric Spaces: in a compact set, a countable collection of non empty nested closed sets has a non empty intersection.

A dynamical property stronger than topological transitivity is the following, which is the first mathematical definition of the intuitive idea of *mixing* (we will see in Chapter 4 another definition of mixing in the context of ergodic theory).

Definition 2.2.4. A topological dynamical system $f : X \to X$ is called *topologically mixing* if for any pair U, V of *non-empty* open sets there exists $N \in \mathbb{N}$ such that for all $n \geq N$ we have $f^n(U) \cap V \neq \emptyset$.

Topological mixing conveys the idea that each set U, after iterations of f, become *spread everywhere*: for each V, for all n sufficiently large, $f^n(U)$ intersects V.

If f is topologically mixing, in particular it is topologically transitive. This follows from the characterization of topologically transitive in Proposition 1: if $f^n(U) \cap V \neq \emptyset$ for all $n \geq N$, in particular there is an n such that $f^n(U) \cap V \neq \emptyset$. Topologically mixing though, requires that the sets intersect for *all* large enough n. Let us start by giving a non-example, that is an example of a map which is topologically transitive but *not* topologically mixing.

Example 2.2.3. Rotations $R_\alpha : S^1 \to S^1$ are *not* topologically mixing. For simplicity take $\alpha < 1/2$. Take for example U, V to be two arcs, each of arc length less than $\pi\alpha$. Then one can see there are infinitely many k such that the images $R_\alpha^k(U)$ does not intersect V: for every $[1/\alpha]$ iterates of R_α ($[1/\alpha]$, that is the integer part of $2\pi/2\alpha$ is the number of iterates to turn once around the cirle), there is at most one iterate such that $R_\alpha(U)$ intersects V (drawing a picture of the iterates of R_α will help you understand it).

On the other hand, if α is irrational, R_α is minimal (by Theorem 1.2.1 in § 1.2) and in particular topologically transitive. So irrational rotations are topologically transitive but not topologically mixing.

Let us now give an example of a topologically mixing dynamical system. In the following we make $[0,1]^2$ a metric space by using the standard Euclidean metric (we could also use the Manhattan distance $d(x, y) = \max(|x_1 - y_1|, |x_2 - y_2|)$; see what this changes in the discussion below.)

Proposition 2. *The baker map* $F : [0,1]^2 \to [0,1]^2$ *is topologically mixing.*

Proof. Let U, V be any two non empty open sets in X. Since U contains a small ball, it also contain a small dyadic square Q, that is, a square with sides of length $1/2^n$ which has the form

$$Q = \left(\frac{i}{2^n}, \frac{i+1}{2^n}\right) \times \left(\frac{j}{2^n}, \frac{j+1}{2^n}\right), \qquad 0 \leq i, j \leq 2^n - 1.$$

Let us describe the iterates $F^n(Q)$.

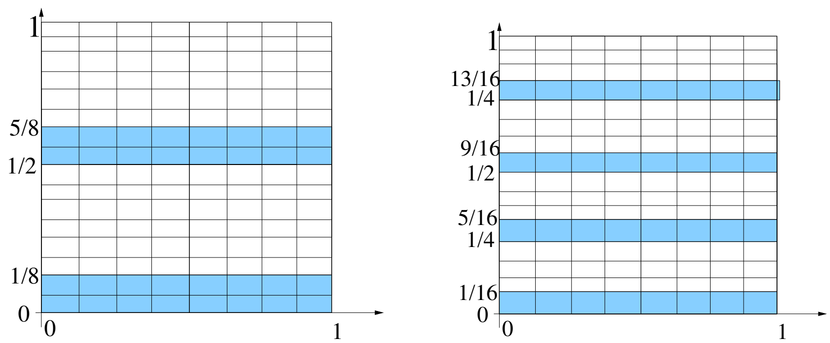


Figure 2.1: Here $n = 1$, $i = j = 0$. The figures show 2^k horizontal strips in $F^{n+k}(Q)$ with spacing $1/2^k$ for $k = 1$ and $k = 2$.

For each $k = 1, 2, \ldots, n$, F acts on Q by doubling the horizontal width and halving the vertical height, so that $F^k(Q)$ is a dyadic rectangle of width $2^k(1/2^n) = 1/2^{n-k}$ and height $(1/2)^k(1/2^n) = 1/2^{n+k}$. In particular, for $n = k$, $F^n(Q)$ is thin horizontal rectangle of full width equal to 1. The image by F of a full horizontal rectangle consists of *two* full horizontal rectangles, whose vertical distance is at most $1/2$ (recall how F acts geometrically and draw a picture to understand it). Since each iterate of F splits a full horizontal rectangle into two, $F^{k+n}(Q)$ (for $k \in \mathbb{N}$) consists of 2^k horizontal rectangles of width 1, whose vertical spacing is no more than $1/2^k$ (where by vertical spacing we mean the distance between for example the centers of two consecutive strips), see for example Figure 2.1. Now let $B(y, \epsilon)$ be a ball contained in V, with $\epsilon > 0$ sufficiently small. Then, for all k such that $1/2^k < \epsilon$, we have $F^{n+k}(Q) \cap B(y, \epsilon) \neq \emptyset$ and hence $F^{n+k}(U) \cap V \neq \emptyset$. $\square$

The doubling map and the cat map are also topologically mixing dynamical systems (see Exercises below) and provide examples in which one can prove that a small set (for example a dyadic rectangle for a baker map or a rectangle whose sides are in eigenvalues directions for the cat map) is *spread* under the dynamics.

Exercise 2.2.1. Let $X = \mathbb{R}/\mathbb{Z}$ and let $f(x) = 2x \mod 1$ be the doubling map. You can use that if $d(x, y) < 1/4$, then $d(f(x), f(y)) = 2d(x, y)$.

(a) Let I be a dyadic interval, that is an interval of the form

$$I = \left(\frac{i}{2^N}, \frac{i+1}{2^N}\right), \qquad N \in \mathbb{N}^+, \quad 0 \leq i \leq 2^N - 1.$$

Describe the iterates $f^n(I)$ for $n \in \mathbb{N}$: how many dyadic intervals do they consist of? of which size? [*Hint:* You will need to consider separately what happens if $n \leq N$ and if $n > N$.]

(b) Show that for any non-empty open set U there exists $N \in \mathbb{N}$ such that $f^N(U) = X$. *continues...*

(c) Show that the doubling map is topologically mixing.

Exercise 2.2.2. Let $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ be the cat map. Let $\lambda_1 > 1, \lambda_2 < 1$ be the eigenvalues of A and let

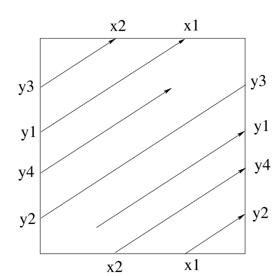
$$\underline{v}_1 = \begin{pmatrix} \frac{1+\sqrt{5}}{2} \\ 1 \end{pmatrix}, \qquad \underline{v}_2 = \begin{pmatrix} \frac{1-\sqrt{5}}{2} \\ 1 \end{pmatrix}.$$

be the corresponding eigenvectors. Check that $\underline{v}_1$ and $\underline{v}_2$ are orthogonal.

(a) Let Q be a small rectangle whose sides have direction v_1 and v_2 respectively. Describe how the iterates $f_A^n(Q)$ look like.

(b) Show that the cat map is topologically mixing.

[Hint: In both Part (a) and (b) you can use that the directions of v_1 and v_2 are irrational and that if $\pi(L)$ is the projection via $\pi : \mathbb{R}^2 \to \mathbb{T}^2$ of a line L with irrational slope (see figure below), then $\pi(S)$ is *dense* in $\mathbb{T}^2$, that is for any non empty open set $U \subset \mathbb{T}^2$ there is a point of $\pi(S)$ inside U.]



2.3 Topological conjugacies and more topological dynamical properties

We already defined the notion of conjugacy and semi-conjugacy between dynamical systems. In the setting of topological dynamics, it is natural to ask more from the conjugacy, so that the properties of a topological dynamical systems are preserved by conjugacy.

2.3.1 Topological conjugacy

Definition 2.3.1. A map $\psi : Y \to X$ is a *homeomorphism* if it is continuous and invertible so that the inverse $\psi^{-1} : X \to Y$ is continuous.

Definition 2.3.2. Two topological dynamical systems $f : X \to X$ and $g : Y \to Y$ are *topologically conjugate* if they are conjugate and the conjugacy map $\psi : Y \to X$ is a *homeomorphism*. We will call ψ a *topological conjugacy.*

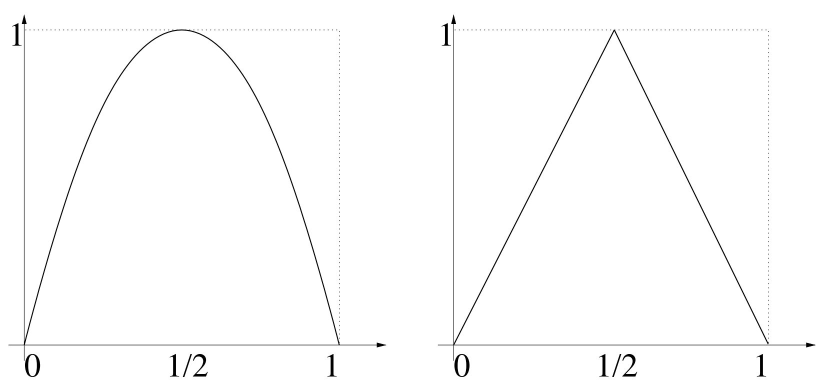
Definition 2.3.3. Two topological dynamical systems $f : X \to X$ and $g : Y \to Y$ are *topologically semi-conjugate* if they are semi-conjugate and the semi-conjugacy map $\psi : Y \to X$ is not only surjective but also *continuous*. W call ψ a *topological semi-conjugacy.*

The definition is the same of conjugacy or semi-conjugacy, just that we require additionally that ψ is continuous to have a topological semi-conjugacy or a homeomorphism (thus that also the inverse is continuous) to have a topological conjugacy.

Example 2.3.1. Let $g : [0,1] \to [0,1]$ be the logistic map $f(x) = 4x(1-x)$ (graph in Figure 2.2(a)) and let $g : [0,1] \to [0,1]$ be the *tent map*, that is

$$g(x) = 1 - |2x-1| = \begin{cases} 2x & \text{if } 0 \le x \le 1/2, \\ 2-2x & \text{if } 1/2 \le x \le 1 \end{cases}$$

The graph of the tent map is shown in Figure 2.2(b).



(a) The logistic map f (b) The tent map g

Figure 2.2: The tent map g and the logistic map f are topologically conjugate.

The logistic map f and the tent map g are topologically conjugate. Let us show that the topological conjugacy is the map $\psi : [0,1] \to [0,1]$ given by

$$\psi(x) = \sin^2\left(\frac{\pi}{2}x\right). \tag{2.2}$$

Let us first show that the following diagram commutes

$$\begin{tikzcd} {[0,1]} \arrow[r, "g"] \arrow[d, "\psi"] & {[0,1]} \arrow[d, "\psi"] \\ {[0,1]} \arrow[r, "f"] & {[0,1]} \end{tikzcd}$$

Recall that we have the trigonometric identities

$$\sin 2\theta = 2\sin\theta\cos\theta, \qquad \sin(\pi-\theta) = \sin(\theta). \tag{2.3}$$

Thus, using this trigonometric identities we have

$$f(\psi(x)) = 4\psi(x)(1-\psi(x)) = 4\sin^2\left(\frac{\pi}{2}x\right)\left(1-\sin^2\left(\frac{\pi}{2}x\right)\right) = \left(2\sin\left(\frac{\pi}{2}x\right)\cos\left(\frac{\pi}{2}x\right)\right)^2 =$$
$$\sin^2\left(2\frac{\pi}{2}x\right) = \sin^2(\pi-\pi x) = \sin^2\left(\frac{\pi}{2}(2-2x)\right) = \psi(g(x)).$$

Thus $f \circ \psi = \psi \circ g$.

Let us show that ψ is a homeomorphism. It is clearly continuous, since it is composition of continuous functions. Since $\psi(0) = 0$ and $\psi(1) = 1$, by continuity and intermediate value theorem it assumes all values in $[0, 1]$, thus it is surjective. Since

$$\psi'(x) = 2\sin\left(\frac{\pi}{2}x\right)\cos\left(\frac{\pi}{2}x\right)\frac{\pi}{2} > 0 \quad \text{for } 0 < x < 1,$$

ψ is monotonic, thus injective. Thus, it is invertible. The inverse is given by

$$\psi^{-1}(y) = \frac{2}{\pi}\arcsin(\sqrt{y}),$$

as you can check by computing $\psi\psi^{-1}$ and showing that it gives the identity map. Hence, also the inverse is continuous. Thus ψ is a homeomorphism and it gives a topological conjugacy.

Topological conjugacies preserve many topological dynamical properties (see Proposition 3 below). Thus, if one finds a topolological conjugacy of a map f with a simpler map g, one can analyse the simpler map g to obtain information about dynamical properties of the original map f. This is for example the case in the previous example: logistic maps (also called logistic family) appear very often as models of real-life dynamical systems, for example in biology. Many chaotic properties of the logistic map f_μ for $\mu = 4$ can be studied through the topological-conjugacy with the tent map: the tent map turns out to be easier to analyse, since it can be studied using binary expansions similarly to the doubling map (in particular, one can find all periodic points and construct dense orbits, see Exercise at the end of this section).

We have already seen that a conjugacy maps periodic points of period n to periodic points of period n. All the topological dynamical properties that we saw in this lecture are preserved by topological conjugacy.

Proposition 3. *Assume that the topological dynamical systems $f : X \to X$ and $g : Y \to Y$ are topologically conjugate. Then:*

(1) f is topologically transitive *if and only if g is topologically transitive;*

(2) f is minimal *if and only if g is minimal;*

(3) f is topologically mixing *if and only if g is topologically mixing.*

Thus, if two topologically conjugate dynamical systems have the same topological dynamical properties.

Remark 2.3.1. If the two systems are semi-conjugate, only one of the implications in (1), (2), (3) follows (see Exercise 2.3.2 and Exercise 2.3.3).

Lemma 2.3.1. *If $f : X \to X$ and $g : Y \to Y$ are topologically conjugated by the homeomorphism $\psi : Y \to X$, then for all $y \in Y$ we have: the orbit $\mathcal{O}_g^+(y)$ is dense in Y if and only if the orbit $\mathcal{O}_f^+(\psi(y))$ is dense in X.*

Proof of Lemma 2.3.1. Assume that $\mathcal{O}_g(y)^+$ is dense and let us show that $\mathcal{O}_f^+(\psi(y))$ is dense. For any $U \subset X$ non-empty open set, $\psi^{-1}(U)$ is an open set in Y since ψ^{-1} is continuous because ψ is an homeomorphism and it is non-empty since ψ is surjective. By density of $\mathcal{O}_g(y)^+$, there exists $k \in \mathbb{N}$ such that

$$g^k(y) \in \psi^{-1}(U) \qquad \Leftrightarrow \qquad \psi(g^k(y)) \in U.$$

Since ψ is a conjugacy, $f^k\psi = \psi g^k$ so $f^k(\psi(y)) = \psi(g^k(y)) \in U$, so $\mathcal{O}_f^+(\psi(y))$ intersects U. This holds for any non-empty open set U and thus shows that $\mathcal{O}_f^+(\psi(y))$ is dense. The other implication follows simply by exchanging the role of f and g, and ψ and ψ^{-1}. $\square$

Exercise 2.3.1. Check that if the topological dynamical systems $f : X \to X$ and $g : Y \to Y$ are topologically semi-conjugate, one can still prove that if the orbit $\mathcal{O}^+_g(y)$ of $y \in Y$ is dense, then the orbit $\mathcal{O}^+_f(\psi(y))$ of $\psi(y)$ is dense in X.

Proof of Proposition 3. Part (1) and (2) follow from the definitions as a consequence of the Lemma 2.3.1, since dense orbits are mapped to dense orbits by the conjugacy. Indeed, if g is topologically transitive, there exists $y \in Y$ so that $\mathscr{O}^+_g(y)$ is dense and by Lemma 2.3.1 $\mathscr{O}^+_f(\psi(y))$ is dense in X so also f is topologically transitive. Similarly, if g is minimal, for any $x \in X$, let $y = \psi^{-1}(x) \in Y$ (which is well defined since ψ is intertible). By minimality of g $\mathscr{O}^+_g(y)$ is dense and by Lemma 2.3.1 $\mathscr{O}^+_f(x)$ is dense in X so also f is minimal. The converse implications follow by reversing the role of f and g and noting that if $\psi : Y \to X$ is a topological conjugacy also $\psi^{-1} : X \to Y$ is a topological conjugacy.

