Symbolic dynamics

Jarkko Kari

Spring semester 2025

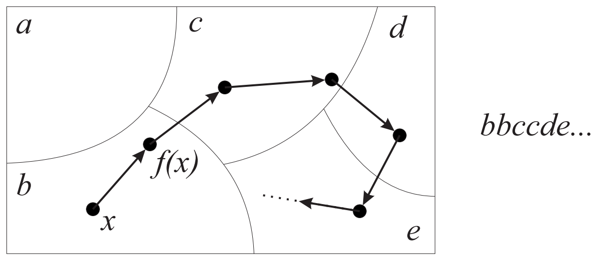
University of Turku

Contents

1 Preliminaries

1.1 Introduction

Discrete-time topological dynamics studies iterates of continuous maps $f : X \longrightarrow X$ on compact metric spaces X. If X is partitioned into a finite number of parts, one obtains information about trajectories of points $x \in X$ from the infinite sequences that record at all times t the part that contains $f^t(x)$:



Depending on f, some finite sequences of parts may not occur in a trajectory of any point. For example, if no point in set a is mapped by f to such a point in c which is mapped into d then the word acd does not occur in any trajectory. This approach leads to the study of infinite words – i.e., infinite sequences of symbols from a finite alphabet – obtained by forbidding some finite subwords. Sets of infinite words obtained by forbidding some finite subwords are called **subshifts**, and they are a central object of study in the field of symbolic dynamics.

If one studies trajectories from initial points, that is, considers time steps $t = 0, 1, 2, \ldots,$ then one obtains words that are one-way infinite and the corresponding subshifts are called one-sided. If f is bijective, i.e. a homeomorphism, one may as well consider bi-infinite trajectories

$$\ldots f^{-2}(x), f^{-1}(x), x, f(x), f^2(x) \ldots,$$

which leads to bi-infinite words and two-sided subshifts.

More generally, we may consider dynamical systems with several transformations $f_1, \ldots, f_n$ that can be applied in arbitrary order. The setup is formalized by viewing the system as a continuous group or semigroup action.

Depending on the complexity of the set of forbidden subwords, one obtains various families of subshifts. A subshift determined by forbidding a finite set of words is called **subshift of finite type**, or SFT. A **sofic subshift** is obtained by forbidding a regular language, that is, words recognized by a finite automaton. If we only assume there is an algorithm that lists the forbidden words, the subshift determined by these words is an **effective subshift**.

The course is structured as follows: We start with a short review of the properties of compact metric spaces as they play a crucial role in the field. We introduce the shift space, the compact metric space of infinite words. We then continue to basic questions in discrete-time topological dynamics, including concepts related to chaos: sensitivity of the system to initial conditions and mixing properties. These are introduced in the setting of arbitrary continuous (semi-)group actions. After these (rather lengthy) preliminaries we finally are able to move on to symbolic dynamics and various types of subshifts.

A few words about our notations: The natural numbers $\mathbb{N} = \{0, 1, 2, \dots\}$ contain 0. For sets A and B, we denote by B^A the set of functions $A \longrightarrow B$. Composition of functions is from right-to-left so $(g \circ f)(x) = g(f(x))$. For any function $f : A \longrightarrow B$, not necessarily injective, and for any $S \subseteq A$ and $T \subseteq B$, we denote $f^{-1}(T) = \{a \in A \mid f(a) \in T\}$ and $f(S) = \{f(s) \mid s \in S\}$ for the pre-image and the image sets of T and S under f, respectively. The restriction of f to $S \subseteq A$ is denoted by $f_{|S}$. For any set A, we denote by $\mathbf{id}_A$ the identity function $A \longrightarrow A$.

Recall also the following algebraic concepts. A **semigroup** is a set A with an associative binary operation $\star$, i.e., $(a \star b) \star c = a \star (b \star c)$ for all $a, b, c \in A$. A semigroup A is a **monoid** if it also has an identity element e that satisfies $a \star e = a$ and $e \star a = a$ for all $a \in A$. A **group** is a monoid A where every element $a \in A$ has an inverse element $a^{-1} \in A$ such that $a \star a^{-1} = a^{-1} \star a = e$, the unique identity element of A.

1.2 A short review of compact metric spaces

Open and closed sets

A **metric space** is a pair (X, d) where X is a set and $d : X \times X \longrightarrow \mathbb{R}$ is a **metric**, a function that measures distances between elements of X. The metric d has to satisfy the following axioms: for all $x, y, z \in X$

(a) $d(x, y) \geq 0$, and $d(x, y) = 0$ if and only if $x = y$; (positivity),

(b) $d(x, y) = d(y, x)$; (symmetry) and

(c) $d(x, y) \leq d(x, z) + d(z, y)$; (the triangle inequality).

For every $\varepsilon > 0$ and $x \in X$ we denote

$$B_\varepsilon(x) = \{y \in X \mid d(x, y) < \varepsilon\}$$

and call $B_\varepsilon(x)$ the (open) ε**-ball** with center x. Let us call a set $U \subseteq X$ **open** iff

$$\forall x \in U,\ \exists \varepsilon > 0\ :\ B_\varepsilon(x) \subseteq U.$$

A set is **closed** if its complement is open, and a set is **clopen** if it is both open and closed.

Proposition 1.1 *Let (X, d) be a metric space. Then*

(i) *$\emptyset$ and X are open,*

(ii) *arbitrary unions of open sets are open, and*

(iii) *intersections of **finitely** many open sets are open.*

Proof. Property (i) is trivial. To see (ii), let $\mathcal{V}$ be an arbitrary family of open sets, and let

$$U = \bigcup_{V \in \mathcal{V}} V$$

be the union of sets V in $\mathcal{V}$. To prove that U is open, consider an arbitrary $x \in U$. We have that $x \in V$ for some $V \in \mathcal{V}$. As V is open there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq V$. But $V \subseteq U$ so that also $B_\varepsilon(x) \subseteq U$.

To prove (iii) consider a set $U = V_1 \cap \cdots \cap V_n$ where V_i are open. Let $x \in U$ be arbitrary. For all $i \in \{1, \ldots, n\}$ we have $x \in V_i$, and because V_i is open there is $\varepsilon_i > 0$ such that $B_{\varepsilon_i}(x) \subseteq V_i$. Taking $\varepsilon = \min\{\varepsilon_1, \ldots, \varepsilon_n\}$ we have that $\varepsilon > 0$ and $B_\varepsilon(x) \subseteq B_{\varepsilon_i}(x) \subseteq V_i$ for all i. Thus $B_\varepsilon(x) \subseteq U$, proving that U is open. □

Corollary 1.2 *A set is open if and only if it is a union of open balls.*

Proof. Let us first show that every open set is a union of open balls: Let U be open. For every $x \in U$ there exists an open ball $V_x \subseteq U$ that contains x. Then U is the union of open balls V_x over all $x \in U$.

For the converse direction it is enough to show that open balls are open, so that by property (ii) of Proposition 1.1 also arbitrary unions of open balls are open: let $U = B_\varepsilon(x)$ be an open ball, and let $y \in U$ be arbitrary. Denote $\varepsilon' = \varepsilon - d(x, y) > 0$. If $z \in B_{\varepsilon'}(y)$ then $d(z, x) \leq d(z, y) + d(y, x) < \varepsilon' + d(x, y) = \varepsilon$, so that $z \in U$. So $B_{\varepsilon'}(y) \subseteq U$, proving that U is open. □

Example 1.3. Let $X = \mathbb{R}$ be the set of real numbers and define $d(x, y) = |x - y|$. Then d is a metric on $\mathbb{R}$, the **usual metric** of real numbers. Open balls are precisely the open intervals (a, b) for $a < b$, and thus open sets are unions on open intervals. Closed intervals $[a, b]$ are examples of closed sets. Set $\mathbb{Q}$ of rational numbers is neither open nor closed because every open interval contains both rational and irrational numbers. The only clopen sets are $\emptyset$ and $\mathbb{R}$. (To see this, note that if $A \subseteq \mathbb{R}$ and $a < b$ are such that $a \in A$ and $b \notin A$ then $c = \sup\{x \mid x \in A, x < b\}$ is such that that for every $\varepsilon > 0$ the set $B_\varepsilon(c)$ contains both elements of A and elements of the complement of A. Thus either A or its complement is not open.) □

A pair $(X, \mathcal{T})$ where X is a set and $\mathcal{T}$ is a family of subsets of X is a **topological space**, and $\mathcal{T}$ is called a **topology** on X, if it satisfies the following three axioms:

(i) $\emptyset \in \mathcal{T}$ and $X \in \mathcal{T}$,

(ii) the union of any number of sets in $\mathcal{T}$ is in $\mathcal{T}$, and

(iii) the intersection of **finitely** many elements of $\mathcal{T}$ is always in $\mathcal{T}$.