Let us now prove Part (3). Assume that g is topologically mixing and let us deduce that f is also topologically mixing. Given U, V open and non empty, $\psi^{-1}(U)$ and $\psi^{-1}(V)$ are also open, since ψ is continuous, and non-empty, since ψ is surjective. Thus there exists N such that for any $n \geq N$ we have that $g^n(\psi^{-1}(U)) \cap \psi^{-1}(V) \neq \emptyset$. Let $y \in g^n(\psi^{-1}(U)) \cap \psi^{-1}(V)$. Recall that by definition of conjugacy, since ψ is invertible, we have $\psi \circ g \circ \psi^{-1} = f$ and hence by induction $\psi \circ g^n \circ \psi^{-1} = f^n$. Thus, if we consider $\psi(y)$ we have that

$$\psi(y) \in \psi\left(g^n(\psi^{-1}(U))\right) \cap V = f^n(U) \cap V.$$

Hence for any $n \geq N$ we have $f^n(U) \cap V \neq \emptyset$, which shows that f is topologically mixing. The other implication follows again by reversing the role of f and g. $\square$

The next Exercise follows from Exercise 2.3.1 as Proposition 3 follows from Lemma 2.3.1:

Exercise 2.3.2. If $f : X \to X$ and $g : Y \to Y$ are topologically semi-conjugated by $\psi : Y \to X$, then if g is topologically transitive, f is topologically transitive and if g is minimal then f is minimal.

Exercise 2.3.3. Prove that if $f : X \to X$ and $g : Y \to Y$ are topologically semi-conjugated by $\psi : Y \to X$ then if g is topologically mixing then f is topologically mixing.

2.3.2 More topological dynamical properties

Let us defined two more topological dynamical properties, sensitive dependence and expansivity, and then let us give a mathematical definition of chaotic systems.

The following definition which captures mathematically the notion of sensitive dependence on initial conditions: in certain systems, even if two points start arbitrally close, their orbits will become eventually macroscopically far apart[8].

Definition 2.3.4. A topological dynamical system $f : X \to X$ on a metric space (X, d) has *sensitive dependence on initial conditions* (or simply sensitive dependence) if there exists $\Delta > 0$, called the *sensitivity constant* such that for all $x \in X$ and all $\delta > 0$, there exists $y \in B(x, \delta)$ and $n \in \mathbb{N}$ such that

$$d(f^n(x), f^n(y)) \geq \Delta.$$

Sensitive dependence means that arbitrarily close to each point of the space there are points whose future itereates will become Δ-apart from the iterates of the given point. This phenomenon cause high unpredictability: if for example one tries to use a computer to understand the dynamics of a system, one needs to approximating the initial conditions by rounding it off and if the system has sensitive dependence on initial conditions, this might cause a huge difference: the simulated orbit might be completely different than the real evolution.

Example 2.3.2. Show that if $\Delta > 0$ is a sensitivity constant for $f : X \to X$, any constant $0 < \Delta' < \Delta$ is also a sensitivity constant.

[8] A popular image for this phenomenon is the so called *butterfly effect*: "'a butterfly in China could cause a hurricane in mexico"'. A small difference in initial conditions, like the presence of a buttarfly swinging, in a chaotic system could create a huge difference in the long-term evolution, as the creation of a hurricane.

Note that we do not require that all points in $B(x,\delta)$ have different evolution, but only that in each ball there is at least one point with that property. A stronger requirement, which implies sensitive dependence, is the following:

Definition 2.3.5. A topological dynamical system $f : X \to X$ on a metric space (X,d) is called *expansive* if there exists $\nu > 0$, called the *expansive constant* such that for all $x,y \in X$ such that $x \neq y$ there exists $n \in \mathbb{N}$ such that

$$d(f^n(x), f^n(y)) \geq \nu.$$

If f is expansive, the orbits of *all* points nearby a given point $x \in N$ have some iterate which eventually become far apart. Thus, if f is expansive with constant ν, in particular it has sensitive dependence with constant $\Delta = \nu$.

Example 2.3.3. The doubling map is expansive. We can take $\nu = 1/4$. We have already seen in §1.3 that if $d(x,y) < 1/4$, then $d(f(x), f(y)) = 2d(x,y)$. Let $x \neq y$ be distinct points. If $d(x,y) \geq \frac{1}{4}$, then the definition of expansive holds with $n = 0$ for x and y. Otherwise, by the above relation $d(f(x), f(y)) = 2d(x,y)$. If $d(f(x), f(y)) \geq \frac{1}{4}$, we are again done since the definition holds with $n = 1$. Otherwise, $d(f^2(x), f^2(y)) = 2d(f(x), f(y)) = 2^2 d(x,y)$. Thus, continuing in this way, if $d(f^k(x), f^k(y)) < \frac{1}{4}$, for all $0 \leq k \leq n-1$, then

$$d(f^n(x), f^n(y)) = 2^n d(x,y).$$

Note that $d(x,y) \neq 0$ by the properties of a distance if $x \neq y$. Thus, for given $x \neq y$ there exists $n \in \mathbb{N}$ such that $2^n d(x,y) \geq \frac{1}{4}$ (namely any $n \geq \frac{\log \frac{1}{4d(x,y)}}{\log 2}$), and therefore $d(f^n(x), f^n(y)) \geq 1/4$. We conclude the doubling map has sensitive dependence on initial conditions with constant $1/4$.

Exercise 2.3.4. The linear maps $E_m : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ given by $E_m(x) = mx \mod 1$ (where $m \in \mathbb{N}$, $m > 1$) are expansive with expansive constant $\nu = 1/2m$.

Let us now give an example of a map which has sensitive dependence but is not expansive.

Example 2.3.4. Let $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ be the cat map. Let us first show that f_A has sensitive dependence on initial conditions. Take for example $\Delta = 1/2$. For any $\underline{x} \in \mathbb{T}^2$ and any $\delta > 0$, consider all $\underline{y} \in B_d(\underline{x}, \delta)$ on the line through $\underline{x}$ in direction of the eigenvector v_1 with eigenvalue $\lambda_1 > 1$. Since f_A expands points in direction v_1 with factor λ_1, we have that

$$d(f_A^n \underline{x}, f_A^n \underline{y}) = \lambda_1^n d(\underline{x}, \underline{y})$$

at least for all n such that $\lambda_1^n d(\underline{x}, \underline{y}) \leq 1/2$. Since $\lambda_1 > 1$, there exists n_0 such that $1/2\lambda_1^{n_0} < \delta$. We can now choose $\underline{y} \in B_d(\underline{x}, \delta)$ so that $d(\underline{y}, \underline{x}) = 1/2\lambda_1^{n_0}$. Then $d(f_A^{n_0} \underline{x}, f_A^{n_0} \underline{y}) = 1/2$. This shows that $\Delta = 1/2$ is a sensitivity constant.

Let us now show that f_A is not expansive. Fix any $\nu \in (0, \frac{1}{2}]$. Then fix any $\underline{x} \in \mathbb{T}^2$ and consider $\underline{y} \in B_d(\underline{x}, \nu)$ on the line through $\underline{x}$ in direction of the eigenvector v_2 with eigenvalue $\lambda_2 < 1$. Since f_A contracts distances in direction v_2 with factor λ_2, we have that

$$d(f_A^n \underline{x}, f_A^n \underline{y}) = \lambda_2^n d(\underline{x}, \underline{y}) < d(\underline{x}, \underline{y}) < \nu.$$

for all $n \in \mathbb{N}$. We have thus shown that for any $\nu \in (0, \frac{1}{2}]$ there exist $\underline{x} \neq \underline{y}$ such that $d(f_A^n \underline{x}, f_A^n \underline{y}) < \nu$ for all $n \in \mathbb{N}$. This implies f_A is not expansive.

Exercise 2.3.5. Show that any hyperbolic toral automorphism $f_X : \mathbb{T}^2 \to \mathbb{T}^2$ has sensitive dependence but is not expansive.

Exercise 2.3.6. Let $F : [0,1]^2 \to [0,1]^2$ be the Baker map.

(a) Show that the baker map has sensitive dependence on initial conditions with $\Delta = 1/4$;

(b) Show that it is not expansive, that is that there is no $\nu > 0$ such that for every two points $(x_1, y_1), (x_2, y_2) \in [0,1]^2$ there is n such that $d(F^n((x_1, y_1), F^n(x_2, y_2)) \geq \nu$.

[*Hint*: Look at nearby points on the same horizontal line for (a) and at nearby points on the same vertical line for (b)]

In various books[9] the following is taken as definition of *chaos*.

Definition 2.3.6. A topological dynamical system $f : X \to X$ is called *chaotic* if

(C0) f has sensitive dependence on initial conditions;

(C1) f is topologically transitive;

(C2) The set $Per(f)$ of perodic points for f is dense in X.

We have already seen some examples of chaotic dynamical systems according to this definition.

Example 2.3.5. The doubling map is chaotic. We have already seen that it is expansive (Eg 2.3.3), that it is topologically transitive (Theorem 1.3.1 in §1.3) and that periodic points are dense (Exercise 1.3.4 in §1.3).

Example 2.3.6. Rotations R_α of the circle are *not* chaotic according to this definition. If α is irrational, there are no periodic points. If α is rational, there are no dense orbits. Moreover, for any α, R_α does not have sensitive dependence on initial conditions since R_α is an isometry and the iterates of nearby points remain close by.

Other maps that can be proved to be chaotic are the baker map, the cat map and the logistic map. For the latter, one can use the fact that it is topologically conjugate to the doubling map. Indeed, let us remark that if $f : X \to X$ and $g : Y \to Y$ are topologically conjugate one can prove, similarly to what we did for the other properties, that f is chaotic if and only if g is chaotic.

* **Exercise 2.3.7.** Consider the *linear twist* $T : \mathbb{T}^2 \to \mathbb{T}^2$, that is the map given by

$$T(x, y) = (x, y) \to (x + y \mod 1, y \mod 1).$$

(a) Show that T has sensitive dependence on initial conditions.

(b) Show that if y is rational than (x, y) is periodic. Conclude that $Per(T)$ is dense.

(c) Is T chaotic?

* **Exercise 2.3.8.** Consider the tent map f and the logistic map g defined above in this lecture.

- Show that the tent map f is topologically mixing;
- Show that periodic points $Per(f)$ are dense;
- Prove that the logistic map g is chaotic.

[9] The first to adopt this as definition of chaos was Devaney, in *An Introduction to Chaotic Dynamical Systems*. Sometimes this definition is also called *Devanay chaotic*.

2.4 Topological Entropy

We will define in this section a very important and deep concept in dynamics, the concept of entropy. We will consider here topological entropy. There is also a notion of metric entropy, that is related to topological entropy but we will not see in this course (metric entropy is very important also in information theory and the concept of entropy was defined independently in dynamics and in information theory[10]). Topological entropy is a positive number assigned to each topological dynamical system, that roughly tells *how much* chaotic a system is. Let us try to give first an intuitive idea of what it measures.

Imagine to look at your space X with a finite scale resolution (for example, if you look at your system from far away you will not distinguish points that are very close; similarly if you do computer simulations of your system, the finite precision of a computer forces you to divide the space in small bits). Consider finite pieces of orbits. Then you will be able to distinguish only a finite number of orbits up to time n. As the length n of the orbits that you consider increases (and also as the resolution increases), the number of orbits will increase. In systems where the entropy is positive, the number of orbits increase *exponentially.*

Topological entropy gives the exponential rate of growth of the number of orbits distinguishable with finite but arbitrary precision.

There are various equivalent ways of defining topological entropy and sometimes one is more convenient than the others to compute it. We will present the one introduced by Bowen[11] (using $(n, \epsilon)-$separated, §2.4.1, and $(n, \epsilon)-$spanning sets, section §2.4.2). In section §2.4 we also give the definition of entropy via covers[12].

2.4.1 Entropy via separated sets

Let (X, d) be a metric space and let $f : X \to X$ be a topological dynamical system. In this section we will always assume that X is *compact.*

For each $n \in N$ let us define a new distance, $d_n : X \times X \to [0, \infty)$ given by

$$d_n(x, y) = \max_{0 \leq k < n} d(f^k(x), f^k(y)).$$

Thus, two points are $\epsilon-$close with respect to the distance d_n if their iterates under f stay $\epsilon-$close until time n. Note that the definition of d_n depends on the trasformation f. Thus, the balls with respect to this metric

$$B_{d_n}(x, \epsilon) = \{y \in X \text{ such } \text{ that } d(f^k(x), f^k y) < \epsilon \text{ for all } \ 0 \leq k < n\}$$

consist of all points whose trajectories up to time n stay ϵ-close to the finite orbit segment $\{x, f(x), \dots, f^{n-1}(x)\}$. Another way to express it is

$$B_{d_n}(x, \epsilon) = B_d(x, \epsilon) \cap f^{-1}(B_d(f(x), \epsilon) \cap \cdots \cap f^{-n+1} B_d(f^{n-1}(x), \epsilon).$$

Indeed, the points y in the intersections are exactly the points such that $f^k(y) \in B_d(f^k(x), \epsilon)$ for all $0 \leq k \leq n - 1$.

Definition 2.4.1. Let $\epsilon > 0$, $n \in \mathbb{N}$. A set $S \subset X$ is (ϵ, n)-*separated* if for all distinct point $x, y \in S$, $x \neq y$, we have $d_n(x, y) \geq \epsilon$.

Points in an (ϵ, n)-separated set have trajectories that, with a finite scale resolution ϵ, can be recognized as different in time n. Let us give an example of an (n, ϵ)-separated set.

[10]The definition of metric entropy, often called *Kolmogorov-Sinai entropy*, was introduced for the first time by Kolmogorov, one of the fathers of ergodic theory, in a paper in 1958 and was successively developed by Sinai, another crucial figure in ergodic theory, who at the time was his graduate student (the entry on entropy on Scholarpedia was actually written by Sinai himself). Entropy in information theory, usually called Shannon entropy, was introduced by Claude Shannon in his 1948 paper Mathematical Theory of Communication.

[11]In metric spaces this definition of entropy was introduced by Bowen in 1971 and independently by Dinaburg in 1970. The definition of entropy via covers for any topological space already existed before. Equivalence between these two notions was proved by Bowen in 1971.

[12]Historically the definition of topological entropy via covers came first and is due to introduced in 1965 by Adler, Konheim and McAndrew. Their definition for topological dynamical systems is modelled on the definition of metric entropy by Kolmogorov and Sinai.

Example 2.4.1. **[Separated sets for the doubling map]** Let $f : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ be the doubling map, $f(x) = 2x \mod 1$. Let $\epsilon > 0$ and assume that $\epsilon < 1/4$. Find k such that $1/2^{k+1} < \epsilon \leq 1/2^k$. By assumption on ϵ, $k \geq 2$. Consider the set S_n of dyadic fractions with denominator 2^n, that is

$$S_n = \left\{ \frac{i}{2^n}, \quad 0 \leq i \leq 2^n - 1 \right\}. \tag{2.4}$$

Let us prove that the set S_{n-1+k} is (n, ϵ)-separated. Let $x, y \in S_{n-1+k}$, with $x \neq y$. We need to show that $d_n(x, y) \geq \epsilon$, that means that there exists $0 \leq l \leq n-1$ such that $d(f^l(x), f^l(y)) \geq \epsilon$. In an exercise, we saw that for any $u, v \in \mathbb{R}/\mathbb{Z}$

$$d(u, v) \leq \frac{1}{4} \quad \Rightarrow \quad d(f(u), f(v)) = 2d(u, v). \tag{2.5}$$

If there exists $0 \leq l \leq n-1$ such that $d(f^l(x), f^l(y)) \geq 1/4$, we are done (since $k \geq 2$, $1/4 \geq 1/2^k \geq \epsilon$). Otherwise, we can apply (2.5) repeatingly for $n-1$ times and get

$$d(f^{n-1}(x), f^{n-1}(y)) = 2^{n-1} d(x, y).$$

Since $x \neq y$ and $x, y \in S_{n-1+k}$, we have $d(x, y) \geq 1/2^{n-1+k}$, so that we get

$$d(f^{n-1}(x), f^{n-1}(y)) = 2^{n-1} d(x, y) \geq \frac{2^{n-1}}{2^{n-1+k}} = \frac{1}{2^k} \geq \epsilon.$$

This proves that S_{n-1+k} is $(n, \epsilon)-$separated. Note that the cardinality of S_{n-1+k} is 2^{n-1+k}.

Remark 2.4.1. One can show, using the assumption that X is compact, (n, ϵ)-separated sets exist and are *finite* (see Extra 1 in the Extras on Topological Entropy linked after the following class).

To determine how many different orbits can be recognized at resolution ϵ one can maximize the size of an $(n, \epsilon)-$separated set.

Definition 2.4.2. Let $Sep(f, n, \epsilon)$ (or simply $Sep(n, \epsilon)$ when there is no ambiguity on f) be the *maximal* (that is, the largest) cardinality of an $(n, \epsilon)-$separated set in X.

Clearly as n grows, the maximum number of $(n, \epsilon)-$separated points will grow. In our Example 2.4.1 with the doubling map, the cardinality of the $(n, \epsilon)-$separated set S_{n-1+k} that we exhibited grows as 2^{k+n-1}, thus exponentially in n.

Note that if a quantity grows exponentially, for example if $a_n = e^{\kappa n}$, the exponential rate of growth, or the exponent κ, can be obtained as

$$\kappa = \lim_{n\to\infty} \frac{\log e^{\kappa n}}{n} = \lim_{n\to\infty} \frac{\log a_n}{n}.$$

This is still true if the exponential growth is not purely exponential, but there are other subexponential factors, for example if $a_n = n^2 e^{\kappa n}$

$$\lim_{n\to\infty} \frac{\log a_n}{n} = \lim_{n\to\infty} \frac{\log n^2 e^{\kappa n}}{n} = \lim_{n\to\infty} \frac{2\log n}{n} + \lim_{n\to\infty} \frac{\log e^{\kappa n}}{n} = 0 + \kappa = \kappa.$$

More in general, if $a_n = f(n)e^{\kappa n}$ where $f(n)$ is *subexponential*, that is, $\lim_{n\to\infty} \log f(n)/n = 0$, the limit $\lim_{n\to\infty} \log a_n/n$ still gives κ.

Thus, to consider the exponential growth rate of $Sep(f, n, \epsilon)$, for any $\epsilon > 0$, consider the quantity

$$h_{top}(f, \epsilon) = \limsup_{n\to\infty} \frac{\log(Sep(f, n, \epsilon))}{n}.$$

We need to consider the *lim sup* because we do not know if the limit a priori exists. In all the examples that we will compute the limit actually exists. If we now change resolution, thus let $\epsilon \to 0$, it is again clear that $Sep(f, n, \epsilon)$ cannot decrease as ϵ tends to zero (as the resolution becomes finer and finer, one can distinguish at least the same orbits, and probabily more). Moreover one can show that the growth rate of distinguishable orbits will stay the same or increase, thus also $h_{top}(f, \epsilon)$ is not decreasing. Since a monotone function has a limit, the limit of $h_{top}(f, \epsilon)$ as ϵ tends to 0 exists.

Definition 2.4.3. [Topological entropy] The *topological entropy* $h_{top}(f, \epsilon)$ of a topological dynamical system $f : X \to X$ is given by

$$h_{top}(f) = \lim_{\epsilon \to 0} h_{top}(f, \epsilon) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log(Sep(f, n, \epsilon))}{n}.$$

Thus, $h_{top}(f, \epsilon)$ is the exponential growth rate of the maximum number of orbits of length n which are distinguishable with finite precision ϵ and $h_{top}(f)$, which is the limit for arbitrarly small ϵ, is the exponential growth rate of the maximum number of orbits of length n which are distinguishable with finite *but arbitrary* precision.

Example 2.4.2. Let f be again the doubling map. We showed that if $1/2^{k+1} < \epsilon \leq 1/2^k$ then the set S_{n+k} is an $(n, \epsilon)-$separated set. Thus the maximal cardinality $Sep(n, \epsilon)$ of an $(n, \epsilon)-$separated set is at least the cardinality of S_{n+k}, which is 2^{n+k}. Thus, remarking that here ϵ, and hence k, are fixed and the limit is taken only in n, we get

$$h_{top}(f, \epsilon) = \limsup_{n \to \infty} \frac{\log(Sep(n, \epsilon))}{n} \geq \limsup_{n \to \infty} \frac{\log 2^{n+k}}{n} = \limsup_{n \to \infty} \frac{(n + k) \log 2}{n} = \log 2.$$

This quantity is independent on ϵ. Thus

$$h_{top}(f) = \lim_{\epsilon \to 0} h_{top}(f, \epsilon) \geq \log 2.$$

This shows that the doubling map has *positive entropy.*

2.4.2 Entropy via spanning sets

While it is relatively easy to construct an $(n, \epsilon)-$separated set, it might be hard to prove that an $(n, \epsilon)-$separated set is maximal. Exhibiting an $(n, \epsilon)-$separated sets for each n gives a lower bound on topological entropy. To give upper bounds, it is easier to produce sets which are in some sense opposite to $(n, \epsilon)-$separated sets and are called $(n, \epsilon)-$spanning sets.

Definition 2.4.4. Let $\epsilon > 0$, $n \in \mathbb{N}$. A set $S \subset X$ is $(\epsilon, n)-$*spanning* if for all $x \in X$ there is an $y \in S$ such that $d_n(x, y) < \epsilon$.

In other words, a set S is $(\epsilon, n)-$spanning if any point of the space can be approximated with a point of S whose orbit up to time n is indistinguishable up to time n with finite resolution ϵ. Equivalently, S is $(\epsilon, n)-$spanning if and only if

$$X \subset \bigcup_{y \in S} B_{d_n}(y, \epsilon), \qquad \text{where } B_{d_n}(y, \epsilon) = \{x \in X, \text{ such that } d_n(x, y) < \epsilon\}.$$

One can show that in any compact metric space (X, d), for any $\epsilon > 0$ and $n \in \mathbb{N}$ there exist (n, ϵ)-spanning sets (see Extra 1 in the Extras on Topological Entropy linked after the following class).

Example 2.4.3. [**Spanning set for the doubling map**] Let f be again the doubling map and S_k the set of dyadic fractions with denominator 2^k, see (2.4). Let $1/2^{k+1} < \epsilon \leq 1/2^k$. Then S_{k+n} is $(\epsilon, n)-$spanning. Indeed, if $x \in [0, 1]$, there exists i such that

$$x \in \left[\frac{i}{2^{n+k}}, \frac{i+1}{2^{n+k}}\right], \qquad \text{where} \quad 0 \leq i \leq 2^{n+k} - 1.$$

Then, if $y \in S_{n+k}$ is one of the endpoints of the interval, we have

$$d(x, y) \leq \frac{1}{2^{n+k}} \quad \Rightarrow \quad d(f^j(x), f^j(y)) \leq \frac{2^j}{2^{n+k}} \leq \frac{2^{n-1}}{2^{n+k}} = \frac{1}{2^{k+1}} < \epsilon, \quad \text{for all } \ 0 \leq j < n.$$

Thus, $d_n(x, y) < \epsilon$.

Note that the cardinality of S_{n+k+1} is 2^{n+k+1}, so it grows exponentially in n.

This time it makes sense to see what is the smallest possible (ϵ, n)-spanning set.

Definition 2.4.5. Let $Span(f, n, \epsilon)$ (or $Span(n, \epsilon)$ if there is no ambiguity on f) be the *minimal* (that is, the smallest) cardinality of an (n, ϵ)-spanning set in X.

In other words, $Span(f, n, \epsilon)$ is the minimum number of initial conditions needed to approximate with resolution ϵ all orbits in the space up to time n. We can consider the exponential growth rate in n of $Span(f, n, \epsilon)$ for fixed ϵ and then take the limit as ϵ tends to zero. It turns out that this gives the same result than using $Sep(f, n, \epsilon)$ and yields again the topological entropy:

Theorem 2.4.1. *For any topological dynamical system $f : X \to X$ we have*

$$h_{top}(f) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log(Span(f, n, \epsilon))}{n}.$$

Note that since we have already defined $h_{top}(f)$ in terms of (n, ϵ)-separated sets, so by definition

$$h_{top}(f) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log(Sep(f, n, \epsilon))}{n}.$$

Thus, the content of the theorem, is that the two limsups, obtained using separated and spanning sets respectively, are the same, that is:

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log(Sep(f, n, \epsilon))}{n} = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log(Span(f, n, \epsilon))}{n}.$$

The proof of the theorem is given below. Let us first remark that the theorem states that to compute topological entropy, we can either use maximal separated sets or minimal spanning sets. While the expression with $Sep(n, \epsilon)$ is useful to give *lower bounds* on $h_{top}(f)$, the expression with $Span(n, \epsilon)$ is useful to give *upper bounds* on $h_{top}(f)$. Indeed, once we find an (n, ϵ)-separated set, we have a *lower* bound on on the *maximal* cardinality of (n, ϵ)-separated sets, that is on $Sep(n, \epsilon)$. Conversely, once we find an (n, ϵ)-spanning set, we have an *upper* bound on the *minimal* cardinality of (n, ϵ)-spanning sets, that is on $Sep(n, \epsilon)$. When the upper bound and the lower bound coincide, the common value is the topological entropy.

Let us show two applications of Theorem 2.4.1.

Example 2.4.4. [Entropy of the doubling map] Consider the doubling map f. Let $\epsilon < 1/4$ and let k be such that $1/2^{k+1} < \epsilon \leq 1/2^k$. Since S_{n+k} is an (n, ϵ)-spanning set of cardinality 2^{n+k}, the minimal cardinality $Span(f, n, \epsilon)$ of an (n, ϵ)-spanning set is at most 2^{n+k} so

$$\limsup_{n \to \infty} \frac{\log(Span(f, n, \epsilon))}{n} \leq \limsup_{n \to \infty} \frac{\log 2^{n+k}}{n} = \lim_{n \to \infty} \frac{(n + k) \log 2}{n} = \log 2.$$

Note that k is fixed when taking the limit in n. Taking now the limit in ϵ of a quantity which is independent on ϵ, by Theorem 2.4.1 we have $h_{top}(f) \leq \log 2$. Since we have already shown using (n, ϵ)-separating sets that $h_{top}(f) \geq \log 2$, we conclude that $h_{top}(f) = \log 2$.

Example 2.4.5. [Entropy of rotations] Let $R_\alpha : S^1 \to S^1$ be a rotation. Fix ϵ and N such that $1/N \leq \epsilon$. Let us consider N equispaced points on S^1:

$$S = \{ e^{2\pi i \frac{m}{N}}, \qquad m = 0, 1, \ldots, N - 1\},$$

so that the arc length distance between two successive points in S is $1/N$. Clearly S is $(1, \epsilon)-$spanning, since for any $z \in S^1$ there exists $z' \in S$ such that $d(z, z') \leq 1/N \leq \epsilon$. Let us show that the same set S is also (n, ϵ)-spanning for any $n \in \mathbb{N}$. It is enough to remark that for any $j \in \mathbb{N}$, since R_α preserves the arc length distance,

$$d(R^j_\alpha(z), R^j_\alpha(z')) = d(z, z') \leq 1/N \leq \epsilon. \tag{2.6}$$

Thus, for any $n \in \mathbb{N}$, $d_n(z, z') \leq \epsilon$. This shows that S is also (n, ϵ)-spanning for any $n \in \mathbb{N}$.

Hence, the minimal cardinality $Span(f, n, \epsilon)$ is less than the cardinality of S, which is N, independently on n. Thus, for any ϵ, if we choose $1/N \leq \epsilon$ and keep N fixed as $n \to \infty$, we get

$$\limsup_{n\to\infty} \frac{\log(Span(f, n, \epsilon))}{n} \leq \limsup_{n\to\infty} \frac{\log N}{n} = 0.$$

Notice that the groth rate is non negative, so that if it is less than 0, i it indeed 0. By the Theorem 2.4.1

$$0 \leq h_{top}(R_\alpha) = \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{\log(Span(f, n, \epsilon))}{n} = 0.$$

This shows that rotations have zero topological entropy.

Remark 2.4.2. Note that the only property of rotations used in the proof their entropy is zero is that rotations are isometries, so that (2.6) holds. More generally, any isometry of a compact metric space has zero entropy.

Exercise 2.4.1. Let (X, d) be a compact metric space and $f : X \to X$ a continuous isometry.

(a) Show that for any $\epsilon > 0$ there exists a $(1, \epsilon)$ spanning set.

[*Hint*: Use the definition of compactness by covers, see §2.1.1]

(b) Prove that $h_{top}(f) = 0$.

Let us know prove Theorem 2.4.1. We will use the following:

Lemma 2.4.1. *For any topological dynamical system $f : X \to X$ on a metric space (X, d) we have:*

(1) $Span(n, \epsilon) \leq Sep(n, \epsilon)$;

(2) $Sep(n, 2\epsilon) \leq Span(n, \epsilon)$.

Proof. Let us prove (1) first. Let S be a (n, ϵ)-separated set of maximal cardinality, so that $Sep(n, \epsilon) = \mathrm{Card}(S)$. We claim that S is also (n, ϵ)-spanning. Indeed, given $x \in X \setminus S$, since S has maximal cardinality, the set $S \cup \{x\}$ is not (n, ϵ)-separated. Thus, since for any distinct $y, y' \in S$ we have $d_n(y, y') \geq \epsilon$ (S is (n, ϵ)-separated), but $S \cup \{x\}$ is not (n, ϵ)-separated, there must exists $y \in S$ such that $d_n(x, y) < \epsilon$. This shows that S is (n, ϵ)-spanning. Thus, the minimal cardinality

$$Span(n, \epsilon) \leq \mathrm{Card}(S) = Sep(n, \epsilon).$$

Let us now prove (2). Let S be a (n, ϵ)-spanning set of minimal cardinality, so that $Span(n, \epsilon) = \mathrm{Card}(S)$. By definition of (n, ϵ)-spanning, we know that

$$X \subset \bigcup_{y\in S} B_{d_n}(y, \epsilon).$$

Let S' be any $(n, 2\epsilon)-$separated set. We claim that no two distinct points $x_1 \neq x_2$ of S' can belong to the same d_n-ball. Indeed, if both $x_1, x_2 \in B_{d_n}(y, \epsilon)$ for some $y \in S$, by triangle inequality

$$d_n(x_1, x_2) \leq d_n(x_1, y) + d_n(x_2, y) < \epsilon + \epsilon = 2\epsilon,$$

which contradicts the fact that S' is $(n, 2\epsilon)$-separated. So the number of points in S' cannot be more than the number of balls, that is the cardinality of S. Let us now choose S' the $(n, 2\epsilon)$-separated set of maximal cardinality. Then

$$Sep(n, 2\epsilon) = \mathrm{Card}(S') \leq \mathrm{Card}(S) = Span(n, \epsilon).$$

□

Proof of Theorem 2.4.1. By the Lemma, for any $\epsilon > 0$ and $n \in \mathbb{N}$, we have

$$Sep(n, 2\epsilon) \leq Span(n, \epsilon) \leq Sep(n, \epsilon).$$

Since the inequalities hold for any $n \in \mathbb{N}$, we also have

$$\limsup_{n\to\infty} \frac{\log(Sep(n, 2\epsilon))}{n} \leq \limsup_{n\to\infty} \frac{\log(Span(n, \epsilon))}{n} \leq \limsup_{n\to\infty} \frac{\log(Sep(n, \epsilon))}{n},$$

where the first inequality follows from (2) and the second inequality follows from (1). By sandwitch Theorem (also called pinching theorem or two policemen theorem), if we now take the limit as ϵ tends to zero, both the left and the righ hand side converge by definition to $h_{top}(f)$, so

$$h_{top}(f) \leq \lim_{\epsilon\to 0} \limsup_{n\to\infty} \frac{\log(Span(f, n, \epsilon))}{n} \leq h_{top}(f).$$

Thus, we conclude as desired that the limit as ϵ tends to zero of the exponential growth rate of $Span(f, n, \epsilon)$ is equal to $h_{top}(f)$. □

Extra: Compactness and existence of separated and spanning sets.

We conclude by recalling that to define topological entropy, we assumed thoughout this section that (X, d) is a *compact* metric space. We also claimed that both (n, ϵ)-sets and (n, ϵ)-spanning sets exist and that (n, ϵ)-separated sets have finite cardinality, so that both $Sep(n, \epsilon)$ and $Span(n, \epsilon)$ are well defined. Compactness is used to prove these two claims, as shown in the following exercise.

Exercise 2.4.2. Let (X, d) be a compact metric space. Let $n \in \mathbb{N}$ and $\epsilon > 0$.

(a) Show that there exists an (n, ϵ)-spanning set $S \subset X$;
[*Hint*: Use the definition of compactness by covers, see §2.1.1]

(b) Show that there exists an (n, ϵ)-separated set.
[*Hint*: You do not need any assumption here.]

(c) Show that the cardinality of any (n, ϵ)-separated set is always finite and bounded above uniformely, that is, there exists a constant $C > 0$ (which depends on n and ϵ) such that

$$\sup\{Card(S), \text{ where } S \text{ is an } (n, \epsilon)\text{-separated set }\} \leq C.$$

[*Hint*: You can use part of the proof of Lemma 2.3.1.]

(d) Conclude that $Sep(n, \epsilon)$ and $Span(n, \epsilon)$ are well-defined and finite.

2.5 More on Topological Entropy: toral automorphisms and entropy via covers

Using both equivalent definitions of entropy via (n, ϵ)-separated and (n, ϵ)-spanning sets one can often compute the topological entropy of a topological dynamical system $f : X \to X$ on a metric space (X, d). Let us summarize how to use them together.

One strategy to compute topological entropy is the following:

1. If for any fixed $\epsilon > 0$ you can construct sets S_n, for each $n \in \mathbb{N}$, which are (n, ϵ)-*separating* and whose exponential growth rate is $\underline{h}$, then, since $Sep(n, \epsilon) \geq Card(S_n)$, you can conclude that $h_{top}(f) \geq \underline{h}$;
2. If for any fixed $\epsilon > 0$ you can construct sets S_n, for each $n \in \mathbb{N}$, which are (n, ϵ)-*spanning* and whose exponential growth rate is $\overline{h}$, then, since $Span(n, \epsilon) \leq Card(S_n)$, you can conclude that $h_{top}(f) \leq \overline{h}$;
3. If the exponential growth rates $\underline{h}$ and $\overline{h}$ in the two previous points the same, then you can conclude that $h_{top}(f) = \underline{h} = \overline{h}$.