By Proposition 1.1 the family of open sets of a metric space (X, d) forms a topology on X. It is called a metric topology. There are also topologies that are not metrizable, i.e., not defined by any metric.

Example 1.4. For any X, let $\mathcal{T}$ contain all subsets of X. Then $\mathcal{T}$ is a topology, the **discrete** topology of X. The discrete topology is metrizable as it is defined by the discrete metric $d(x, y) = 1$ for $x \neq y$ and $d(x, y) = 0$ for $x = y$. This metric satisfies the triangular inequality even in a stronger form

$$d(x, y) \leq \max\{d(x, z), d(z, y)\}.$$

Under this metric all singleton sets $\{x\}$ are open balls.

Also $\{X, \emptyset\}$ is a topology, the **trivial** topology of X. If X contains at least two points then the trivial topology is not defined by any metric: for $x \neq y$, under any metric d, we have $y \notin B_\varepsilon(x)$ for $\varepsilon = d(x, y)$, and thus $B_\varepsilon(x)$ is an open set containing x but not containing y. □

The concept of a topological space emerges as a generalization of a metric space, and it turns out that in many proofs one only needs the axioms (i), (ii) and (iii) of open sets rather than the full power of axioms (a), (b) and (c) of metric spaces. Consistently with this, sets in a topology $\mathcal{T}$ are called open, and their complements are called closed. In topological and symbolic dynamics we mostly consider metrizable spaces, so in the following we focus on properties of metric spaces although several results will be proved for more general topological spaces.

Let (X, d) be a metric space and let $A \subseteq X$. The restriction of d on $A \times A$ is a metric on A, the **induced metric**. Open sets under the induced metric are then precisely the intersections of A and the open sets of X under the original metric d. More generally then, for any topological space $(X, \mathcal{T})$ and for any $A \subseteq X$ we define the **induced topology** on A to be the family $\{A \cap U \mid U \in \mathcal{T}\}$ of the intersections of A and the open sets in the original topology $\mathcal{T}$. In case of a metric space, the induced topology is then the topology defined by the induced metric. By default, we assume the induced metric and the induced topology whenever considering a subset of a metric space or a topological space.

Example 1.5. The usual metric of $\mathbb{R}$ induces on $\mathbb{N}$ a metric that defines the discrete topology. It is not the discrete metric of Example 1.4 but defines the same open sets. □

Note that the concept of a closed set is dual with the concept of an open set. The following properties easily follow using de Morgan's laws.

Proposition 1.6 *Let $(X, \mathcal{T})$ be a topological space.*

(i) The empty set $\emptyset$ is closed, and X is closed,

(ii) the intersection of any number of closed sets is closed, and

(iii) the union of a finite number of closed sets is closed.

A point $x \in X$ is called **isolated** if $\{x\}$ is open. In the case of a metric space this means that there exists $\varepsilon > 0$ such that there are no other elements of X within distance ε from x. A topological space is **perfect** if it has no isolated points.

Example 1.7. The usual topology of $\mathbb{R}$ is perfect because every open ball contains infinitely many points. The discrete topology is far from perfect because every point is isolated. $\square$

Let $A \subseteq X$. The **closure** $\overline{A}$ of A is the intersection of all closed sets that contain A. It is then the smallest closed set that contains A. That is: $\overline{A}$ is closed, and if $A \subseteq F$ for a closed set F then also $\overline{A} \subseteq F$. Notice that A itself is closed if and only if $\overline{A} = A$. Set A is called **dense** in X if $\overline{A} = X$.

Example 1.8. Consider the usual topology of $\mathbb{R}$. The closure of $\mathbb{Q}$ is $\mathbb{R}$, so $\mathbb{Q}$ is dense in $\mathbb{R}$. The closure of the open interval $(0, 1)$ is the closed interval $[0, 1]$, while the set $\mathbb{Z}$ is its own closure; indeed, $\mathbb{Z}$ is a closed set. $\square$

Dual to the closure is the **interior** A° of $A \subseteq X$. It is the union of all open subsets of A. It is the greatest open subset of A. That is: A° is open, and for any open $U \subseteq A$ holds that $U \subseteq A^\circ$. A set A is open if and only if $A^\circ = A$. We call a set A a **neighborhood** of point x if $x \in A^\circ$, that is, if there exists an open set U such that $x \in U \subseteq A$.

Convergence of sequences

A topological space is a **Hausdorff** space if for every $x \neq y$ there are open U_x and U_y such that $x \in U_x$, $y \in U_y$ and $U_x \cap U_y = \emptyset$. In other words, any two distinct points have non-intersecting neighborhoods.

Example 1.9. Every metric topology is Hausdorff. Indeed, if $x \neq y$ then $d(x, y) > 0$. If we choose $\varepsilon = \frac{1}{2} d(x, y)$ then $B_\varepsilon(x)$ and $B_\varepsilon(y)$ are non-intersecting neighborhoods of x and y. The trivial topology $\{\emptyset, X\}$ is not Hausdorff when $|X| \geq 2$. $\square$

A sequence $x_1, x_2, \ldots$ of points of X **converges** to point $x \in X$ if for every open $U \subseteq X$ that contains x there is positive integer n such that $x_i \in U$ for all $i \geq n$. In the metric setting this is equivalent to saying that for every $\varepsilon > 0$ there is n such that $d(x_i, x) < \varepsilon$ for all $i \geq n$.

Note that in general topological spaces a converging sequence may converge to several different points, but if the topology is Hausdorff (e.g. metric) the limit is unique.

Proposition 1.10 *In Hausdorff topology every converging sequence converges to a unique point.*

Proof. Suppose $x_1, x_2, \ldots$ converges to x and y where $x \neq y$. Since X is Hausdorff, there are open sets U and V such that $x \in U$, $y \in V$ and $U \cap V = \emptyset$. By the definition of convergence, $x_i \in U$ and $x_i \in V$ for all sufficiently large i, a contradiction. $\square$

Note: the proposition does not hold in all topological spaces – the Hausdorff assumption is needed. For example, in the trivial topology $\mathcal{T} = \{\emptyset, X\}$ every sequence converges to every point. In a Hausdorff topology we denote by $\lim_{i\to\infty} x_i$ the unique point into which the sequence $x_1, x_2, \ldots$ converges, if it exists. This point is the limit of the sequence.

Another useful property of Hausdorff spaces is that single element sets $\{x\}$ are closed. Indeed, for every $y \neq x$ there exists an open set U_y such that $x \notin U_y$ so that the complement of $\{x\}$ is open as the union of open sets U_y over all $y \neq x$.

Base of a topology

A family $\mathcal{B}$ of open sets is called a **base** (or a **basis**) of a topology iff every open set is the union of some members of $\mathcal{B}$.

Example 1.11. In any metric space (X, d), by Corollary 1.2, open sets are precisely unions of open balls. Thus the family $\{B_\varepsilon(x) \mid x \in X, \varepsilon > 0\}$ of all open balls is a base. □

The following proposition gives a simple condition to check if a family $\mathcal{B}$ is a base.

Proposition 1.12 *Let $(X, \mathcal{T})$ be a topological space. A family $\mathcal{B} \subseteq \mathcal{T}$ of open sets is a base of $\mathcal{T}$ if and only if for every open set U and every $x \in U$ there exists some $B \in \mathcal{B}$ such that $x \in B \subseteq U$.*

Proof. If $\mathcal{B}$ is a base then any open set U is a union of sets in $\mathcal{B}$ and thus for any $x \in U$ there is $B \in \mathcal{B}$ such that $x \in B \subseteq U$.

Conversely, let $\mathcal{B}$ be a family of open sets with the property that for every open U and $x \in U$ there exists $B \in \mathcal{B}$ such that $x \in B \subseteq U$. Unions of sets in $\mathcal{B}$ are clearly open as unions of open sets. To prove that $\mathcal{B}$ is a base of $\mathcal{T}$ consider an arbitrary open set U. For every $x \in U$ there is some $B_x \in \mathcal{B}$ such that $x \in B_x \subseteq U$, and then clearly U is the union of sets B_x over all $x \in U$. □

Compactness

Next we define compactness, a central concept in these lectures. Let $A \subseteq X$ where X is a topological space. A family $\mathcal{U}$ of open sets is called an **open cover** of A if every element of A belongs to some $V \in \mathcal{U}$, that is, if

$$A \subseteq \bigcup_{V \in \mathcal{U}} V.$$

A subfamily $\mathcal{U}' \subseteq \mathcal{U}$ of an open cover $\mathcal{U}$ of A is called a **subcover** if it is also a cover of A.