We have already seen examples of this principle in the previous section. We will now compute the topological entropy of hyperbolic toral automorpshisms using this strategy.

2.5.1 Topological entropy of hyperbolic toral automorpshisms

We defined hyperbolic toral automorphisms in § 1. 6. Recall that they are maps on $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ defined by a matrix A with integer entries, determinant ± 1 and no eigenvalue of modulus 1. The cat map is an example of hyperbolic toral automorphisms.

Theorem 2.5.1. *Let $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ be a hyperbolic toral automorphisms, with eigenvalues $\lambda_1 > 1$ and $\lambda_1 < 1$. The topological entropy of f_A is given by*

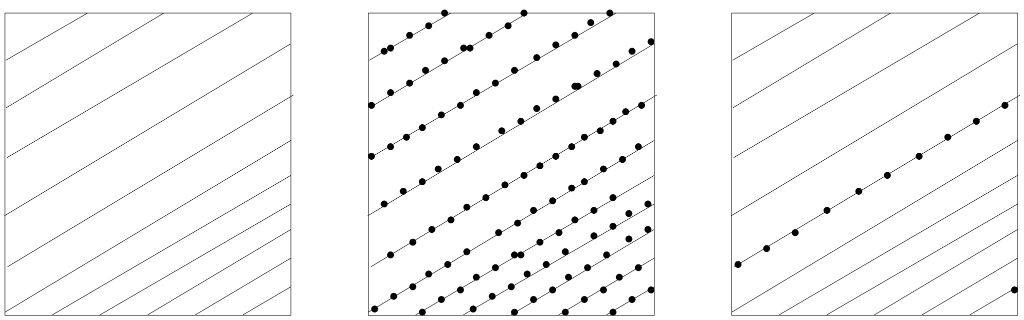
$$h_{top}(f_A) = \log \lambda_1.$$

Note that only the expanding eigenvalue λ_1 enters in the entropy. The direction which is contracted does not play any role in the value of the entropy. This is a general phenomenon: topological entropy sees only points which *diverge* exponentially and hence only the *expanding* directions matter.

Let us denote by $\underline{v}_1$ and $\underline{v}_2$ the two eigenvalues of A, that is

$$A\ \underline{v}_1 = \lambda_1\ \underline{v}_1, \qquad A\ \underline{v}_2 = \lambda_2\ \underline{v}_2,$$

and let us assume that they are renormalized so that they have unit length: $\|\underline{v}_1\| = \|\underline{v}_2\| = 1$.



(a) Lines in direction $\underline{v}_1$ (b) An (n,ϵ)-spanning set (c) An (n,ϵ)-separating set

Figure 2.3: (n,ϵ)-spanning and (n,ϵ)-separating sets fo $f_A : \mathbb{T}^2 \to \mathbb{T}^2$.

Proof. Let us first construct (n,ϵ)-spanning sets for f_A. Fix $\epsilon > 0$ and choose $N \in \mathbb{N}$ such that $1/N \leq \epsilon/2$. Draw inside the unit square $[0,1) \times [0,1)$ segments of lines in direction of $\underline{v}_1$, which cross the square fully and have spacing $1/N$ both on the horizontal and on the vertical side, as in Figure 2.3(a). Note that the distance between two successive lines is less than $1/N \leq \epsilon/2$, since the distance between lines is the side of a right triangle whose hypothenus is the spacing $1/N$ between lines.

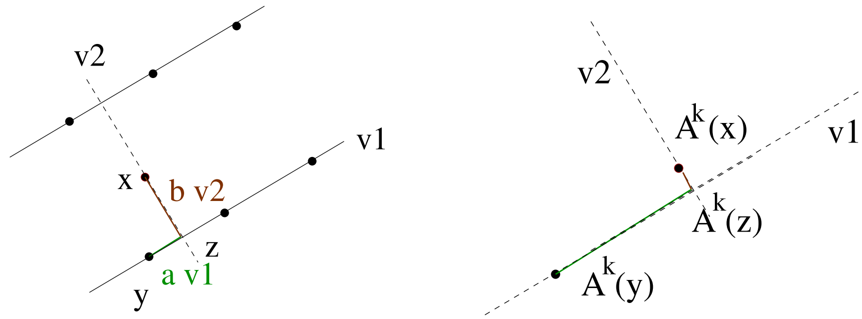
On each line, consider points whose spacing is $\epsilon/2\lambda_1^{n-1}$. The reason of this choice will be clear later. Let $S \subset \mathbb{T}^2$ be the set which consists of the union over the lines of these points (see Figure 2.3(b)). Let us prove that S is (n,ϵ)-spanning.

Let $\underline{x} \in \mathbb{T}^2$ (note that a point on $\mathbb{T}^2$ has two coordinates (x_1, x_2) and we can think of it as a vector, that is why why write $\underline{x}$). Let $\underline{y} \in S$ be the closest point to $\underline{x}$ among points in S. We can write between $\underline{x} - \underline{y}$ as sum of a vector proportional to $\underline{v}_2$ and a vector proportional to $\underline{v}_1$ (see Figure 2.4(a)). Note that the distance between $\underline{x}$ and $\underline{y}$ along the $\underline{v}_1$ direction is less than the spacing of points on each line, which is $\frac{\epsilon}{2\lambda_1^{n-1}}$ (see Figure 2.4(a)); the distance between $\underline{x}$ and $\underline{y}$ in direction $\underline{v}_2$ is less than the distance between $\underline{x}$ and a line, which is less than the distance $\epsilon/2$ between lines. Thus can write (2.7).

$$\underline{x} = \underline{y} + a\underline{v}_1 + b\underline{v}_2, \qquad \text{where} \quad |a| \leq \frac{\epsilon}{2\lambda_1^{n-1}}, \quad |b| \leq \frac{\epsilon}{2}. \tag{2.7}$$

Let us now compute the distance between the orbits of $\underline{x}$ and $\underline{y}$ under f_A. Recall that f_A^k is obtained by first acting linearly by the matrix A^k and then taking the result modulo 1 (which correspond to cutting and pasting the affine image of $[0,1]^2$ under A to map it back again a unit square.) Since A acts linearly, for each iterate $k \in \mathbb{N}$

$$A^k(\underline{x} - \underline{y}) = A^k(a\underline{v}_1 + b\underline{v}_2) = aA^k\underline{v}_1 + bA^k\underline{v}_2 = a\lambda_1^k\underline{v}_1 + b\lambda_2^k\underline{v}_2,$$



(a) $\underline{x}$ and $\underline{y}$

(b) $A^k(\underline{x})$ and $A^k(\underline{y})$

Figure 2.4: Distance from points of the (n, ϵ)-spanning set.

where in the latter equality we used that $\underline{v}_1$, $\underline{v}_2$ are eigenvectors. Thus, since the operation of cutting and pasting does not increase the distances (actually, if the distance is less than 1, it is preserved when taking the result modulo $\mathbb{Z}^2$), by setting $\underline{z} = \underline{y} + av_1 = \underline{x} - bv_2$ (see again Figure 2.4(a)), we have by triangle inequality that

$$d(f_A^k(\underline{x}), f_A^k(\underline{y})) \leq d(f_A^k(\underline{x}), f_A^k(\underline{z})) + d(f_A^k(\underline{z}), f_A^k(\underline{y})) \leq |a|\lambda_1^k + |b|\lambda_2^k.$$

Using now the bounds on $|a|$ and $|b|$ from (2.4(a)) and then recalling that $k \leq n-1$ and that $\lambda_2 < 1$, we get that

$$|a|\lambda_1^k + |b|\lambda_2^k \leq \frac{\epsilon}{2\lambda_1^{n-1}}\lambda_1^k + \frac{\epsilon}{2}\lambda_2^k \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus $d(f_A^k(\underline{x}), f_A^k(\underline{y})) \leq \epsilon$ for each $0 \leq k \leq n-1$, which means that $d_n(\underline{x}, \underline{y}) \leq \epsilon$ and concludes the proof that S is (n, ϵ)-spanning.

Let us bound the cardinality of S. Let L be the length of the longest line segment. Then, we can bound the number of points in each line by L divided by the spacing and since there are at most $2N$ lines, we get

$$Card(S) \leq (\text{number of lines })(\text{points on each line}) \leq 2N\left(\frac{L}{\frac{\epsilon}{2\lambda_1^{n-1}}}\right) = \frac{4NL}{\epsilon}\lambda_1^{n-1}.$$

Let us compute the exponential growth rate of $Span(n, \epsilon)$, recalling that N and ϵ are fixed and only n grows:

$$\limsup_{n\to\infty} \frac{\log(Span(n,\epsilon))}{n} \leq \limsup_{n\to\infty} \frac{\log Card(S)}{n} \leq \limsup_{n\to\infty} \frac{\log(\frac{4NL}{\epsilon}\ \lambda_1^{n-1})}{n}$$
$$= \limsup_{n\to\infty} \frac{\log\frac{4NL}{\epsilon}}{n} + \limsup_{n\to\infty} \frac{(n-1)\log\lambda_1}{n} = 0 + \log\lambda_1.$$

Thus, using the Theorem which expresses the topological entropy using the growth rate of $Span(n, \epsilon)$, we have $h_{top}(f_A) \leq \log\lambda_1$.

Let us now construct (n, ϵ)*-separating sets for* f_A. Fix an $\epsilon < 1/2$. To construct an (n, ϵ) separated set S, it is enought to consider now only one of the lines, for example the line of maximal length, whose length will be denoted by L and let $S \subset \mathbb{T}^2$ be the set which consists of points on the line whose spacing is ϵ/λ_1^{n-1}, as in Figure 2.3(c). Let us prove that S is (n, ϵ)-separating, that is, that, given two distinct points $\underline{x}, \underline{y}$ in S, there exists $0 \leq l \leq n-1$ such that $d(f_A^l(\underline{x}), f_A^l(\underline{y})) \geq \epsilon$, i.e. the two points can be distinguished with resolution ϵ in time n. Let us check that the closest points, that is two consecutive points on the line, can be distinguished. If $\underline{x}$ and $\underline{y}$ are consecutive points, they can be written as

$$\underline{y} = \underline{x} + \frac{\epsilon}{\lambda_1^{n-1}}\underline{v}_1.$$

Since, reasoning as before, we have

$$A^k(\underline{x} - \underline{y}) = \frac{\epsilon}{\lambda_1^{n-1}}A^k\underline{v}_1 = \frac{\epsilon}{\lambda_1^{n-1}}\lambda_1^k\underline{v}_1,$$

and the operation of cutting and pasting to consider the result $\mod \mathbb{Z}^2$ does not change distances until the distance is at least 1, we have, for $k = n-1$

$$d(f_A^{n-1}(\underline{x}), f_A^{n-1}(\underline{y})) = \frac{\epsilon}{\lambda_1^{n-1}}\lambda_1^{n-1} = \epsilon.$$

Thus, $d_n(\underline{x}, \underline{y}) \geq \epsilon$. If one considers another pair of distinct points $\underline{x}', \underline{y}'$ on the line, $d(\underline{x}', \underline{y}') \geq d(\underline{x}, \underline{y})$ and moreover, since A^k acts linearly, $d(A^k\underline{x}', A^k\underline{y}') \geq d(A^k\underline{x}, A^k\underline{y})$ for any k. Hence there will also be a $0 \leq k < n$ such that

$$d(f_A^k(\underline{x}'), f_A^k(\underline{y}')) = d(A^k(\underline{x}'), A^k(\underline{y}')) \geq \epsilon.$$

Thus we conclude that $d_n(\underline{x}', \underline{y}')$ and that S is (n, ϵ)-separating.

Since cardinality of S is the number of points on the line is at least

$$Card(S) \geq \frac{L}{\frac{\epsilon}{\lambda_1^{n-1}}} = \frac{L}{\epsilon}\lambda_1^{n-1}.$$

Using this time the definition of $h_{top}(f_A)$ with $Sep(n, \epsilon)$ we get

$$\begin{aligned} h_{top}(f_A, \epsilon) &= \limsup_{n\to\infty} \frac{\log(Sep(n,\epsilon))}{n} \geq \limsup_{n\to\infty} \frac{\log Card(S)}{n} \geq \limsup_{n\to\infty} \frac{\log(\frac{L}{\epsilon}\ \lambda_1^{n-1})}{n} \\ &= \limsup_{n\to\infty} \frac{\log \frac{L}{\epsilon}}{n} + \limsup_{n\to\infty} \frac{(n-1)\log\lambda_1}{n} = 0 + \log\lambda_1. \end{aligned}$$

Thus, $h_{top}(f_A) \geq \log\lambda_1$. Combining with $h_{top}(f_A) \leq \log\lambda_1$, we proved $h_{top}(f_A) = \log\lambda_1$. □

Remark 2.5.1. If f and g are topologically conjugate, then

$$h_{top}(f) = h_{top}(g).$$

Thus, topological entropy is what is called an *invariant* under topological conjugacy.

To prove Remark 2.5.1 one can use an alternative definition of entropy using covers (see below and see Exercise).

2.5.2 Topological entropy via covers

The combination of (n, ϵ)-separating and (n, ϵ)-spanning sets allows to calculate the entropy in many examples, but in some examples another alternative definition, topological entropy using covers, is easier to apply.

Let (X, d) be a compact metric space and $f : X \to X$ a topological dynamical system. Let us give an alternative way to calculate topological entropy using covers.

Given a set U, the diameter of U with respect to the metric d is

$$diam_d(U) = \sup_{x,y\in U} d(x, y).$$

For example, the ball $B_d(x, \epsilon)$ has diameter 2ϵ. Let d_n be as usual be the metric

$$d_n(x, y) = \max_{0\leq k<n} d(f^k(x), f^k(y)).$$

Let $\mathscr{U} = \{U_1, U_2, \ldots, U_n\}$ be a cover of X with open sets with d_n-diameter less than ϵ, that is $diam_{d_n}(U_i) \leq \epsilon$ for any $1 \leq i \leq n$. Note that since X is compact we can assume that the cover consists of finitely many open sets (since from any open cover we can extract a finite subcover).

Example 2.5.1. If S is $(n, \epsilon/2)-$spanning, since (see the definition of $(n, \epsilon/2)-$spanning),

$$X \subset \bigcup_{y\in S} B_{d_n}(y, \frac{\epsilon}{2}),$$

and the balls have d_n-diameter less than ϵ, the collection

$$\left\{ B_{d_n}\left(y, \frac{\epsilon}{2}\right), \qquad y \in S \right\}$$

is such a cover.

Example 2.5.2. Let $\mathscr{U} = \{U_1, U_2, \dots, U_N\}$ be a cover with open sets with d-diameter less than ϵ. Consider the following sets defined using $f : X \to X$:

$$\left\{ U_{k_0} \cap f^{-1}(U_{k_1}) \cap \dots f^{-(n-1)}(U_{k_{n-1}}), \quad U_{k_0}, U_{k_1}, \dots, U_{k_{n-1}} \in \mathscr{U} \right\}.$$

This collection is a *cover*, since given any $x \in X$, since $\mathscr{U}$ is a cover, $x \in U_{l_0}$ for some $0 \le l_0 \le N$, $f(x) \in U_{l_1}$ for some $0 \le l_1 \le N$ and so on up to $f^{n-1}(x) \in U_{l_{n-1}}$ for some $0 \le l_{n-1} \le N$, so that

$$x \in U_{l_0} \cap f^{-1}(U_{l_1}) \cap \dots f^{-(n-1)}(U_{l_{n-1}}).$$

(Note that here $l_0, l_1, \dots, l_{n-1}, \dots$ is a symbolic coding of the itinerary of the orbit $\mathcal{O}_f^+(x)$ with respect to the cover $\{U_1, U_2, \dots, U_N\}$. Moreover, the d_n-diameter of each set is less than ϵ, since if $x, y \in U_{k_0} \cap f^{-1}(U_{k_1}) \cap \dots f^{-(n-1)}(U_{k_{n-1}})$, then

$$x, y \in U_{k_0}, \quad f(x), f(y) \in f(U_{k_1}), \qquad \dots, \qquad f^{n-1}(x), f^{n-1}(y) \in f^{n-1}(U_{k_{n-1}})$$

and since $diam_d(U_k) \le \epsilon$ for each $0 \le k \le N$,

$$d(x, y) \le \epsilon,\ d(f(x), f(y)) \le \epsilon,\ \dots, d(f^{n-1}(x), f^{n-1}(y)) \le \epsilon \quad \Rightarrow \quad d_n(x, y) \le \epsilon.$$

Definition 2.5.1. Let $Cov(f, n, \epsilon)$ (or simply $Cov(n, \epsilon)$ is the map f is clear from the context) be the minimal cardinality of open covers $\mathscr{U} = \{U_1, U_2, \dots, U_N\}$ with sets whose diameter in the d_n metric satisfy $diam_{d_n}(U_i) \le \epsilon$.