Set $A \subseteq X$ is called **compact** if every open cover of A has a finite subcover of A. The topology is called compact if the whole space X is compact. In other words, a topology is compact iff every family of open sets whose union is X has a finite subfamily whose union is X.

Example 1.13. In the usual topology of $\mathbb{R}$ the set

$$A = \{0\} \cup \{\frac{1}{n} \mid n \in \mathbb{Z}_+\}$$

is compact. Namely, an open set that contains 0 covers all but finitely many elements of A. So any open cover of A contains a finite subcover: Open set U that covers 0 together with a finite number of open sets that cover the finitely many elements of A that are outside of U.

On the other hand, set $B = \{\frac{1}{n} \mid n \in \mathbb{Z}_+\}$ is not compact. It has an open cover in which every open set covers exactly one element of B. Such cover has no finite subcover. □

The following proposition states the finite intersection property. It is dual to the open cover property we used as the definition. We state the property for the whole space X:

Proposition 1.14 *Topology of X is compact if and only if every family of closed sets whose intersection is empty has a finite subfamily whose intersection is empty.*

Proof. This follows directly from the definition of compactness and de Morgan's laws: A family of open sets is a cover of X if and only if the family of their complements has empty intersection. □

We typically apply the previous proposition in the following set-up:

Corollary 1.15 *Let $F_1 \supseteq F_2 \supseteq F_3 \supseteq \dots$ be an infinite chain of closed sets in a compact space X. If*

$$\bigcap_{i=1}^{\infty} F_i = \emptyset,$$

then $F_i = \emptyset$ for some i. □

The next proposition states that in compact metric spaces sequences of points can not "escape to infinity". In fact, this property characterizes compactness of a metric space, but we only need the implication in one direction. Recall that a **subsequence** of $x_1, x_2, \dots$ is another sequence $x_{i_1}, x_{i_2}, \dots$ where $i_1 < i_2 < \dots$. A subsequence is hence obtained by picking infinitely many elements of the sequence, preserving their relative order.

Proposition 1.16 *Let X be a compact metric space. Every sequence has a converging subsequence.*

Proof. Let $x_1, x_2, \dots$ be arbitrary sequence, $x_i \in X$.

Suppose first that for every $x \in X$ there is some $\varepsilon_x > 0$ such that $x_k \in B_{\varepsilon_x}(x)$ only for finitely many indices k. Clearly the family of $B_{\varepsilon_x}(x)$ over all $x \in X$ is an open cover of X, so by compactness it has a finite subcover. So there is a finite set $A \subseteq X$ such that

$\{B_{\varepsilon_x}(x) \mid x \in A\}$ covers X. But this contradicts the fact that for large enough indices k the elements x_k are not in $B_{\varepsilon_x}(x)$ for any $x \in A$.

Thus we see that there must exist $x \in X$ such that for every $\varepsilon > 0$ one has $x_i \in B_\varepsilon(x)$ for infinitely many indices i. But then the sequence $x_1, x_2, \dots$ has a subsequence that converges to x: There namely is a subsequence whose n'th element belongs to $B_{\frac{1}{n}}(x)$. □

Next we show that in compact metric spaces compact sets are exactly the closed sets.

Proposition 1.17 *If X is a compact topological space then every closed $A \subseteq X$ is compact.*

Proof. Let $A \subseteq X$ be closed. Consider an open cover of A. Together with the complement of A it forms an open cover of X. By compactness of X this has a finite subcover of X, from which we obtain a finite subcover of A by removing the complement of A (if present). Hence A is compact. □

Proposition 1.18 *If X is Hausdorff then every compact $A \subseteq X$ is closed.*

Proof. Let $A \subseteq X$ be compact. Let $x \in X \setminus A$. By the Hausdorff property, for every $a \in A$ there are open sets U_a and V_a such that $a \in U_a$, $x \in V_a$ and $U_a \cap V_a = \emptyset$. Sets U_a form an open cover of A so by compactness of A there is a finite subcover $U_{a_1}, \dots, U_{a_m}$ of A. But then the intersection

$$V_x = V_{a_1} \cap \dots \cap V_{a_m}$$

of the corresponding sets V_{a_i} is an open set satisfying $x \in V_x$ and $V_x \cap A = \emptyset$. The union of sets V_x over all $x \in X \setminus A$ is the complement of A. Since the union is open, we see that A is closed. □

Corollary 1.19 *Let X be a compact metric space. Then $A \subseteq X$ is compact if and only if it is closed.* □

Countability

A topological space is **separable** if it has a countable dense subset, and it is **second countable** if it has a countable base. Compact metric spaces are separable and second countable.

Proposition 1.20 *A compact metric space has a countable base and a countable dense set of points.*

Proof. For every n the cover of X by the open balls $B_{1/n}(x)$ has a finite subcover $\mathcal{B}_n$. The open balls in these finite subcovers for $n = 1, 2, 3, \dots$ form a countable set $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots$ of open sets.

Let us prove that $\mathcal{B}$ is a base. For every open U and every $x \in U$ there exists $\varepsilon > 0$ such that $B_\varepsilon(x) \subseteq U$. Choose an integer $n > 2/\varepsilon$. Because $\mathcal{B}_n$ covers X, there exists $B_{1/n}(y) \in \mathcal{B}_n$ such that $x \in B_{1/n}(y)$. Because $1/n < \varepsilon/2$ we have

$$B_{1/n}(y) \subseteq B_\varepsilon(x) \subseteq U.$$

It follows from Proposition 1.12 that $\mathcal{B}$ is a base.

To find a countable dense set of points, simply take an element from each base set. □

Baire property

A set $A \subseteq X$ is called **residual** if it is the intersection of countably many dense open sets. A topological space X is called a **Baire space** if every residual set is dense. That is, if $U_1, U_2, \ldots$ are open sets such that for all i hold $\overline{U}_i = X$, then the set

$$A = \bigcap_{i=1}^{\infty} U_i$$

is dense. In particular, A is then non-empty.

Intuitively, dense open sets are "big" and contain almost all elements of X, and the Baire property corresponds to the idea in measure theory that countable intersections of full measure sets have full measure.

Proposition 1.21 *Every compact metric space is a Baire space.*

Proof. Let $U_1, U_2, \ldots$ be open dense sets, and let A be their intersection. Let U be an arbitrary non-empty open set. It is enough to prove that $U \cap A \neq \emptyset$. Let us define a sequence $V_0, V_1, V_2, \ldots$ of non-empty open sets as follows: $V_0 = U$, and for every $n \geq 1$, we choose as V_n a non-empty, open set whose closure is a subset of $V_{n-1} \cap U_n$. Such V_n exists for the following reasons: Set $V_{n-1} \cap U_n$ is open, and non-empty by the denseness of U_n. This means that $B_\varepsilon(x) \subseteq V_{n-1} \cap U_n$ for some $\varepsilon > 0$ and $x \in X$. Then $B_{\varepsilon/2}(x)$ can be selected as V_n, because its closure is a subset of $B_\varepsilon(x)$.

Closures of V_n form a decreasing chain

$$\overline{V}_0 \supseteq \overline{V}_1 \supseteq \overline{V}_2 \supseteq \ldots$$

of non-empty closed sets. By Corollary 1.15 their intersection is non-empty. The intersection is a subset of every U_n and also of U, so we conclude that $A \cap U \neq \emptyset$. □

Continuity

Finally, a few words about continuous functions. Let X and Y be two topological spaces. A function $f : X \longrightarrow Y$ is **continuous** at point $x \in X$ if for every open $V \subseteq Y$ that contains $f(x)$ there exists an open neighborhood $U \subseteq X$ of x such that $f(U) \subseteq V$. We call function $f : X \longrightarrow Y$ **continuous** if it is continuous at every $x \in X$.

Example 1.22. If X has the discrete topology then every function $f : X \longrightarrow Y$ is continuous. Also, if Y has the trivial topology $\{\emptyset, Y\}$ then every $f : X \longrightarrow Y$ is continuous. In all topological spaces X and Y all constant functions $f : X \longrightarrow Y$ are continuous. If X has the trivial topology and Y has the discrete topology then the constant functions are the only continuous functions. □

Proposition 1.23 *Let $f : X \longrightarrow Y$ be a function between two topological spaces. The following conditions are equivalent:*

(i) Function $f : X \longrightarrow Y$ is continuous,

(ii) pre-image $f^{-1}(V)$ is open in X for every open $V \subseteq Y$,

(iii) pre-image $f^{-1}(C)$ is closed in X for every closed $C \subseteq Y$.

Proof. (i) $\Longrightarrow$ (ii): Suppose f is continuous and let $V \subseteq Y$ be open. Let $x \in f^{-1}(V)$ be arbitrary, so $f(x) \in V$. From continuity it follows that there is an open $U \subseteq X$ such that $f(U) \subseteq V$ and $x \in U$. This means that $x \in U \subseteq f^{-1}(V)$, which implies that $f^{-1}(V)$ is open.

(ii) $\Longrightarrow$ (i): Suppose $f^{-1}(V)$ is open for every open $V \subseteq Y$. Let $x \in X$ be arbitrary. Let us show that f is continuous at point x. Let $f(x) \in V$ for open $V \subseteq Y$. Then $U = f^{-1}(V)$ is an open set that satisfies $x \in U$ and $f(U) \subseteq V$. So f is continuous at x.

(ii) $\Longleftrightarrow$ (iii): Follows directly from the fact that for every $A \subseteq Y$ holds

$$X \setminus f^{-1}(A) = f^{-1}(Y \setminus A).$$

□

If X and Y are metric spaces with metrics d_X and d_Y, respectively, then continuity of f can be stated as follows:

$$\forall \varepsilon > 0,\ \forall x \in X,\ \exists \delta > 0\ :\ f(B_\delta(x)) \subseteq B_\varepsilon(f(x)).$$

Here, number δ may depend not only on $\varepsilon > 0$ but also on point x. If δ can be chosen independently of x then function f is called **uniformly continuous**:

$$\forall \varepsilon > 0,\ \exists \delta > 0,\ \forall x \in X\ :\ f(B_\delta(x)) \subseteq B_\varepsilon(f(x)).$$

Even more restrictively, if there exists a positive constant r such that $\delta = r \cdot \varepsilon$ works for all x and ε, then f is **Lipschitz continuous**. An **isometry** satisfies $d_Y(f(x), f(y)) = d_X(x, y)$ for all $x, y \in X$, and is hence Lipschitz continuous. So in metric spaces we have the implications

$$f \text{ isometry } \implies f \text{ Lipschitz continuous } \implies f \text{ uniformly continuous } \implies f \text{ continuous.}$$

Note that apart from continuity the concepts above are defined for metric spaces only; continuity is defined also on non-metric topological spaces. For us uniform continuity is particularly important since on compact metric spaces it turns out to be equivalent to continuity:

Proposition 1.24 *Let (X, d_X) and (Y, d_Y) be compact metric spaces. A continuous function $f : X \longrightarrow Y$ is uniformly continuous.*

Proof. Homework. □

Note that compositions of continuous functions are continuous, and compositions of uniformly continuous functions are uniformly continuous.

The following proposition characterizes continuous functions from a metric space in terms of converging sequences. We use this characterization later to study continuity in shift spaces, e.g., in Example 2.7.

Proposition 1.25 *Let X be a metric space and Y a topological space. Then $f : X \longrightarrow Y$ is continuous if and only if for every converging sequence $x_1, x_2, \ldots$ the sequence $f(x_1), f(x_2), \ldots$ converges and*

$$\lim_{i\to\infty} f(x_i) = f(\lim_{i\to\infty} x_i).$$

Proof. "$\Longrightarrow$": Suppose that f is continuous and let $x_1, x_2, \ldots$ be a converging sequence of elements of X. Let $x = \lim_{i\to\infty} x_i$. Let us prove that $f(x_1), f(x_2), \ldots$ converges to $f(x)$. Let U be an open set that contains $f(x)$. Then $f^{-1}(U)$ is open and $x \in f^{-1}(U)$. Because $x_1, x_2, \ldots$ converges to x there is n such that $x_i \in f^{-1}(U)$ for all $i \geq n$. But then $f(x_i) \in U$ for all $i \geq n$.

"$\Longleftarrow$": Let $x \in X$. To prove that f is continuous at point x, we assume the contrary and derive a contradiction. So suppose there is an open $V \subseteq Y$ that contains $f(x)$ such that for every δ there is a point y in $B_\delta(x)$ such that $f(y) \notin V$. Using $\delta = \frac{1}{i}$ for positive integers i we see that for every i there is $x_i \in X$ such that $d(x_i, x) < \frac{1}{i}$ and $f(x_i) \notin V$. Sequence $x_1, x_2, \ldots$ converges to x so by the hypothesis sequence $f(x_1), f(x_2), \ldots$ converges to $f(x) \in V$. But this is not possible since all $f(x_i) \notin V$. □

In all topological spaces, images of compact sets under continuous maps are compact:

Proposition 1.26 *Suppose function $f : X \longrightarrow Y$ is continuous. For every compact A the set $f(A)$ is compact.*

Proof. Consider an open cover of $f(A)$ by open sets V_i. Then, by Proposition 1.23 the sets $f^{-1}(V_i)$ form an open cover of A. By compactness of A there is a finite subcover of A by $f^{-1}(V_i)$ where $i \in F$ for some finite set F. But then the corresponding sets V_i for $i \in F$ form a finite subcover of $f(A)$. Hence $f(A)$ is compact. $\square$

The next proposition implies that continuous bijections between compact metric spaces have continuous inverse functions.

Proposition 1.27 *Let $f : X \longrightarrow Y$ be a continuous bijection where X is a compact and Y is a Hausdorff topological space. Then the inverse function $f^{-1} : Y \longrightarrow X$ is also continuous.*

Proof. By Proposition 1.23 it is enough to show that for every closed $A \subseteq X$ also $f(A)$ is closed. But if $A \subseteq X$ is closed then by Proposition 1.17 it is also compact. By Proposition 1.26 set $f(A)$ is also compact, and then by Proposition 1.18 set $f(A)$ is closed. $\square$

A continuous bijection $f : X \longrightarrow Y$ between topological spaces is a **homeomorphism** if the inverse function $f^{-1} : Y \longrightarrow X$ is also continuous. Proposition 1.27 implies that continuous bijections between compact metric spaces are homeomorphisms. We say that two topological spaces are **homeomorphic** if there is a homeomorphism between them. Homeomorphic spaces are "topologically isomorphic" with each other: they have identical topological properties as the homeomorphism bijectively relates their open sets.

1.3 The Cantor space

The relevant compact metric space in symbolic dynamics is the shift space, which is homeomorphic to the Cantor set. In fact, we prove in this section that all non-empty compact metric spaces with a countable clopen base and without isolated points are homeomorphic to the Cantor set, and hence they can be called **the** Cantor space.

We start the section by introducing the most basic notations and concepts related to languages and words. Let A be a finite set containing at least two elements. We call A an **alphabet**. The elements of the alphabet are called **letters**, and a **word** is a finite sequence of letters. If w is a word then $|w|$ denotes its **length**. (Note the same notation $|S|$ is used for the cardinality of a set S but this should not cause confusion.) The empty word has length 0, and it is denoted by ε. For $n \in \mathbb{N}$, we denote $A^n = \{u_0 \ldots u_{n-1} \mid a_i \in A\}$ for the set of words of length n over the alphabet A. The i'th letter of word u is u_i where the indexing starts with 0 on the leftmost letter. Note that $A^0 = \{\varepsilon\}$ contains exactly one element. The set of all words over alphabet A is $A^* = A^0 \cup A^1 \cup A^2 \cup \ldots$. A **language** is a set of words over a fixed alphabet. The language is **finite** if it contains only a finite number of words.

The **concatenation** uv of two words u and v is the word obtained by writing the first word followed by the second one as a single word. The empty word ε is the identity for concatenation. The n-fold repetition of word u is u^n, e.g. $u^3 = uuu$. In particular, $u^0 = \varepsilon$.

A **prefix** of a word is any sequence of leading symbols of the word, and a **suffix** is any sequence of trailing symbols of the word. A **subword** is any sequence of consecutive symbols that appears in the word. For a word u of length n and integers m, k such that $0 \le m \le k < n$ we denote by $u_{[m,k]} = u_m u_{m+1} \dots u_k$ the subword from position m to position k. By $u_{[m,k)} = u_m u_{m+1} \dots u_{k-1}$ we denote the subword up to (but not including) position k, and we define analogously $u_{(m,k]}$ and $u_{(m,k)}$. Note that $u_{[m,m)}$ is the empty word.