Remark 2.5.2. One can show that the exponential growth rate of $Cov(n, \epsilon)$, that is

$$\lim_{n\to\infty} \frac{\log(Cov(f, n, \epsilon))}{n},$$

exists (see Extra), so there is no need to use lim inf to define it.

Theorem 2.5.2. *Let (X, d) be a compact metric space and $f : X \to X$ a topological dynamical systems. The topological entropy $h_{top}(f)$ can be computed using $Cov(f, n, \epsilon)$ and is given by*

$$h_{top}(f) = \lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{\log(Cov(f, n, \epsilon))}{n}.$$

We will see an example of computation of topological entropy using this definition in the next sections.

Let us now give the proof of Theorem 2.5.2, which follows immediately from the following Lemma (which is similar to Lemma 2.3.1).

Lemma 2.5.1. *For any topological dynamical system $f : X \to X$ on a metric space (X, d) we have:*

$$Span(n, \epsilon) \le Cov(n, \epsilon) \le Span(n, \epsilon/2). \tag{2.8}$$

Proof. In Example 2.5.1, we showed that if S in $(n, \epsilon/2)-$spanning, that there it gives a cover with balls of d_n-diameter less than ϵ and of cardinality $Card(S)$. Thus, if S is such that $Span(n, \epsilon/2) = Card(S)$, this shows that the minimal cardinality $Cov(n, \epsilon)$ is less than the cardinality of S, thus $Cov(n, \epsilon) \le Span(n, \epsilon/2)$, which is the second inequality.

For the other inequality, note that if $\mathscr{U}$ is a cover with open sets of d_n-diameter less than ϵ, then each open set U is contained in a ball $B_{d_n}(x, \epsilon)$ centered at any point $x \in U$. Indeed, if $x \in U$, any other point $y \in U$ is such that $d_n(x, y) \le \epsilon$ by definition of diameter. Thus, we can find a cover with d_n-balls of radius ϵ of the cardinality $Card(\mathscr{U})$ and the collection S of the centers of the balls give an $(n, \epsilon)-$spanning set of cardinality $Card(\mathscr{U})$. If we choose $\mathscr{U}$ such that $Card(\mathscr{U}) = Cov(n, \epsilon)$, this shows that the minimal cardinality $Span(n, \epsilon)$ is at less than $Card(S) = Card(\mathscr{U})$. So we have $Span(n, \epsilon) \le Cov(n, \epsilon)$. ◻

Proof of Theorem 2.5.2. For each $\epsilon > 0$ and each $n \in \mathbb{N}$, by Lemma 2.5.1, we have that

$$Span(n, \epsilon) \leq Cov(n, \epsilon) \leq Span(n, \epsilon/2).$$

Thus, for every $\epsilon > 0$ we have that

$$\limsup_{n\to\infty} \frac{\log Span(n,\epsilon)}{n} \leq \limsup_{n\to\infty} \frac{\log Cov(n,\epsilon)}{n} = \lim_{n\to\infty} \frac{\log Cov(n,\epsilon)}{n} \leq \limsup_{n\to\infty} \frac{\log Span(n,\epsilon/2)}{n}.$$

Since as $\epsilon \to 0$, by Theorem 2.3.1 in the previous section the right an the left hand side tend to $h_{top}(f)$, we have, by the sandwitch theorem, that

$$h_{top}(f) = \lim_{\epsilon\to 0} \lim_{n\to\infty} \frac{\log Cov(n,\epsilon)}{n}.$$

$\square$

See the Extras for this section for for the proof of the existence of the exponential growth rate of $Cov(f, n, \epsilon)$ and see Exercise 2.5.3 in the Extras for the proof of Remark 2.5.1 on topological entropy as a conjugacy invariant.

Extra: Existence of the exponential growth rate of $Cov(n, \epsilon)$

Let $f : X \to X$ be a topological dynamical system on a compact metric space (X, d). Recall that we defined in Section 2.5. the quantity $Cov(n, \epsilon)$ as the smallest cardinality of an open cover of X with open sets with diameter less than ϵ in the d_n metric. Let us prove Remark 2.5.2, that is let us show that the following limit exists:

$$\lim_{n\to\infty} \frac{\log(Cov(f, n, \epsilon))}{n}$$

Let us show that $Cov(n, \epsilon)$ is a subadditive sequence in n, that is it satisfies

$$Cov(n + m, \epsilon) \leq Cov(n, \epsilon)Cov(m, \epsilon), \qquad \text{for all } n, m \in \mathbb{N}. \tag{2.9}$$

Let $\mathscr{U}$ be a cover with open sets of d_n-diameter less than ϵ and $Card(\mathscr{U}) = Cov(n, \epsilon)$ and let $\mathscr{V}$ be a cover with open sets of d_m-diameter less than ϵ and $Card(\mathscr{U}) = Cov(\epsilon, m)$. Then, if $U \in \mathscr{U}$ and $V \in \mathscr{V}$, then the set

$$U \cap f^{-n}(V)$$

is open and has $d_{n+m}-$diameter less than ϵ, since if $x, y \in U$, $\max_{0\leq i\leq n-1} d(f^i(x), f^i(y)) \leq \epsilon$ and since $f^n(x)$ and $f^n(y)$ are both in V, also

$$\max_{0\leq i\leq m-1} d(f^i(f^n(x)), f^i(f^n(y)) \leq \epsilon \quad \Leftrightarrow \quad \max_{n\leq i\leq m+n-1} d(f^i(x)), f^i(y) \leq \epsilon,$$

so that $d_{n+m}(x, y) \leq \epsilon$.

Thus, if we consider the collection of all the sets of the form

$$\mathscr{W} = \{U \cap f^{-n}(V), \qquad U \in \mathscr{U}, \quad U \in \mathscr{U}\}$$

it is an open cover with $d_{n+m}-$diameter less than ϵ which consists of at most $Card(\mathscr{U})Card(\mathscr{V})$ sets (less if some intersection is empty). Thus, the minimal cardinality $Cov(n + m, \epsilon)$ satisfies

$$Cov(n + m, \epsilon) \leq Card(\mathscr{W}) \leq Card(\mathscr{U})Card(\mathscr{V}) = Cov(n, \epsilon)Cov(\epsilon, m).$$

To conclude the proof, one can use the following exercise that shows that subadditive sequences always have a limit[13]:

[13] You can find the solution for example in Katok-Hasselblatt book, Lemma 4.3.7.

*** Exercise 2.5.1.** *If* $(a_n)_{n\in\mathbb{N}}$ *is a* subadditive sequence*, that is*

$$0 \le a_{m+n} \le a_n + a_m, \qquad \textit{for all } n, m \in \mathbb{N},$$

then the following limit exists

$$\lim_{n\to\infty} \frac{a_n}{n} \qquad \textit{and} \qquad \lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\to\infty} \frac{a_n}{n}.$$

If we consider now the sequence $a_n = \log Cov(n, \epsilon)$, property (2.9) becomes

$$a_{n+m} = \log Cov(n+m, \epsilon) \le \log Cov(n, \epsilon) Cov(\epsilon, m) \le \log Cov(n, \epsilon) + \log Cov(\epsilon, m) = a_n + a_m$$

for any $n, m \in \mathbb{N}$. Thus a_n is subadditive and by Exercise 2.5.1, the limit

$$\lim_{n\to\infty} \frac{\log Cov(n, \epsilon)}{n}$$

exists.

Extra: Some properties of topological entropy

The following properties of topological entropies can be proved using the definition of entropy via covers.

Exercise 2.5.2. Let (X, d_X), (Y, d_Y) be two compact metric space and $f : X \to X$, $g : Y \to Y$ two topological dynamical systems. Using the definition of topological entropy via covers, show that: Let (X, d) be a compact metric space and let $f : X \to X$ be a homeomorphism. Show that $h_{top}(f) = h_{top}(f^{-1})$.

[*Hint*: Use the definition of topological entropy via covers. Show that if d_n^f and $d_n^{f^{-1}}$ are the metrics using to define entropy, one has

$$d_n^f(x, y) = d_n^{f^{-1}}(f^{n-1}(x), f^{n-1}(y)).$$

and use covers by sets of the form $f^{n-1}(U_i)$ were U_i belong to a cover of X.]

Exercise 2.5.3. Let (X, d_X), (Y, d_Y) be two compact metric space and $f : X \to X$, $g : Y \to Y$ two topological dynamical systems. Using the definition of topological entropy via covers, show that if f is topologically conjugate to g (that is there is a topological conjugacy $\psi : Y \to X$) then $h_{top}(f) = h_{top}(g)$;

[*Hint*: Since X is compact, you can prove (or just assume and use) that for any $\epsilon > 0$ there is $\delta > 0$ such that for any $y_1, y_2 \in Y$, if $d_Y(y_1, y_2) < \delta$, then $d_X(\psi(y_1), \psi(y_2)) < \epsilon$. Consider covers of X by sets of the form $\psi(U_i)$ were U_i belong to a cover of Y.]

2.6 Shift spaces and coding

In various of the examples of dynamical systems that we saw so far (the doubling map, the baker map and the Gauss map) we described an orbit through its *itinerary*. In this section we introduce some symbolic spaces that allow to describe the dynamics of more maps using itineraries.

Let us decompose the space X in finitely many pieces (more in general, one could consider countably many pieces), that is

$$X = P_1 \cup P_2 \cup \cdots \cup P_N.$$

If P_i are pairwise disjoint, we say that $\{P_1, \ldots, P_N\}$ is a finite partition of X. Let $x \in X$. Since P_i cover X, for each $i \in \mathbb{N}$ there exists $1 \leq a_i \leq N$ such that $f^i(x) \in P_{a_i}$. The sequence $a_0, a_1, \ldots, a_n, \ldots$, is the itinerary of x with respect to $\{P_1, \ldots, P_N\}$.

We can *code* the (forward) orbit $\mathcal{O}_f^+(x)$ with the sequence $\underline{a} = (a_i)_{i=0}^{\infty}$. The sequence belongs to

$$\Sigma_N^+ = \{1, \ldots, N\}^{\mathbb{N}} = \{ \ \underline{a} = (a_i)_{i=0}^{\infty}, \qquad 1 \leq a_i \leq N\},$$

that is the space of (one-sided) sequences in the digits $1, \ldots, N$.

If f is invertible, for each $i \in \mathbb{Z}$ there exists $1 \leq a_i \leq N$ such that $f^i(x) \in P_{a_i}$ and we can code the full orbit $\mathcal{O}_f^+(x)$ with the full (past and future) itinerary $\underline{a} = (a_i)_{i=-\infty}^{\infty}$, which belongs to the space

$$\Sigma_N = \{1, \ldots, N\}^{\mathbb{Z}} = \{ \ \underline{a} = (a_i)_{i=-\infty}^{\infty}, \qquad 1 \leq a_i \leq N\},$$

that is the space of bi-sided sequences in the digits $1, \ldots, N$.

In both cases, $f(x)$ is coded by the *shifted sequence*: since $f^i(f(x)) = f^{i+1}(x) \in P_{a_{i+1}}$ by definition of itinerary of x, the itinerary of $f(x)$, and hence the coding of $\mathscr{O}_f^+(f(x))$ is given by

$$\sigma^+((a_i)_{i=0}^{+\infty}) = (a_{i+1})_{i=0}^{+\infty}, \qquad \text{or, when } f \text{ is invertible, by} \qquad \sigma((a_i)_{i=-\infty}^{+\infty}) = (a_{i+1})_{i=-\infty}^{+\infty}.$$

The maps

$$\sigma^+ : \Sigma_N^+ \to \Sigma_N^+, \qquad \sigma : \Sigma_N \to \Sigma_N,$$

are known as *full* (one-sided) *shift* on N symbols and *full bi-sided shift* on N

If $\psi : X \to \Sigma_N^+$ (or $\psi : X \to \Sigma_N$ in the invertible case) is the *coding map* which assign to each point its itinerary, the previous relation shows that for all $x \in X$

$$\psi(f(x)) = \sigma^+(\psi(x)) \qquad (\text{or} \quad \psi(f(x)) = \sigma^+(\psi(x)) \quad \text{if } f \text{ is invertible}).$$

In order to give a conjugacy, though, the coding map ψ should be both injective and surjective. Thus, it is natural to ask:

(Q1) Is the coding unique?

(Q2) Do all sequences in Σ_N^+ (or in Σ_N) occur as possible itineraries?

The answer to both these questions is generally NO. In all the cases that we saw (doubling map, baker map, Gauss map) so far, all possible *finite*[14] sequences (in Σ_2^+ for the doubling map, in Σ_2 for the baker map and in countably many digits $\{1, \ldots, n, \ldots, \}^{\mathbb{N}}$ for the Gauss map) do occur, but as the Example 2.6.1 below shows, this is often not the case.

Example 2.6.1. Let $f : [0, 1] \to [0, 1]$ be the map give by

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x < \frac{1}{2}, \\ x - \frac{1}{2} & \text{if } \frac{1}{2} \leq x < 1. \end{cases}$$

whose graph is shown in Figure 2.5. Let $I_1 = [0, 1/2)$ and $I_2 = [1/2, 1]$. It is clear that if $x \in I_2$, then $f(x) \in I_1$. On the other hand, if $x \in I_1$, one could have either $f(x) \in I_1$ (if $x < 1/4$) or $f(x) \in I_2$ (if $1 < 4 \leq x < 1/4$). Thus, one will never see two consecutive digits $2, 2$ in the itinerary, while all combinations $1, 1$, $1, 2$ and $2, 1$ can occur.

[14] Also in these examples there are countably many infinite sequences that do not occurr as itinerararies: for example, for the doubling map, if the coding partition is $[0, 1/2)$ and $[1/2, 1]$, all sequences which end with a tail of 1s do not appear as itinerary of any point.

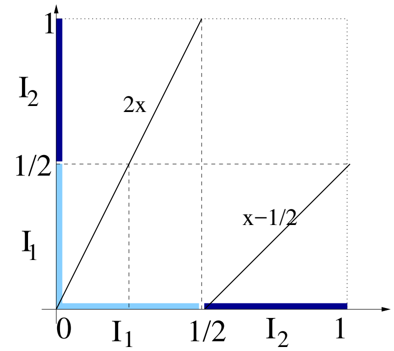


Figure 2.5: The map f in Example 2.6.1.

To be able to describe the subset of the shift that describes itineraries of this form is one of the reasons to study *subshifts of finite type* of the following form.

Definition 2.6.1. An $N \times N$ matrix is called an *transition matrix* (also called *incidence matrix*) if all entries A_{ij}, $1 \leq i, j \leq N$, are either 0 or 1.

One can use a matrix A to encore the information of which pairs of consecutive digits can appear in an itinerary: the digit i can be followed by the digit j if and only if the entry A_{ij} is equal to 1. More formally, we can consider the following subspaces $\Sigma_A^+ \subset \Sigma_N^+$ and $\Sigma_A \subset \Sigma_N$ of sequences

Definition 2.6.2. The shift spaces associated to a transition matrix A are:

$$\Sigma_A^+ = \{(a_i)_{i=0}^{+\infty} \in \Sigma_N^+, \quad A_{a_i a_{i+1}} = 1 \quad \text{for} \quad \text{all } i \in \mathbb{N}\},$$
$$\Sigma_A = \{(a_i)_{i=-\infty}^{+\infty} \in \Sigma_N, \quad A_{a_i a_{i+1}} = 1 \quad \text{for} \quad \text{all } i \in \mathbb{Z}\}.$$

Example 2.6.2. For example if A is the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},$$

since the only zero entry is $A_{22} = 0$, the digit 2 cannot be followed by another digit 2, while all the other pairs of successive digits 12, 11 and 21 are allowed. Thus the sequences in Σ_A^+ (respectively Σ_A) are all sequences in the digits $1, 2$ (respectively all the bisided sequences) without any pairs of consecutive digits 2.

If $(a_i)_{i=0}^{+\infty} \in \Sigma_A^+$, also the shifted sequence $\sigma^+((a_i)_{i=0}^{+\infty})$ belongs to Σ_A^+, since if $A_{a_i a_{i+1}} = 1$ for all $i \in \mathbb{N}$, clearly also $A_{a_{i+1} a_{i+2}} = 1$ for all $i \in \mathbb{N}$ (in other words, if a pair of consecutive digits did not occurr in $\underline{a}$, it clearly does not occurr either in the shifted sequence). The same is true for bisided sequences: if $(a_i)_{i=-\infty}^{+\infty} \in \Sigma_A$, also the shifted sequence $\sigma((a_i)_{i=-\infty}^{+\infty}) \in \Sigma_A$. Thus, the spaces Σ_A and Σ_A^+ are *invariant* under the shift and we can consider the restriction of σ^+ and σ to this subspaces.

Definition 2.6.3. The restriction of the shift maps to

$$\sigma^+ : \Sigma_A^+ \to \Sigma_A^+, \qquad \sigma : \Sigma_A \to \Sigma_A,$$

are called a *topological Markov chain*[15] (or also a *subshift of finite type*) associated to the matrix A.