Let us define next infinite words, that we also call **configurations**. For an alphabet A, the set $A^{\mathbb{N}}$ consists of all assignments $\mathbb{N} \longrightarrow A$ of letters to natural numbers. These are one-way infinite words over A. Analogously, elements of $A^{\mathbb{Z}}$ are the bi-infinite words. For $M = \mathbb{N}$ or $M = \mathbb{Z}$, and for any $u \in A^M$, we use the analogous notations u_i, $u_{[m,k]}$, $u_{[m,k)}$ etc. as for the finite words for the letter in position $i \in M$, for the word $u_m \dots u_k$, the word $u_m \dots u_{k-1}$ and so on. When writing down a bi-infinite word we may place a dot before position zero, as in $u = \dots u_{-2}u_{-1}.u_0u_1u_2 \dots$

For a non-empty finite word u we denote by u^∞ the one-way infinite periodic word $uuu\dots$, by ${}^\infty u$ the analogous left-infinite repetition $\dots uuu$ (which is an element in $A^{-\mathbb{N}}$), and by ${}^\infty u^\infty$ the bi-infinite periodic word $\dots uu.uu\dots$. These notations may be combined so that, for example, ${}^\infty(01)1.110^\infty$ is the word $\dots 010101\ 1.11\ 000 \dots$

In the following we consider the set $X = A^{\mathbb{N}}$ of right-infinite words, and define a metric on $A^{\mathbb{N}}$ that makes it a compact metric space. (Analogous considerations can be made on $A^{\mathbb{Z}}$, and actually on A^M for any countable set M.) For $x, y \in A^{\mathbb{N}}$ we define

$$d(x,y) = \begin{cases} 0, & \text{if } x = y, \\ 2^{-\min\{k \mid x_k \neq y_k\}}, & \text{if } x \neq y. \end{cases}$$

The metric considers two configurations to be close to each other if they have a long common prefix and one needs to look far to see the first difference.

Lemma 1.28 *Function $d : A^{\mathbb{N}} \times A^{\mathbb{N}} \longrightarrow \mathbb{R}$ is a metric.*

Proof. We have to check the three defining properties of a metric:

(a) $d(x,y) \ge 0$, and $d(x,y) = 0$ if and only if $x = y$,

(b) $d(x,y) = d(y,x)$, and

(c) $d(x,y) \le d(x,z) + d(z,y)$.

The first two conditions (a) and (b) are immediate. The triangle inequality (c) follows from the fact that for every $k \in \mathbb{N}$, if $x_k \neq y_k$ then $x_k \neq z_k$ or $z_k \neq y_k$. This means that $d(x,z) \ge d(x,y)$ or $d(z,y) \ge d(x,y)$, so even the strong form $d(x,y) \le \max\{d(x,z), d(z,y)\}$ of the triangle inequality holds. □

We call $A^{\mathbb{N}}$ endowed with metric d a one-sided **shift space**. (The reason for this name becomes clear in Examples 2.5 and 2.6.) The open ball of radius $\varepsilon = 2^{-r}$ centered at $x \in A^{\mathbb{N}}$ is

$$B_\varepsilon(x) = \{y \in A^{\mathbb{N}} \mid y_k = x_k \text{ for all } k \le r\}.$$

These sets form a base of our topology. More generally, for any finite domain $D \subseteq \mathbb{N}$ and **finite pattern** $w \in A^D$ we define the **cylinder**

$$[w] = \{x \in A^{\mathbb{N}} \mid x_{|D} = w\}$$

of configurations that contain pattern w in domain D. Note that if $D \subseteq E$ are two finite domains then every cylinder $[w]$ with domain D is a finite union of cylinders with domain E:

$$[w] = \bigcup_{\substack{v \in A^E \\ v_{|D} = w}} [v].$$

Open balls are cylinders. Moreover, any finite domain $D \subseteq \mathbb{N}$ is a subset of the domain $E = \{0, 1, \ldots, r\}$ for a big enough r. Cylinders with domain E are open balls of radius $\varepsilon = 2^{-r}$. Every cylinder is thus a finite union of open balls, and therefore open.

Cylinders with any fixed domain D form a finite partitioning of $A^{\mathbb{N}}$. It follows that every cylinder is also closed. We conclude that **cylinders form a countable clopen base** of the topology. Topological space $A^{\mathbb{N}}$ is also **perfect**: it has no isolated points. Recall that a point x is **isolated** if the singleton $\{x\}$ is open. However, cylinders are infinite so no point can be isolated.

Next we prove that our metric topology on $A^{\mathbb{N}}$ is compact.

Theorem 1.29 *The metric space $(A^{\mathbb{N}}, d)$ is compact.*

Proof. Because there is a countable base $\mathcal{B}$ of cylinders, every open cover $\mathcal{C}$ of $A^{\mathbb{N}}$ has a countable subcover $\mathcal{C}'$: indeed, for every $B \in \mathcal{B}$, choose in the subcover $\mathcal{C}'$ a set $C \in \mathcal{C}$ such that $B \subseteq C$, if such a C exists. Collection $\mathcal{C}'$ has at most as many elements as $\mathcal{B}$, and thus it is a countable subfamily of $\mathcal{C}$. Also $\mathcal{C}'$ covers $A^{\mathbb{N}}$: because $\mathcal{C}$ is a cover and $\mathcal{B}$ is a base, for every $c \in A^{\mathbb{N}}$ there are $B \in \mathcal{B}$ and $C \in \mathcal{C}$ such that $c \in B \subseteq C$. By the construction of $\mathcal{C}'$ there is some $C' \in \mathcal{C}'$ such that $B \subseteq C'$.

Next we show that a countable open cover $\mathcal{C}' = \{V_0, V_1, \ldots\}$ has a finite subcover. For every $n \in \mathbb{N}$, let F_n be the complement of $V_0 \cup \cdots \cup V_n$. Sets F_n are closed and form a decreasing chain $F_0 \supseteq F_1 \supseteq \ldots$. If some F_n is empty then $\{V_0, \ldots, V_n\}$ is a finite subcover of $\mathcal{C}'$, and the proof is complete.

Suppose then that $F_n \neq \emptyset$ for all $n \in \mathbb{N}$. Let us prove that the intersection $F_0 \cap F_1 \cap \ldots$ is non-empty, contradicting the fact that $\mathcal{C}'$ covers $A^{\mathbb{N}}$. (Note that the complement of $F_0 \cap F_1 \cap \ldots$ is the union $V_0 \cup V_1 \cup \ldots$.)

Denote $D_n = \{0, 1, \ldots, n\}$ for all $n \in \mathbb{N}$. There are finitely many patterns with domain D_0 so for some such pattern $p_0 \in A^{D_0}$ the intersections $F_i \cap [p_0]$ are non-empty for all $i \in \mathbb{N}$.

By the same argument, using sets $F_i \cap [p_0]$ in place of sets F_i, there is a pattern p_1 with domain D_1 such that all $F_i \cap [p_0] \cap [p_1]$ are non-empty. Repeating likewise, we obtain for all $j \in \mathbb{N}$ patterns p_j with domain D_j such that

$$F_i \cap [p_1] \cap [p_2] \cap \cdots \cap [p_j]$$

are non-empty for all $i \in \mathbb{N}$.

Define a configuration $c \in A^{\mathbb{N}}$ by $c(k) = p_k(k)$ for all $k \in \mathbb{N}$. The non-emptiness of intersections $[p_1] \cap [p_2] \cap \cdots \cap [p_j]$ means that $p_j(k) = p_k(k) = c(k)$ for all $j \geq k \geq 0$. Thus $c \in [p_j]$ for all $j \in \mathbb{N}$.

Let us prove that $c \in F_i$ for all $i \in \mathbb{N}$. Suppose the contrary: $c \notin F_i$ for some i. Because the complement of F_i is open, there is a cylinder $[c|_D] \subseteq A^{\mathbb{N}} \setminus F_i$. But $D \subseteq D_j$ for some large enough j, and then

$$[p_j] = [c|_{D_j}] \subseteq [c|_D] \subseteq A^{\mathbb{Z}^d} \setminus F_i,$$

contradicting $[p_j] \cap F_i \neq \emptyset$. Thus $c \in F_0 \cap F_1 \cap \ldots$. □

The compactness of $A^{\mathbb{N}}$ can also be obtained as a direct corollary to Tychonoff's theorem, which states that cartesian products of compact spaces are again compact. Our topology on $A^{\mathbb{N}}$ is namely the product of discrete topologies on A, and the finite space A is of course compact.

We have established the following result.

Theorem 1.30 *Metric space $A^{\mathbb{N}}$ is compact. It has a countable base of clopen sets and no isolated points.* □

It follows from compactness, using Proposition 1.29, that every sequence of configuration has a converging subsequence. Next lemma states an intuitive meaning of convergence.

Lemma 1.31 *Let $x^{(1)}, x^{(2)}, \ldots$ be a sequence of configurations, $x^{(i)} \in A^{\mathbb{N}}$. The sequence converges to $x \in A^{\mathbb{N}}$ if and only if for every $k \in \mathbb{N}$ there exists n_k such that $x_k^{(n)} = x_k$ for all $n > n_k$.*

Proof. First suppose that $x^{(1)}, x^{(2)}, \ldots$ converges to x, and let $k \in \mathbb{N}$ be arbitrary. Consider the cylinder $U = [x_k]$ with singleton domain $D = \{k\}$. It is a neighborhood of x so the convergence to x implies that for all sufficiently large n holds $x^{(n)} \in U$, that is, $x_k^{(n)} = x_k$.

Conversely, assume that for every k holds $x_k^{(n)} = x_k$ for all sufficiently large n. It follows then that for any finite $D \subseteq \mathbb{N}$ also holds that $x_{|D}^{(n)} = x_{|D}$ for all sufficiently large n. This means that, for sufficiently large n, the configuration $x^{(n)}$ is in the (unique) cylinder of domain D that contains x. Since cylinders are a base of the topology, it follows that $\lim_{n\to\infty} x^{(n)} = x$, as claimed. □

Example 1.32. Let $x^{(n)} = 1^n 0^\infty$. The sequence $x^{(1)}, x^{(2)}, \ldots$ converges to 1^∞. Let then $y^{(n)} = a0^\infty$ where $a = 0$ for even n and $a = 1$ for odd n. This sequence does not converge. □

Above we have discussed a topology on $A^{\mathbb{N}}$ but the analogous topology can be defined on A^M for any countable infinite set M. **Cylinders** of A^M are defined by **patterns** $w \in A^D$ on finite domains $D \subseteq M$ in the same manner as above:

$$[w] = \{x \in A^M \mid x_{|D} = w\}.$$

Any bijection $\alpha : \mathbb{N} \longrightarrow M$ defines a bijection $\hat{\alpha} : A^M \longrightarrow A^{\mathbb{N}}$ that maps coordinates of configurations according to α, that is, by $\hat{\alpha}(x)_m = x_{\alpha(m)}$ for all $m \in \mathbb{N}, x \in A^M$. Thus the symbols written by x in M are "linearized" into the sequence

$$\hat{\alpha}(x) = x_{\alpha(0)}x_{\alpha(1)}x_{\alpha(2)}\ldots$$

The bijection $\hat{\alpha}$ carries our metric d on $A^{\mathbb{N}}$ over to A^M in the obvious way: the distance of any two configurations $x, y \in A^M$ is $d(\hat{\alpha}(x), \hat{\alpha}(y))$. Thus $\hat{\alpha}$ is an isometry. Isometry $\hat{\alpha}$ is also a bijection between cylinders of $A^{\mathbb{N}}$ and A^M, so the topology on A^M has a clopen base consisting of the cylinders of A^M. For each A^M, the cylinders provide a most convenient base of the topology. Note that the concept of a cylinder does not depend on the choice of α, so that while different choices of α may give different metrics on A^M, the topologies they define are the same.

Convergence of sequences of elements of A^M in the topology has the analogous meaning as in $A^{\mathbb{N}}$: A sequence of configurations converges if and only if for each position $m \in M$ the symbol assigned to m stabilizes to a fixed value after a finite initial segment of the sequence.

So we get, for example, a compact metrizable topology on $A^{\mathbb{Z}}$, the set of two-way infinite configurations, as well as on multidimensional configuration spaces $A^{\mathbb{Z}^d}$ studied in multidimensional symbolic dynamics. These spaces are homeomorphic to $A^{\mathbb{N}}$.

Let us finish the section by showing that all compact metric spaces that have a countable clopen base and that have no isolated points are homeomorphic to each other. A famous example of such a space is the Cantor's middle-thirds set: the subset S of $\mathbb{R}$ containing numbers in the interval $[0, 1]$ that have a ternary (that is, base-3) expansion without digit 1. It is easy to see that the ternary representation of numbers gives a bijection between S and $\{0, 2\}^{\mathbb{N}}$ that takes the usual topology of $\mathbb{R}$ to our topology on $\{0, 2\}^{\mathbb{N}}$.

Theorem 1.33 *Let X be a non-empty compact Hausdorff space that has a countable clopen base and has no isolated points. Then X is homeomorphic to $\{0, 1\}^{\mathbb{N}}$.*

Proof. Let $B_1, B_2, \ldots$ be a clopen base of X, and let $A = \{0, 1\}$. We inductively assign to each finite word $u \in A^*$ a non-empty clopen subset C_u of X as follows:

(i) $C_\varepsilon = X$.

(ii) Assume a non-empty clopen C_u has been defined. We next construct C_{u0} and C_{u1}. Let $k \in \mathbb{N}$ be minimum such that $C_u \cap B_k \neq \emptyset$ and $C_u \cap (X \setminus B_k) \neq \emptyset$. Such k exists because C_u contains at least two points (there are no isolated points so a singleton set is not open), and by the Hausdorff property there is a base set B_k that separates these two points. We set $C_{u0} = C_u \cap B_k$ and $C_{u1} = C_u \cap (X \setminus B_k)$ using the smallest available k. Clearly these sets are again clopen and non-empty.

Note that if u is a prefix of v then $C_v \subseteq C_u$, and note that, for every n, the sets C_u for $u \in A^n$ form a partitioning of X. Note also that if $|u| = k$ then either $C_u \subseteq B_k$ or $C_u \subseteq X \setminus B_k$.

This follows from the fact that on the i'th round of the algorithm, the chosen k in step (ii) must satisfy $k \geq i$ as otherwise that k would have been chosen on an earlier round.

For infinite words $w \in A^{\mathbb{N}}$ we define

$$C_w = \bigcap_{k=0}^{\infty} C_{w_0 \dots w_k}.$$

By Corollary 1.15 the set C_w is non-empty as the intersection of a decreasing sequence of non-empty closed sets. But C_w cannot contain two distinct points: By the Hausdorff property there is a base set B_k that separates these two points, which implies that $C_{w_0 \dots w_k}$ cannot contain both these points. Recall that the recursive construction would have chosen this B_k in step (ii) at latest on the k'th round.

We conclude that for all $w \in A^{\mathbb{N}}$ we have $C_w = \{x_w\}$ for some $x_w \in X$. Now define a function $f : A^{\mathbb{N}} \longrightarrow X$ by $f(w) = x_w$. It is one-to-one (because $C_{u0} \cap C_{u1} = \emptyset$ for all finite words u) and surjective (because for every n, the sets C_u for $u \in A^n$ partition X). Let us prove that f is continuous. If $f(w) \in B_k$ then $C_{w_0 \dots w_k} \cap B_k \neq \emptyset$. But then $C_{w_0 \dots w_k} \subseteq B_k$ so that $f(y) \in B_k$ for any word y that has prefix $w_0 \dots w_k$. In other words, an open neighborhood of w is mapped into B_k. We conclude that $f^{-1}(B_k)$ is open for all base sets B_k, which is enough to show that f is continuous.

A continuous bijection between compact Hausdorff spaces is a homeomorphism (Proposition 1.27). □

Based on the theorem we call any non-empty compact metric space that has a countable clopen base and no isolated points **the Cantor space**.

2 Discrete-time dynamical systems

2.1 Basic concepts and examples

A dynamical system (X, f) consists of a compact metric space X and a continuous function $f : X \longrightarrow X$. Set X is the **phase space** of the system and f is its transformation. If f is bijective (i.e. a homeomorphism) then the system is **invertible**, and the dynamical system (X, f^{-1}) is the **inverse** system.