These are special examples of *subshifts*, that is restrictions of the shift to closed invariant spaces of Σ_N^+ (or Σ_N). In a topological Markov chain, the only type of restrictions on the sequences is of the form *i cannot be followed by j*, thus depend only on the previous digit[16].

[15] In probability, one studies Markov chains, which consists of a topological Markov chain with in addition a measure. We will define a measure which is invariant under the shift in one of the next lectures.

[16] More in general, one can define for example invariant spaces where certain combination of digits, also called *words* in the digits, are not allowed (for example forbidden words could be 2212 and 111, so there cannot be occurrences of 11, but not three consecutive digits 1). A subshift can be equivalently defined in terms of countably many forbidden words: no sequence in the subshift contains a forbidden word and any sequence in the complement does contain a forbidden word. If only a finite number of words are forbidden, we have a *subshift of finite type*. If the maximal length of forbidden words is $k+1$, the subshift is called a *k-step* subshift of finite type. Thus, topological Markov chains are $1-$step subshifts of finite type.

It is very convenient to visualize sequences sequences in Σ_A as paths on a graph.

Definition 2.6.4. The graph $\mathscr{G}_A$ associated to the $N \times N$ transition matrix A is a graph with vertices are $v_1, \ldots, v_N$, where v_i and v_j are connected by an arrow from v_i to v_j if and only if $A_{ij} = 1$.

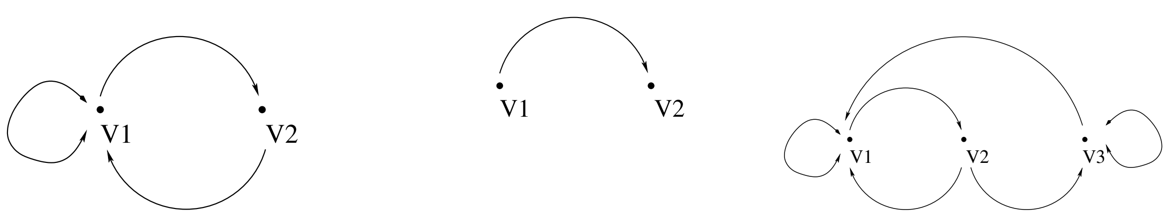
Then the following fact is immediate:

Lemma 2.6.1. *A sequence $(a_i)_{i=0}^{+\infty} \in \Sigma_N^+$ belongs to Σ_A^+ if and only if it describes a path on $\mathscr{G}_A$. Similarly a sequence $(a_i)_{i=-\infty}^{+\infty} \in \Sigma_N$ belongs to Σ_A if and only if it describes a bi-infinite path on $\mathscr{G}_A$.*

Example 2.6.3. For example for the following matrices:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \tag{2.10}$$

one obtains the graphs $\mathscr{G}_A$, $\mathscr{G}_B$ and $\mathscr{G}_C$ in Figure 2.6. A paths on $\mathscr{G}_A$ can never go through v_2 and immediately after v_2 again. Since there are no infinite paths on $\mathscr{G}_B$, we see that $\Sigma_B = \emptyset$.



(a) $\mathscr{G}_A$ (b) $\mathscr{G}_B$ (c) $\mathscr{G}_C$

Figure 2.6: The graphs associated to the transition matrices A, B, C in (2.10).

Number of paths and periodic points

Recall that in the previous section we interpreted the subshift $\Sigma_A \subset \{1, \ldots, N\}^{\mathbb{Z}}$ as the space of bi-infinite paths on the graph $\mathscr{G}_A$ associated to the transition matrix A. The following Lemma turns out to be very helpful to study dynamical properties.

Lemma 2.6.2 (Number of paths). *For any $1 \leq i, j \leq N$, the number of paths of length n on $\mathscr{G}_A$ (that is, paths obtained composing n arrows) starting from the vertex v_i and ending in the vertex v_j is given by $(A^n)_{ij}$ (recall that A^n_{ij} is the (i,j) entry of the matrix A^n obtained producing A by itself n times).*

Proof. Let us prove it by induction on n. For $n = 1$, the paths of length 1 connecting v_i to v_j are simply arrows from v_i to v_j. By definition of $\mathscr{G}_A$ there is an arrow from v_i to v_j if and only if $A_{ij} = 1$, thus the statement for $n = 1$ holds.

Assume we proved it for n. Then, the number of paths of length $n+1$ starting from the vertex v_i and ending in the vertex v_j can be obtained by considering all paths of length $n+1$ starting from the vertex v_i and ending in any of the other vertices v_k, where $1 \leq k \leq N$ and extending each to a path of length $n+1$ ending in v_j if there is an arrow from v_k to v_j. Using that by inductive assumption the number of paths of length n starting from v_i and ending in v_k is given by A^n_{ij}, this gives

$$\begin{aligned} &Card\{\text{ paths of length } n+1 \text{ from } v_i \text{ to } v_j\ \} \\ &\quad = \sum_{k=1}^{N} Card\{\text{ paths of length } n \text{ from } v_i \text{ to } v_k\ \}\ Card\{\text{ arrows from } v_k \text{ to } v_j\ \} \\ &\quad = \sum_{k=1}^{N} A^n_{ik} A_{kj} = A^{n+1}_{ij}, \end{aligned}$$

where in the latter equation we simply used the definition of (i,j) entry of the product matrix $A^n A = A^{n+1}$. $\square$

The Lemma has the following immediate Corollary on the number of periodic points of a topological Markov chain.

Corollary 2.6.1. *The cardinality of periodic points of period n for $\sigma : \Sigma_A \to \Sigma_A$ is exactly the trace $Tr(A^n)$.* (Recall that the trace of a matrix $Tr(A) = \sum_i A_{ii}$ is the sum of the diagonal entries of A.)

Proof. If $\underline{x}$ is a periodic points of period n for $\sigma : \Sigma_A \to \Sigma_A$, then $\sigma^n(\underline{x}) = \underline{x}$, which implies that the digits of the sequence $\underline{x} = (x_i)_{i=-\infty}^{+\infty}$ have period n, that is $x_{n+i} = x_i$ for all $i \in \mathbb{Z}$. Thus, the path described by $\underline{x}$ on $\mathscr{G}_A$ is a periodic path, that repeats periodically the path starting from some v_i and coming back to the same v_i. Since the paths of length n connecting v_i to v_i are A^n_{ii} by Lemma 2.6.2, we have

$$Card\{\underline{x} = (x_i)_{i=-\infty}^{+\infty} \quad \text{such that } \sigma^n(\underline{x}) = \underline{x}\} =$$
$$= \sum_{i=1}^{N} Card\{\underline{x} \quad \text{such that } \sigma^n(\underline{x}) = \underline{x} \quad \text{and} \quad x_0 = i\} = \sum_{i=1}^{N} A^n_{ii} = Tr(A^n),$$

which is the desired expression. □

2.7 Dynamical properties of topological Markov chains

Let A be an $N \times N$ transition matrix and let $\sigma : \Sigma_A \to \Sigma_A$ be the corresponding topological Markov chain. In this section we investigate the topological dynamical properties of $\sigma : \Sigma_A \to \Sigma_A$. We first need to know what are the open sets in a shift space.

Metric, cylinders and balls for a shift space.

The shift spaces Σ_N^+ and Σ_N and their subshifts spaces are metric spaces with one of the following distances.

Let $\Sigma_N = \{1, \dots, N\}^{\mathbb{Z}}$ be the *full bi-sided shift space* on N symbols. For $\rho > 1$ consider the distance

$$d_\rho(\underline{x}, \underline{y}) = \sum_{k=-\infty}^{\infty} \frac{|x_k - y_k|}{\rho^{|k|}}, \qquad \text{where} \quad \underline{x} = (x_k)_{k=-\infty}^{+\infty}, \quad \underline{y} = (y_k)_{k=-\infty}^{\infty}. \tag{2.11}$$

The distance is well defined since $|x_k - y_k| \le N - 1$ for any $k \in \mathbb{N}$ (since both $x_k, y_k \in \{1, \dots, N\}$, their difference is at most $N - 1$) and the series $\sum_k 1/\rho^k$ is convergent for any $\rho > 1$, so that

$$d_\rho(\underline{x}, \underline{y}) \le \sum_{k=-\infty}^{\infty} \frac{N-1}{\rho^{|k|}} \le 2(N-1) \sum_{k=0}^{\infty} \frac{1}{\rho^k} < \infty.$$

Similarly, for the full *one-sided* shift space $\Sigma_N^+ = \{1, \dots, N\}^{\mathbb{N}}$ we can consider for any $\rho > 1$ the distance

$$d_\rho^+(\underline{x}, \underline{y}) = \sum_{k=0}^{\infty} \frac{|x_k - y_k|}{\rho^k}, \qquad \text{where} \quad \underline{x} = (x_k)_{k=0}^{\infty}, \quad \underline{y} = (y_k)_{k=0}^{\infty}. \tag{2.12}$$

Note that if $\Sigma_A \subset \Sigma_N$ (respectively $\Sigma_A^+ \subset \Sigma_N^+$) is a subshift, the distance d_ρ in (2.11) (respectively the distance d_ρ^+ in (2.12), gives a distance on Σ_A (respectively Σ_A^+). More generally, all subshift spaces (subsets of Σ_N or Σ_N^+ invariant under the shift map σ) are metric spaces.

The following sets, called cylinders, play an essential role in the study of shift spaces:

Definition 2.7.1. A *cylinder* is a subset of Σ_N of the form

$$C_{-m,n}(a_{-m}, \dots, a_n) = \{ \underline{x} = (x_i)_{i=-\infty}^{+\infty} \in \Sigma_N, \ \text{ such that } \ x_i = a_i \quad \text{for all } -m \le i \le n \},$$

where $m, n \in \mathbb{N}$ and $a_i \in \{1, \dots, N\}$ for $-m \le i \le n$.

A cylinder is called *symmetric* if $n = m$.

Similarly a cylinder in Σ_N^+ is a subset of the form

$$C_n(a_0, \dots, a_n) = \{ \underline{x} = (x_i)_{i=0}^{+\infty} \in \Sigma_N^+, \ \text{ such that } \ x_i = a_i \quad \text{for all } 0 \le i \le n \},$$

where $n \in \mathbb{N}$ and $a_i \in \{1, \dots, N\}$ for $0 \le i \le n$.

Example 2.7.1. The following sequence

$$\underline{x} = \dots, x_{-m-2}, x_{-m-1}, \underbrace{a_{-m}, \dots, \overbrace{a_0}^{i=0}, \dots, a_n}_{fixed\ block}, x_{n+1}, x_{n+2} \cdots$$

belongs to the cylinder $C_{-m,n}(a_{-m}, \dots, a_n)$. All points $(y_i)_{i=-\infty}^{+\infty}$ in $C_{-m,n}(a_{-m}, \dots, a_n)$ contain the fixed block of digits $a_{-m}, \dots, a_n$ centered at a_0 (for $i = 0$), while the tails can be any sequence in the digits $\{1, \dots, N\}$.

It two points belong to the same cylinder, they share a common central block of digits. Thus, it is clear that the distances in (2.11) and (2.12) are small. More is true. If ρ is chosen sufficiently large, than cylinders in one sides shifts spaces and symmetric cylinders in two sided shifts spaces are *exactly* balls with respect to the distances d_ρ^+ and d_ρ, as shown by the following two Lemmas.

Lemma 2.7.1. *If $\rho > N$, for any $\epsilon = 1/\rho^n$ we have:*

$$C_n(x_0, \dots, x_n) = B_{d_\rho^+}\left(\underline{x}, \frac{1}{\rho^n}\right).$$

Proof. Let $C_n(x_0, \dots, x_n)$ be a cylinder in Σ_N^+. Since $\underline{x} = (x_i)_{i=0}^{\infty} \in \Sigma_N^+$ begins with $x_0, \dots, x_n$, the point $\underline{x} \in C_n(x_0, \dots, x_n)$. If also $\underline{y} \in C_n(x_0, \dots, x_n)$, then, since $|x_k - y_k| = 0$ for all $k \in \mathbb{N}$ with $0 \le k \le n$, we have

$$d_\rho(\underline{x}, \underline{y}) = \sum_{k=0}^{\infty} \frac{|x_k - y_k|}{\rho^k} = \sum_{k=n+1}^{\infty} \frac{|x_k - y_k|}{\rho^k}.$$

Note that, since both $x_i, y_i \in \{1, \dots, N\}$, for any $i \in \mathbb{N}$ we have $|x_i - y_i| \le N - 1$. Thus, using also the formula for the sum of the geometric progression, that gives us

$$\sum_{j=0}^{\infty} \frac{1}{\rho^j} = \frac{1}{1 - \frac{1}{\rho}} = \frac{\rho}{\rho - 1},$$

we get

$$d_\rho(\underline{x}, \underline{y}) \le \sum_{k=n+1}^{\infty} \frac{N-1}{\rho^k} \le \frac{N-1}{\rho^{n+1}} \sum_{j=0}^{\infty} \frac{1}{\rho^j} = \frac{N-1}{\rho^{n+1}} \frac{\rho}{\rho - 1} = \frac{N-1}{\rho^n} \frac{1}{\rho - 1}.$$

Thus, since $\rho > N$ and hence $N - 1/(\rho - 1) < 1$,

$$d_\rho(\underline{x}, \underline{y}) < \frac{1}{\rho^n} \quad \Leftrightarrow \quad \underline{y} \in B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right).$$

This proves that if $\rho > N$ we have the inclusion

$$C_n(x_0, \dots, x_n) \subset B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right).$$

Let us check the reverse inclusion. Assume that $\underline{y} \in B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right)$. Then, if by contradiction $\underline{y} \notin C_n(x_0, \dots, x_n)$, there exists $j \in \mathbb{Z}$ with $0 \le j \le n$ such that $x_j \ne y_j$, so that $|x_j - y_j| \ge 1$. But then

$$d_\rho(\underline{x}, \underline{y}) = \sum_{k=0}^{\infty} \frac{|x_k - y_k|}{\rho^k} \ge \frac{|x_j - y_j|}{\rho^j} \ge \frac{1}{\rho^j} \ge \frac{1}{\rho^n}, \qquad \text{since } 0 \le j \le n.$$

Thus, $\underline{y} \notin B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right)$, which is a contradiction. Thus, we also have (without any additional assumption on ρ) the opposite inclusion

$$B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right) \subset C_n(x_0, \dots, x_n).$$

$\square$

In a similar way, one can prove the following lemma.

Lemma 2.7.2. *If $\rho > 2N - 1$, than for any $\epsilon = 1/\rho^n$ we have:*

$$C_{-n,n}(x_{-n}, \dots, x_n) = B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right),$$

where $\underline{x} = (x_i)_{i=-\infty}^{\infty} \in \Sigma_N$ be a sequence which contains the central block $x_{-n}, \dots, x_n$.

Exercise 2.7.1. Prove Lemma 2.7.2.

Consider now a subshift $\Sigma_A \subset \Sigma_N$ determined by the transition matrix A (or the one-sided subshift $\Sigma_A^+ \subset \Sigma_N^+$).

Definition 2.7.2. A cylinder $C_{-m,n}(a_{-m}, \dots, a_n)$ (where $n, m \in \mathbb{N}$ and $a_i \in \{1, \dots, N\}$ for $-m \leq i \leq n$) is called *admissible* if $A_{a_i, a_{i+1}} = 1$ for all $-m \leq i < n$.

Similarly, a cylinder $C_n(a_0, \dots, a_n)$ (where $n \in \mathbb{N}$ and $a_i \in \{1, \dots, N\}$ for $0 \leq i \leq n$) is called *admissible* if $A_{a_i, a_{i+1}} = 1$ for all $0 \leq i < n$.

Cylinder sets are the basic sets which can be used to study topological properties, in virtue of the following remark.

Lemma 2.7.3. *Let A be an $N \times N$ transition matrix. Assume that $\rho > N$. Then*

i. Any non-empty open set U in Σ_N^+ contains a cylinder $C_n(a_0, \dots, a_n)$.

ii. Any non-empty open set U in Σ_A^+ contains an admissible cylinder $C_n(a_0, \dots, a_n)$.

Similarly, if $\rho > 2N - 1$:

iii. Any non-empty open set U in Σ_N contains a symmetric cylinder $C_{-n,n}(a_0, \dots, a_n)$.

iv. Any non-empty open set U in Σ_A contains an admissible symmetric cylinder $C_{-n,n}(a_{-n}, \dots, a_n)$.

Proof. Let U be a non empty open set in Σ_N^+ (respectively in Σ_A^+). Since U is non empty, there exists $\underline{x} \in \Sigma_N^+$ (respectively in $\underline{x} \in \Sigma_A^+$) such that $\underline{x} \in U$. Since U is open, there exists $\epsilon > 0$ such that $B_{d_\rho^+}(\underline{x}, \epsilon) \subset U$. If we choose n such that $1/\rho^n < \epsilon$, since $\rho > N$, by Lemma 2.7.1 we have that

$$C_n(x_0, \dots, x_n) = B_{d_\rho^+}(\underline{x}, \frac{1}{\rho^n}) \subset B_{d_\rho^+}(\underline{x}, \epsilon) \subset U.$$

Furthermore, if $\underline{x} \in \Sigma_A^+$, $A_{x_i, x_{i+1}} = 1$ for all $0 \leq i < n$, so $C_n(x_0, \dots, x_n)$ is admissible. This concludes the proof of i and ii. The other two points are proved similarly. □

Topological transitivity and topological mixing

Let us give two preliminary definitions.