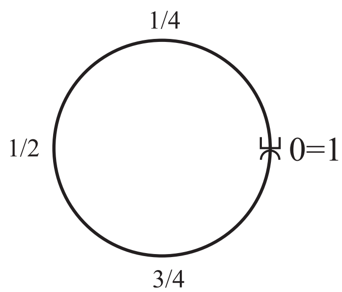
The forward **trajectory** of a point $x \in X$ is the sequence $x, f(x), f^2(x), \dots$. A two-way trajectory is any bi-infinite sequence $(x_i)_{i \in \mathbb{Z}}$ of points $x_i \in X$ that satisfies $x_{i+1} = f(x_i)$ for all $i \in \mathbb{Z}$. The forward **orbit** of x is the set that contains the elements of its forward trajectory, that is, $\mathcal{O}(x) = \{f^n(x) \mid n \in \mathbb{N}\}$. A two-way orbit is a set of points of a two-way trajectory. So we use the term trajectory for a sequence of consecutive points, and the term orbit for the set of these points.

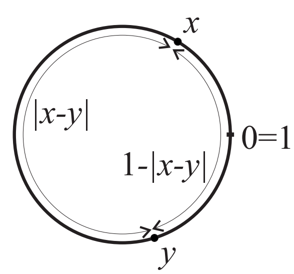
Point $x \in X$ is **periodic** if $f^n(x) = x$, for some $n > 0$. It is **eventually periodic** if $\mathcal{O}(x)$ is finite, that is, if $f^n(x) = f^m(x)$ for some $n \neq m$. A **fixed point** satisfies $f(x) = x$, and an **eventually fixed** point has the property that $f^{n+1}(x) = f^n(x)$, for some $n \in \mathbb{N}$.

In invertible dynamics eventually periodic (or eventually fixed) points are periodic (fixed, respectively).

Let us begin with a number of examples. In addition to the shift spaces $A^{\mathbb{N}}$ and $A^{\mathbb{Z}}$, compact metric spaces used in the examples include the closed unit interval $[0, 1]$ of real numbers and the closed circle $\mathbb{T}$ defined in Example 2.1 below. Recall that compact sets of $\mathbb{R}$ are precisely the bounded closed sets, so $[0, 1]$ is indeed compact.

Example 2.1. Let us denote by $\mathbb{T}$ the set $[0, 1)$ endowed with metric $d(x, y) = \min\{|x - y|, 1 - |x - y|\}$. Set $\mathbb{T}$ is the **circle**: it is useful to think the interval being bent into a ring, gluing together points 0 and 1, as shown on the left illustration below. Distance d is then the shortest distance between given points along the ring, as seen in the picture on the right:





This is indeed a metric (homework). For any $r \in \mathbb{R}$ we denote the integer and the fractional parts of r by

$$\begin{aligned} \lfloor r \rfloor &= \max\{n \in \mathbb{Z} \mid n \leq r\}, \text{ and} \\ \text{frac}(r) &= r - \lfloor r \rfloor. \end{aligned}$$

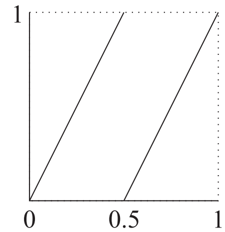
Then $\text{frac}(r) \in \mathbb{T}$ for all $r \in \mathbb{R}$, and the mapping $r \mapsto \text{frac}(r)$ from $\mathbb{R}$ to $\mathbb{T}$ is continuous. Continuity follows from the fact that $d(\text{frac}(x), \text{frac}(y)) \leq |x - y|$, so that $r \mapsto \text{frac}(r)$ is even Lipschitz continuous. It also follows that the circle $\mathbb{T}$ is compact as the image of the compact set $[0, 1]$ under the continuous function $r \mapsto \text{frac}(r)$. □

We denote $r = s \pmod 1$ when $\text{frac}(r) = \text{frac}(s)$.

Example 2.2. Let $X = \mathbb{T}$ be the circle from Example 2.1. For $\alpha \in \mathbb{R}$ define the **rotation** by α as the function $\rho_\alpha : \mathbb{T} \longrightarrow \mathbb{T}$ that maps $x \mapsto \text{frac}(x + \alpha)$. Function $f = \rho_\alpha$ is a homeomorphism of the circle. In fact, it is an isometry as $d(x, y) = d(f(x), f(y))$ holds for all $x, y \in \mathbb{T}$.

If $\alpha = \frac{m}{n}$ is rational then $f = \rho_\alpha$ has finite order because $f^n(x) = x + n\alpha = x + m = x$ (mod 1), for all $x \in \mathbb{T}$. So $f^n = \mathbf{id}_{\mathbb{T}}$ and all points are periodic. If α is irrational then the orbit of every point is dense. To see this notice that $\mathcal{O}(0)$ is infinite so that for every $\varepsilon > 0$ there exist $n_1 < n_2$ such that $d(f^{n_1}(0), f^{n_2}(0)) < \varepsilon$. But then for $n = n_2 - n_1$ we have $0 < d(f^n(0), 0) < \varepsilon$. Function f^n is a rotation by $f^n(0)$ so it is clear that $B_\varepsilon(f^{in}(x))$ for $i \in \mathbb{N}$ cover $\mathbb{T}$, for all initial $x \in \mathbb{T}$. We conclude that for all $x, y \in \mathbb{T}$ and all $\varepsilon > 0$, the orbit $\mathcal{O}(x)$ contains points within distance ε from y. All orbits are hence dense. □

Example 2.3. Again, we consider the circle $\mathbb{T}$. The **doubling map** $\times_2 : \mathbb{T} \longrightarrow \mathbb{T}$ maps $x \mapsto \mathrm{frac}(2x)$:



Function $\times_2$ is (Lipschitz) continuous on $\mathbb{T}$, but it is not injective because $\times_2(x) = \times_2(x+\frac{1}{2})$. Every rational point $x = \frac{m}{n}$ is eventually periodic because its trajectory only contains rational points with denominator n, and there are only finitely many such points. For example, the trajectory $\frac{1}{6}, \frac{1}{3}, \frac{2}{3}, \frac{1}{3}, \ldots$ of $x = \frac{1}{6}$ is eventually periodic but not periodic. Only rational numbers with an odd denominator are periodic. Irrational numbers have infinite orbits.

Analogously, for any integer n, we define the multiplication $\times_n : x \mapsto \mathrm{frac}(nx)$ by n. □

Example 2.4. Let $X = [0, 1]$ under the usual metric, and let $f(x) = x^2$. Function f is a homeomorphism of $[0, 1]$ so we have an invertible dynamical system. In this system the trajectories of all initial points except $x = 1$ converge to limit 0. In the inverse system $([0, 1], f^{-1})$ all trajectories except for $x = 0$ converge to 1. □

Example 2.5. Let $X = A^{\mathbb{Z}}$ be the two-sided shift space defined in Section 1.3 where A is an alphabet with at least two letters. Define the two-sided **left shift** $\sigma : A^{\mathbb{Z}} \longrightarrow A^{\mathbb{Z}}$ as follows: $\sigma(x)_k = x_{k+1}$ for all $x \in A^{\mathbb{Z}}$ and all $k \in \mathbb{Z}$. Function σ translates infinite words one position to the left. The left shift is clearly continuous since the pre-images of cylinders are cylinders. It is also bijective and its inverse is the right shift σ^{-1}. Periodic points are precisely the words ${}^{\infty}u^{\infty}$ where u is a non-empty finite word. The dynamical system $(A^{\mathbb{Z}}, \sigma)$ is called a (two-sided) **full shift**.

□

Example 2.6. Let $X = A^{\mathbb{N}}$ be the one-sided shift space. The one-sided left shift $\sigma : A^{\mathbb{N}} \longrightarrow A^{\mathbb{N}}$ is defined analogously to the two sided-case by $\sigma(x)_k = x_{k+1}$ for all $x \in A^{\mathbb{N}}$ and all $k \in \mathbb{N}$. One-sided shift is not bijective because it erases the first letter of the infinite word: $\sigma(ax) = x$ for all $a \in A$ and $x \in A^{\mathbb{N}}$. The dynamical system $(A^{\mathbb{N}}, \sigma)$ is called a **one-sided full shift**. □

Example 2.7. Let $X = A^{\mathbb{Z}}$ be again the two-sided shift space. Let $D \subseteq \mathbb{Z}$ be finite, the neighborhood, and let $\phi : A^D \longrightarrow A$ be a function, the local rule. Define $f : A^{\mathbb{Z}} \longrightarrow A^{\mathbb{Z}}$ as follows: for all $x \in A^{\mathbb{Z}}$ and $k \in \mathbb{Z}$ we have $f(x)_k = \phi(\sigma^k(x)_{|D})$, where σ is the left shift. In other words, the new symbol in position k is obtained by applying the local rule ϕ on the pattern seen in x in domain $k + D$, that is, the D-pattern around position k. Function f

is a **cellular automaton**. To prove that f is continuous, consider any converging sequence $x^{(1)}, x^{(2)}, \dots$ of configurations, with limit $\lim_{i\to\infty} x^{(i)} = x$. The convergence means that for every $k \in \mathbb{Z}$, the symbols in domain $k + D$ eventually stabilize in the sequence, so that for all large enough n

$$\sigma^k(x^{(n)})_{|D} = x^{(n)}_{|k+D} = x_{|k+D} = \sigma^k(x)_{|D}.$$

But then also $f(x^{(n)})_k = f(x)_k$ for all large enough n. This means that the sequence $f(x^{(1)}), f(x^{(2)}), \dots$ converges to $f(x)$. By Proposition 1.25 function f is continuous. If f is a bijection then f is a **reversible** cellular automaton.