Definition 2.7.3. A transition matrix A is called *irredudible* if for any $1 \leq i, j \leq N$ there exists an $n \in \mathbb{N}$ (possibly dependent on i, j) such that the entry $A^n_{ij} := (A^n)_{ij}$ of the matrix A^n obtained multiplying A by itself n times is *positive* ($(A^n)_{ij} > 0$).

A transition matrix A is called *aperiodic* (or also, in some books, *transitive*) if there exists an $n \in \mathbb{N}$ such that if for any $1 \leq i, j \leq N$ we have $(A^n)_{ij} > 0$.

A matrix A such that all entries $A_{ij} > 0$ is called *positive* (and we write $A > 0$). Thus, A is aperiodic if there exists a power $n \in \mathbb{N}$ such that A^n is positive.

Example 2.7.2. For example

$$A^2 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}; \qquad \text{if } D = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad D^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}, \tag{2.13}$$

so A is irreducible and aperiodic (with $n = 2$) since all entries of A^2 are positive, while D is not irreducible, since for any n the entry $(D^n)_{21} = 0$.

Remark 2.7.1. Irreducible and aperiodic matrices can be easily recognized using the associated graph:

(1) A is *irreducible* if and only if for any two vertices v_i and v_j on $\mathscr{G}_A$ there exists a path connecting v_i to v_j;

(2) A is *aperiodic* if and only if there exists an n such that any two vertices v_i and v_j on $\mathscr{G}_A$ can be connected by a path of the same length n.

Exercise 2.7.2. Prove the above Remark using Lemma 2.6.2.

Example 2.7.3. The graph $\mathscr{G}_A$ in Figure 2.6 shows that A is irreducible, since all vertices can be connected: for example, v_2 can be connected to itself going through v_1. Moreover, it is aperiodic with $n = 2$, since the paths from v_2 to v_2 has length two and one can get paths of length 2 from v_1 to itself, from v_2 to v_1 and from v_1 to v_2 by repeating the loop around v_1. The graph $\mathscr{G}_B$ in Figure 2.6 shows that B is *not* irreducible, since there are no paths connecting for example v_1 to itself (or neither paths connecting v_2 to v_1 and v_2 to itself). The graph $\mathscr{G}_C$ in Figure 2.6 shows that C is irreducible, since one sees immediately that all vertices can be connected to each other.

Exercise 2.7.3. Consider the following transition matrices

$$A_1 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}. \tag{2.14}$$

Draw the corresponding graphs $\mathscr{G}_{A_i}$, $i = 1, 2$, associated to them. For each $i = 1, 2$ is A_i irreducible? is A_i aperiodic?

The dynamical significance of these definitions lie in the following Theorem.

Theorem 2.7.1. *Let $\sigma : \Sigma_A \to \Sigma_A$ be a topological Markov chain.*

(1) The matrix A is irreducible if and only if $\sigma : \Sigma_A \to \Sigma_A$ is topologically transitive;

(2) If A is aperiodic, then $\sigma : \Sigma_A \to \Sigma_A$ is topologically mixing.

Let us first prove a Lemma which will be used in the proof of the Theorem.

Lemma 2.7.4. *If $A^n > 0$ for some $n > 0$, then for any $m \geq n$ we also have $A^m > 0$.*

Proof. Note first that if $A^n > 0$ for some $n > 0$, this means that for each j there exists a k_j such that $A_{k_j j} = 1$. Otherwise, if $A_{kj} = 0$ for all $1 \leq k \leq N$, then the vertex v_j cannot be reached from any other vertex v_k, so there cannot exist any path of length n reaching v_j, in contradiction with the fact that $A^n_{ij} > 0$.

Let us now prove by induction on m that $A^m > 0$ for $m \geq n$. For $m = n$ it is true by assumption. If we verified it for m, take any $1 \leq i, j \leq N$. Then, by the previous remark there exists k_j such that $A_{k_j j} = 1$. Moreover, for all the other k we have $A_{kj} \geq 0$. Hence, we get

$$A^{m+1}_{ij} = \sum_{k=1}^{N} A^m_{ik} A_{kj} \geq A^m_{ik_j} A_{k_j j} = A^m_{ik_j}$$

and $A^m_{ik_j} > 0$ since $A^m > 0$ the inductive assumption. This shows that $A^{m+1} > 0$ and concludes the proof. □

We will also need the following remark.

Remark 2.7.2. If A is *irreducible* (so in particular also if it is *aperiodic*), a cylinder is *admissible* if and only if it is not empty. Indeed, the condition $A_{a_i, a_{i+1}} = 1$ for all $-m \leq i < n$ guarantees that there is a path on $\mathscr{G}_A$ described by $a_{-m}, \dots, a_n$ (that is passing in order through the vertices $v_{a_{-m}}, \dots, v_{a_n}$) and since A is irreducible, one can continue this path to a biinfinite path in $\mathscr{G}_A$ (adding any admissible forward tail starting from a_n and any admissible backward tail before a_{-m}). This path belong to the cylinder and shows that it is not empy.

Proof of Theorem 2.7.1. Let us prove (1). Assume that A is irreducible. We want to show that for each U, V non-empty open sets there exists $M > 0$ such that $\sigma^M(U) \cap V \neq \emptyset$. Fix $\rho > 2N - 1$. Since each open sets contains an open ball of the form $B_{d_\rho}(\underline{x}, \rho^{-k})$ for some large k and by Lemma in the previous section each ball of this form is a symmetric cylinder, there exists two admissible cylinders

$$C_{(-k,k)}(a_{-k}, \dots, a_k) \subset U, \qquad C_{(-l,l)}(b_{-l}, \dots, b_l) \subset V. \tag{2.15}$$

Let us now construct a point $\underline{x}$ which contains both blocks of digits $a_{-k}, \dots, a_k$ and $b_{-l}, \dots, b_l$. By definition of irreducibility, taking $i = a_k$ and $j = b_{-l}$, there exists $n > 0$ such that $A^n_{a_k, b_{-l}} > 0$. This means that there exists a path of length n which connects v_{a_k} to $v_{b_{-1}}$. Let us denote by

$$y_0 = a_k, y_1, y_2, \dots, y_{n-1}, y_n = b_{-l},$$

the digits which describe this path. Clearly $A_{y_i y_{i+1}} = 1$ for all $0 \leq i \leq n-1$. Consider a point $\underline{x} \in \Sigma_A$ such that

$$\underline{x} = \dots a_{-k}, \dots, \underbrace{a_0}_{i=0}, \dots, a_k, y_1, \dots, y_{n-1}, b_{-l}, \dots, b_l, \dots$$

(such point exist since by irreducibility we can choose a backward and a forward tail by choosing any path on $\mathscr{G}_A$ which starts from b_l (for the forward tail) or ends in a_{-k} (for the backward tail). Clearly, since $\underline{x}$ contains as central block of digits $a_{-k}, \dots, a_k$, we have $\underline{x} \in C_{(-k,k)}(a_{-k}, \dots, a_k) \subset U$. Moreover, if we set $M = n+k+l$, shifting the sequence $k + n + l$ times to the left, since $x_{k+n+l} = b_0$, we get

$$\sigma^M(\underline{x}) = \dots b_{-l}, \dots, \underbrace{b_0}_{i=0}, \dots, b_l, \dots,$$

so that $\sigma^M(\underline{x}) \in C_{(-l,l)}(b_{-l}, \dots, b_l) \subset V$. This shows that

$$\underline{x} \in U \cap \sigma^{-M}(V) \neq \emptyset \qquad \Leftrightarrow \qquad \sigma^M(U) \cap V \neq \emptyset.$$

This shows that $\sigma : \Sigma_N \to \Sigma_N$ is topologically transitive.

Let us prove the converse implication in (1). Assume that $\sigma : \Sigma_A \to \Sigma_A$ is topologically transitive. Let us show that A is irreducible. Let $1 \leq i, j \leq N$. Consider as open sets U, V the cylinders giving by fixing only the first digit:

$$U = C_{(0,0)}(i) = \{\underline{x} \in \Sigma_A, \quad x_0 = i\}, \qquad V = C_{(0,0)}(j) = \{\underline{x} \in \Sigma_A, \quad x_0 = j\}.$$

Since $\sigma : \Sigma_A \to \Sigma_A$ is topologically transitive, there exists $n > 0$ such that $\sigma^n(U) \cap V \neq \emptyset$. Equivalently, $U \cap \sigma^{-n}(V) \neq \emptyset$. This means that there exists $\underline{x} \in U \cap \sigma^{-n}(V)$. Since $\underline{x} \in U$, this means that $x_0 = i$. Since $\underline{x} \in \sigma^{-n}(V)$, we have $\sigma^n(\underline{x}) \in V$. But

$$(\sigma^n(\underline{x}))_0 = x_n, \quad \text{so} \quad \sigma^n(\underline{x}) \in V \quad \Leftrightarrow \quad x_n = j.$$

Thus, we found an element $\underline{x} \in \Sigma_A$, that by definition describes a biinfinite path on $\mathscr{G}_A$, such that $x_0 = i$ and $x_n = j$. This gives a path of length n connecting v_i to v_j, showing that $A^n_{ij} > 0$.

Let us now prove (2). Assume that $A^n > 0$. We want to show that $\sigma : \Sigma_A \to \Sigma_A$ is topologically mixing. Let U, V be non empty open sets. We seek M_0 such that for any $M \geq M_0$ we have $\sigma^M(U) \cap V \neq \emptyset$. We can reason very similarly to part (1). Both U, V contain admissible symmetric cylinders of the form (2.15). Let $M_0 = n + k + l$. If $M \geq M_0$, then $M = m + k + l$ with $m \geq n$. Then also $A^m > 0$ by Lemma 2.7.4, so $A^m_{a_k, b_{-l}} > 0$. Thus, there exists a path of length m from v_{a_k} and $v_{b_{-l}}$, so we can construct a point in Σ_A of the form

$$\underline{x} = \ldots a_{-k}, \ldots, \underbrace{a_0}_{i=0}, \ldots, a_k, y_1, \ldots, y_{m-1}, b_{-l}, \ldots, b_l, \ldots$$

Reasoning as in Part (1), $\underline{x} \in U \cap \sigma^{-M}(V)$, so that $\sigma^M(U) \cap V \neq \emptyset$. This can be repeated for any $M \geq M_0$, showing that $\sigma : \Sigma_A \to \Sigma_A$ is topologically mixing. $\square$

Topological entropy of Markov chains

Let us sketch how to compute the topological entropy of a topological Markov chain. We will use here the definition of topological entropy via covers.

Given an $N \times N$ matrix, let $\|A\|$ be the norm given by

$$\|A\| = \sum_{1 \leq i,j \leq N} |A_{ij}|.$$

Theorem 2.7.2 (Entropy of topological Markov chains)**.** *Let A be a $N \times N$ be an* aperiodic *transition matrix and $\sigma : \Sigma_A \to \Sigma_A$ be the associated topological Markov chain. The following limit exists and the topological entropy is given by*

$$h_{top}(\sigma) = \lim_{n \to \infty} \log \frac{\|A^n\|}{n}. \tag{2.16}$$

Remark 2.7.3. The limit

$$\rho(A) = \lim_{n \to \infty} \log \frac{\|A^n\|}{n}.$$

which appears in in (2.16) is called *spectral radius* of the matrix A. One can show that the spectral radius is also given by

$$\rho(A) = \log |\lambda_A^{max}|,$$

where $|\lambda_A^{max}|$ is the maximum modulus of the eigenvalues of A.

Combining the Remark and the Theorem we have the following Corollary, which is more useful for the computations of entropy:

Corollary 2.7.1. *The topological entropy of $\sigma : \Sigma_A \to \Sigma_A$ where A is irreducible is given by*

$$h_{top}(\sigma) = \log |\lambda_A^{max}|.$$

Proof of Theorem 2.4.1. Let $\rho < 2N - 1$ and consider the distance $d = d_\rho$, so that by Lemma 2.7.1 balls of radius $1/\rho^n$ in Σ_N are exacly symmetric cylinders of the form $C_{-n,n}(a_{-n}, \ldots, a_n)$. Let $\epsilon = 1/\rho^k$. Consider the collection $\mathscr{U}_n$ of cylinders which are admissible and of the following form:

$$\mathscr{U}_n = \{\ C_{-k,n-1+k}(a_{-k}, \ldots, a_{n-1+k}), \qquad A_{a_i a_{i+1}} = 1 \text{ for } -k \leq i < n-1+k\ \}. \tag{2.17}$$

Since $(a_{-k},\dots,a_{n-1+k})$ varies over all possible admissible values, $\mathscr{U}_n$ is a cover: for any $\underline{x}\in\Sigma_A$, $\underline{x}\in C_{-k,n-1+k}(x_{-k},\dots,x_{n-1+k})\in\mathscr{U}_n$. Moreover, since cylinders are open balls, $\mathscr{U}_n$ is an open cover. Let us check that the d_n diameter of each $C\in\mathscr{U}_n$ is less than ϵ.

Let $C=C_{-k,n-1+k}(a_{-k},\dots,a_{n-1+k})$. If $\underline{x},\underline{y}\in C$, by definition

$$x_i=y_i=a_i \ \text{ for } -k\le i\le k+n \quad\Rightarrow\quad \sigma^j(\underline{x})_i=\sigma^j(\underline{y})_i \ \text{ for } -k-j\le i\le n+k-j.$$

Thus, for any $0\le j<n$, since $n-1+k-j\le k$, we have in particular

$$\sigma^j(\underline{x})_i=\sigma^j(\underline{y})_i \ \text{ for } -k\le i\le k,$$

so that both $\sigma^j(\underline{x})$ and $\sigma^j(\underline{y})$ belong to the same symmetric cylinder which is a ball or radius $1/\rho^k$. Hence

$$d_n(\underline{x},\underline{y})=\max_{0\le j\le n-1} d(\sigma^j(\underline{x}),\sigma^j(\underline{y}))\le 1/\rho^k=\epsilon.$$

Since this holds for all $\underline{x},\underline{y}\in C$, $diam_d(C)<\epsilon$.

Since we just showed that $\mathscr{U}_n$ is an open cover with balls of d_n-diameter less than ϵ, we have

$$Cov(n,\epsilon)\le Card(\mathscr{U}_n).$$

The cardinality of $\mathscr{U}_n$ is the number of admissible cylinders of the form (2.17), thus the number of admissible paths of length $n+2k$ from any vertex v_i to any other vertex v_j. Thus, by Lemma 2.6.2, we have

$$Card(\mathscr{U}_n)=\sum_{i,j=1}^{N}\{\text{paths of length } n+2k \text{ from } v_i \text{ to } v_j\ \}=\sum_{i,j=1}^{N}A_{ij}^{n+2k}.$$

Moreover, one can see that $\mathscr{U}_n$ is a minimal cover with sets of d_n-diameter less than ϵ. Note that all cylinders in $\mathscr{U}$ are *disjoint* balls in the d_n-metric. A cover has in particular to cover a point in each cylinder $C_i\in\mathscr{U}$ with some open set U_i, but since the $diam_{d_n}(U_i)\le\epsilon$, it cannot contain any other point outside C_i. Thus, the cardinality of any cover with open sets of d_n-diameter less than ϵ has at least the cardinality of $\mathscr{U}_n$. Thus

$$Cov(n,\epsilon)=Card(\mathscr{U}_n)=\sum_{i,j=1}^{N}A_{ij}^{n+2k}=\|A^{n+2k}\|.$$

Using the definition of entropy via covers and Lemma **??**, we get

$$h_{top}(\sigma,\epsilon)=\lim_{n\to\infty}\frac{\log Cov(n,\epsilon)}{n}=\lim_{n\to\infty}\frac{\log\|A^{n+2k}\|}{n}=\lim_{n\to\infty}\frac{\log\|A^{n+2k}\|}{(n+2k)}\frac{(n+2k)}{n}=\lim_{n\to\infty}\frac{\log\|A^{n}\|}{n}.$$

(Recall that when we send n to infinity, k is fixed). Thus, since the quantity is independent on ϵ, $h_{top}(\sigma)=h_{top}(\sigma,\epsilon)$. $\square$

2.8 Symbolic coding for the CAT map

In all the examples of coding that we have seen so far, the choice of the partition to use to describe itineraries was very natural. We will show now an example were the partition is much less obvious.

Let $f_A:\mathbb{T}^2\to\mathbb{T}^2$ be the cat map, that is the hyperbolic toral automorphism associated to the matrix

$$A=\begin{pmatrix}2&1\\1&1\end{pmatrix}.$$

Let $\lambda_1>1,\lambda_2<1$ be the eigenvalues of A and let

$$\underline{v}_1=\begin{pmatrix}\frac{1+\sqrt{5}}{2}\\1\end{pmatrix},\qquad \underline{v}_2=\begin{pmatrix}\frac{1-\sqrt{5}}{2}\\1\end{pmatrix}.$$

be the corresponding eigenvectors, which are orthogonal since their scalar product is

$$< \underline{v}_1, \underline{v}_2 >= \frac{(1+\sqrt{5})}{2}\frac{(1-\sqrt{5})}{2} + 1 \cdot 1 = \frac{1-5}{2} + 1 = -1 + 1 = 0.$$

In Figure 2.7 you can see a partition of $\mathbb{T}^2$ into two sets, R_1 and R_2 (R_1 is the union of the white pieces and R_2 is the union of the dark pieces).

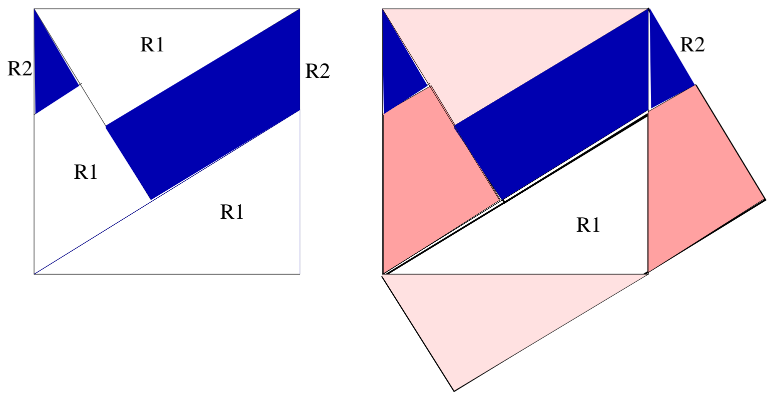


Figure 2.7: A partition of $\mathbb{T}^2$ into rectangles R_1, R_2 .

Each of them is a *rectangle* on the surface of $\mathbb{T}^2$, whose sides are in the orthogonal directions of either v_1 or v_2. In order to see that, it is convenient to cut and paste the triangles as in Figure 2.7 (sets filled with the same shade are translations of each other). When you cut and paste a triangle by moving it by an integer vector $(k, l) \in \mathbb{Z}^2$, you get a different set in $\mathbb{R}^2$, which represent the *same* set on the torus $\mathbb{T}^2$ (recall that $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ is the set of equivalence classes of points in $\mathbb{R}^2$ where $(x, y) \sim (x', y')$ if $(x', y') = (x, y) + (k, l)$). Thus, if instead than the standard copy which lies in the unit square $[0, 1)^2$, we use the cut and pasted copies translated as in Figure 2.7, we can see that both R_1 and R_2 are rectangles.

[Instead than representing $\mathbb{T}^2$ as a unit square with opposite sides identified, we could choose to represent $\mathbb{T}^2$ as union of the two copies of the rectangles R_1 and R_2 where again opposite sides are identified.]

We could try to use the partition $\mathscr{P} = \{R_1, R_2\}$ of $\mathbb{T}^2$ to code f_A. This partition has the nice property that rectangles are mapped to rectangles. Indeed, since the sides of R_1 and R_2 are parallel to eigenvectors, the image of R_1 and R_2 under A have still sides parallel to v_1 and v_2, that is, they are still rectangles.

Let us describe the images $f_A(R_1)$, $f_A(R_2)$ under the cat map, referring to Figure 2.8.

Recall that f_A is obtained by first acting linearly by A and then taking the result modulo one, which correspond to projecting $\mathbb{R}^2$ to $\mathbb{T}^2$ by the projection $\pi : \mathbb{R}^2 \to \mathbb{T}^2$ given by

$$\pi(x, y) = (x \mod 1, y \mod 1). \qquad \text{So} \quad f_A = \pi \circ A.$$

Since the sides of R_1 and R_2 are parallel to eigenvectors, the image of R_1 and R_2 are still rectangles, but since all directions parallel to v_1 are expanded by a factor λ_1 and all directions parallel to v_1 are contracted by a factor λ_2, the rectangles image under A are thinner and longer, as shown in Figure 2.8 (the image of R_1 under A is the lighter shade rectangle, the image of R_2 under A is the darker shade rectangle). The projection π consists of cutting and pasting corresponding pieces of these rectangles back to the unit square, as shown again in Figure 2.8 (sets with the same name are translated copies of each other).

Note that the images of R_1 and R_2 under A crosses different parts of the translated copies of each original rectangle R_1, R_2. In order to describe these intersections precisely, let us write

$$R_1 = P_1 \cup P_2 \cup P_3, \qquad R_2 = P_4 \cup P_5,$$

where $P_1, \ldots, P_5$ are the sets in Figure 2.8 (that we give the same name P_i to all the sets in $\mathbb{R}^2$ that represent a translated copy by an integer vector of $P_i \subset [0, 1)^2$, since they correspond to the same set on $\mathbb{T}^2$).

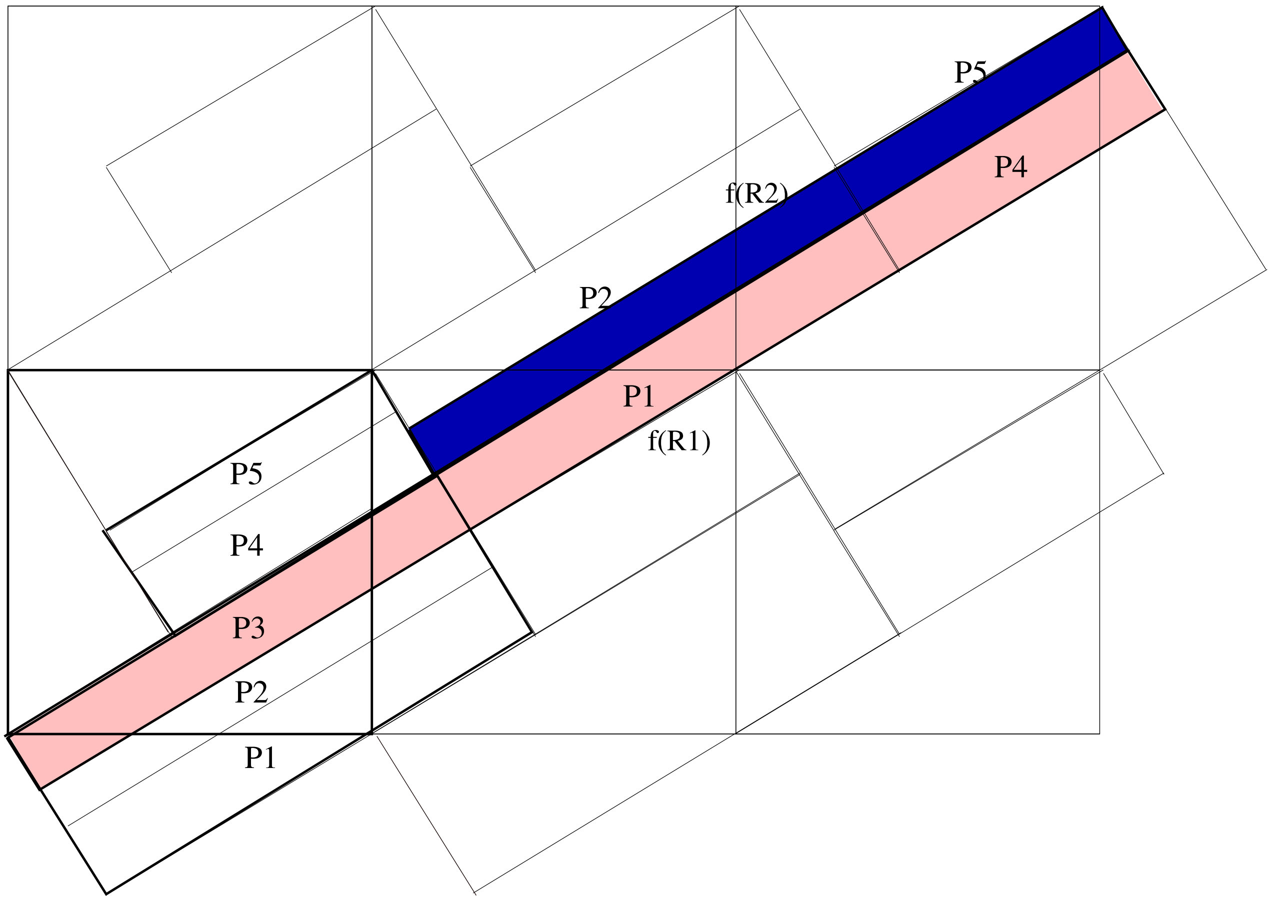


Figure 2.8: The images of the rectangles R_1 and R_2 under $f_A : \mathbb{T}^2 \to \mathbb{T}^2$.

Looking at Figure 2.8, you can see that

$$f(R_1) = P_3 \cup P_1 \cup P_4, \qquad f(R_2) = P_2 \cup P_5. \tag{2.18}$$

Since R_1 crosses itself more than once, it is not a good idea to use the partition $\mathscr{P} = \{R_1, R_2\}$ of $\mathbb{T}^2$ to code f_A. Let us use instead the finer partition

$$\mathscr{P} = \{P_1, P_2, P_3, P_4, P_5\}, \qquad \mathbb{T}^2 = R_1 \cup R_2 = (P_1 \cup P_2 \cup P_3) \cup (P_4 \cup P_5) = \bigcup_{k=1}^{5} P_k$$

It is clear that if we code the orbit of the point (x, y) using a sequence $(a_i)_{i\in\mathbb{Z}} \in \Sigma_5 = \{1, \ldots, 5\}^{\mathbb{Z}}$ in the usual way, so that

$$f_A^k((x, y)) \in P_{a_k}, \qquad \text{for all } k \in \mathbb{Z},$$

not all sequences of Σ_5 will describe itineraries, becouse of the inclusions (2.18). For example, if $a_k = 1$, it means that $f_A^k((x, y)) \in P_1 \subset R_1$. Thus, $f_A^{k+1}((x, y)) \in f(R_1) = P_3 \cup P_1 \cup P_4$, so a_{k+1} can be only $1, 3$ or 4.

Reasoning in a similar way, since if $x \in P_2 \subset R_1$ or $x \in P_3 \subset R_1$, $f(x) \in f(R_1) = P_3 \cup P_1 \cup P_4$, the digits $2, 3$ can be followed only by $1, 3$ or 4. If $x \in P_4 \subset R_2$ or $x \in P_5 \subset R_2$, $f(x) \in f(R_2) = P_2 \cup P_5$, we see that $4, 5$ can be followed only by 2 or 5.

Let encode this information in a 5×5 transition matrix B, by setting $B_{ij} = 1$ if and only if i can be followed by j, 0 otherwise. We get

$$B = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}. \tag{2.19}$$

These considerations lead to the following:

Remark 2.8.1. Full itineraries of orbits $\mathscr{O}^{+}_{f_A}(\underline{x})$ of the cat map $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ belong to the shift space Σ_B.

Conversely, all the transitions that we describe can actually occur. Thus, finite sequences in Σ_B all represent possibile itineararies of orbits of the cat map with respect to the partition $\mathscr{P}$. Moreover, a general property of the coding map, the image of a point under f_A is coded by the shifted sequence in Σ_B. Thus, one can use the shift space Σ_B to construct a (semi-)conjugacy of the cat map with the topological Markov chain $\sigma : \Sigma_B \to \Sigma_B$.

More details can be found in the references quoted in the Extra.

Extra: a conjugacy of the CAT map with a topological Markov chain

More precisely, one can prove the following:

Theorem 2.8.1. *The CAT map $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ and the topological Markov chain $\sigma : \Sigma_B \to \Sigma_B$ are topologically semi-conjugated via a semi-conjugacy $\psi : \Sigma_B \to \mathbb{T}^2$.*

Furthermore, the map ψ is one to one onto the complement of the union of the lines through the origin of $\mathbb{T}^2$ parallel to one of the eingenvalues (which contain the boundaries of the rectangles and their images and premiages via f_A) and one can obtain a topological conjugacy $\psi : \hat{\Sigma_B} \to \mathbb{T}^2$ by inducing the map ψ on the space $\hat{\Sigma_B} = \Sigma_A / \sim$ obtained by identifying the sequences in Σ_B which either

- *share a forward tail of 0s or 1s (i.e. there exists i_0 such that $x_i = 0$ for all $i \geq 0$ or $x_i = 0$ for all $i \geq 0$);*
- *share a backward tail of 4s or 4s (i.e. there exists i_0 such that $x_i = 4$ for all $i \leq 0$ or $x_i = 2$ for all $i \leq 0$);*

Sketch of the Proof. The idea of the proof is to build the map $\underline{x} \to \psi(\underline{x})$ which assign to $\underline{x} \in \Sigma_B$ the point(s) in $(x, y) \in \mathbb{T}^2$ which has $\underline{x}$ as itinerary of the full orbit $\mathscr{O}_{f_A}(x, y)$. To construct this, one can first show that, for any fixed block, there is a rectangle of in $\mathbb{T}^2$ consisting of points which share that block as part of their itinerary (this is because the conditions imposed on sequences Σ_B reflect geometric intersections between rectangles and their images). Then, one can show that the size of these rectangles, respectively the width and the height, shrink exponentially respectively as longer and longer pieces of backward or forward itineraries are prescribed (as in the case of the baker map, where the future of the itinerary was determining the x-coordinate, while the past the y-one). Finally, the identifications of sequences with certain tails are necessary to make the map one to one, since for the boundaries of these rectangles there is an ambiguity of coding.

The details of the proof can be found in the book by Katok and Hasselblatt (see §7.4.5). □

Once we have a topological conjugacy between the CAT map and the shift space, checking topological properties becomes automatic in virtue of the results that we have proved so far. For example, to show that the CAT map is topologically transitive (in particular it has dense orbits) it is sufficient to verify that the matrix B is irreducible.

Finding a good partition for coding is not always easy (as in this case), but once one has constructed a semi-conjugacy with a shift space, it is much easier to prove some dynamical properties, as for example topological transitivity (which, for a subshift given by a transition matrix, follows simply by verifying that the matrix is irreducible). See for example the exercise Ex. 2.8.1 below.

Remark 2.8.2. The partition used to code the cat map is an example of *Markov partition.* Markov partitions can be constructed more in general for hyperbolic dynamical systems, that is systems that have contracting and expanding directions. Coding via a Markov partition, allows to reduce the study of the dynamical system to a symbolic space. This is often a powerful technique to prove dynamical properties of the original system.

In particular, Markov partitions can be constructed for all hyperbolic toral automorphisms similarly to the example we saw for the cat map. We give another example as an Exercise.

* **Exercise 2.8.1.** Check that the toral automorphism $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ given by the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}.$$

is hyperbolic. Figure 2.9 shows a partition of $\mathbb{T}^2$ into rectangles and their images under f_A.

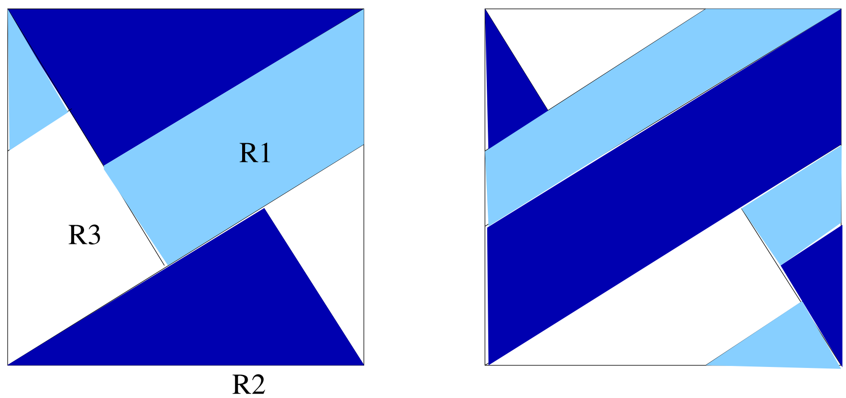


Figure 2.9: A partition of $\mathbb{T}^2$ into rectangles R_0, R_1, R_2 and their images under $f_A : \mathbb{T}^2 \to \mathbb{T}^2$.

(a) Using itineraries with respect to the partition $\{R_1, R_2, R_3\}$ to code a point, find a transition matrix A such that the shift space Σ_A describes all possible itineraries coding orbits $\mathcal{O}^+_{f_A}((x, y))$.

(b) Can there be a point $\underline{x} \in \mathbb{T}^2$ whose orbit $\mathcal{O}^+_{f_A}(\underline{x})$ never visits the rectangle R_2? Justify your answer.

(c) Can there be a point $\underline{y} \in \mathbb{T}^2$ whose orbit $\mathcal{O}^+_{f_A}(\underline{y})$ never visits the rectangle R_3? Explain.

References for further reading

- More details on how to construct a conjugacy between the cat map and a shift space can be found in the book by Katok and Hasselblatt (see §7.4.5).
- More details on the coding of the automorphism in Exercise 2.8.1 can be found in the book by Devanay, *An introduction to Chaotic Dynamical Systems*, see §2.4.
- A definition of the properties of a Markov partition can be found in the book by Katok and Hasselblatt (see §10.3.1).