Notice also that cellular automaton f commutes with the left shift σ of Example 2.5: we have $f \circ \sigma = \sigma \circ f$ because for all $x \in A^{\mathbb{Z}}$ and $k \in \mathbb{Z}$ holds

$$(f \circ \sigma)(x)_k = \phi(\sigma^k(\sigma(x))_{|D}) = \phi(\sigma^{k+1}(x)_{|D}) = f(x)_{k+1} = (\sigma \circ f)(x)_k.$$

We'll see in Section 3 that cellular automata are the only continuous functions $A^{\mathbb{Z}} \longrightarrow A^{\mathbb{Z}}$ that commute with the left shift (Theorem 3.7).

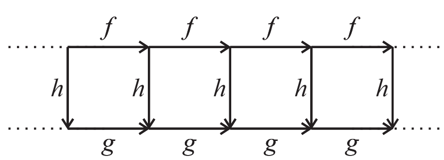
Cellular automata on the one-sided shift space $A^{\mathbb{N}}$ are defined analogously. In this case the neighborhood $D \subseteq \mathbb{N}$ is also one-sided, so a cell does not see any cells on its left. Analogously we can define a multidimensional cellular automaton on the configuration space $A^{\mathbb{Z}^d}$, or more generally, a cellular automaton over group or even a monoid M, in which case the configurations are elements of A^M and the neighborhood is some finite $D \subseteq M$. □

Homomorphisms

Let (X, f) and (Y, g) be dynamical systems. A function $h : X \longrightarrow Y$ is a **homomorphism** if it is continuous and has the commutation property

$$h \circ f = g \circ h.$$

In this case $h \circ f^n = g^n \circ h$ for all $n \in \mathbb{N}$. The following diagram commutes



so the trajectories of f are mapped to trajectories of g. In mathematics in general, homomorphisms are structure preserving maps: In the case of dynamical systems the structures to be preserved are the topology (preserved by continuity) and the transformation (preserved by the commutation property).

A surjective homomorphism is called a **factor map**, and in this case system (Y, g) is a **factor** of (X, f). An injective homomorphism is an **embedding**. A bijective homomorphism is a **conjugacy**. Notice that a conjugacy is a homeomorphism between the metric spaces X and Y. Two dynamical systems are (topologically) **conjugate** if there is a conjugacy between them. Conjugate systems are “the same”: one just has to look into one “through the homeomorphism” h to see the other.

It follows from the commutation property that $f^n(x) = x$ implies $g^n(h(x)) = h(x)$. It means that a homomorphism has to map a periodic point x to a periodic point and, moreover, the shortest period of the image has to divide the period of x. Similar property applies to eventually periodic points: if $f^n(x) = f^m(x)$ then $g^n(h(x)) = g^m(h(x))$.

Example 2.8. We return to rotations from Example 2.2. For every α, the system $(\mathbb{T}, \rho_{2\alpha})$ is a factor of $(\mathbb{T}, \rho_\alpha)$. A factor map is the doubling map $\times_2 : x \mapsto \text{frac}(2x)$ of Example 2.3. Continuity and surjectivity of $\times_2$ are clear, and the commutation property $\times_2 \circ \rho_\alpha = \rho_{2\alpha} \circ \times_2$ is seen by the direct calculation

$$(\times_2 \circ \rho_\alpha)(x) = 2(x + \alpha) = 2x + 2\alpha = (\rho_{2\alpha} \circ \times_2)(x) \pmod 1.$$

□

Example 2.9. Consider then the doubling and the tripling systems $(\mathbb{T}, \times_2)$ and $(\mathbb{T}, \times_3)$. System $(\mathbb{T}, \times_2)$ has only one fixed point $x = 0$ while $(\mathbb{T}, \times_3)$ has two fixed points 0 and $\frac{1}{2}$ so the systems are not conjugate. □

Example 2.10. Consider full shifts $(A^{\mathbb{Z}}, \sigma_A)$ and $(B^{\mathbb{Z}}, \sigma_B)$ over alphabets $A = \{0, 1\}$ and $B = \{0, 1, 2\}$. Here, σ_A and σ_B are the left shifts. We have $A^{\mathbb{Z}} \subseteq B^{\mathbb{Z}}$ and the identity $\mathbf{id}_{|A^{\mathbb{Z}}}$ is an embedding of $A^{\mathbb{Z}}$ to $B^{\mathbb{Z}}$. On the other hand, the function h that changes every symbol 2 in all configurations into symbol 1 is a factor map from $B^{\mathbb{Z}}$ to $A^{\mathbb{Z}}$. There is no embedding from $(B^{\mathbb{Z}}, \sigma_B)$ to $(A^{\mathbb{Z}}, \sigma_A)$ because the first one has three fixed points and the second one only two. There is also no factor map from $(A^{\mathbb{Z}}, \sigma_A)$ to $(B^{\mathbb{Z}}, \sigma_B)$. This follows easily from entropy considerations that we learn later. □

Example 2.11. Let us find a factor map from a full shift to the doubling map. For $x \in \{0, 1\}^{\mathbb{N}}$ we define

$$(x)_2 = \sum_{k=0}^{\infty} x_k 2^{-k-1}$$

to be the real number in interval $[0, 1]$ with binary expansion $0.x_0x_1x_2x_3\ldots$. We further identify the boundaries 0 and 1 of the interval, and define the function $\varphi : \{0, 1\}^{\mathbb{N}} \longrightarrow \mathbb{T}$ by $\varphi(x) = \text{frac}((x)_2)$. If σ is the left shift on $\{0, 1\}^{\mathbb{N}}$ and $\times_2$ is the doubling map on $\mathbb{T}$ then $\varphi \circ \sigma = \times_2 \circ \varphi$. This is verified by a direct calculation: For any $x \in \{0, 1\}^{\mathbb{N}}$

$$(\varphi \circ \sigma)(x) = \sum_{k=0}^{\infty} x_{k+1} 2^{-k-1} = \sum_{i=1}^{\infty} x_i 2^{-i} = \sum_{i=0}^{\infty} x_i 2^{-i} = 2 \sum_{i=0}^{\infty} x_i 2^{-i-1} = (\times_2 \circ \varphi)(x) \pmod 1.$$

Let us also verify that φ is continuous. If $x, y \in \{0, 1\}^{\mathbb{N}}$ start with the same prefix of length n then $|(x)_2 - (y)_2| \le \sum_{k=n}^{\infty} 2^{-k} = 2^{-n+1}$ in the usual metric of $[0, 1]$. It follows that

$x \mapsto (x)_2$ is continuous, and as $\varphi(x)$ is the composition of this function and the continuous function $x \mapsto \text{frac}(x)$, also φ is continuous. Clearly φ is surjective so it is a factor map from $(\{0,1\}^{\mathbb{N}}, \sigma)$ to $(\mathbb{T}, \times_2)$.

Note that φ is not injective because there are numbers with two different binary expansions, one ending $111\ldots$ and one ending $000\ldots$. Systems $(\{0,1\}^{\mathbb{N}}, \sigma)$ and $(\mathbb{T}, \times_2)$ are in fact not conjugate because any conjugacy would be a homeomorphism between metric spaces $\{0,1\}^{\mathbb{N}}$ and $\mathbb{T}$. These spaces are not homeomorphic because the only clopen subsets of $\mathbb{T}$ are $\mathbb{T}$ and $\emptyset$, i.e., $\mathbb{T}$ is not the Cantor space. □

A homomorphism from a dynamical system to itself is an **endomorphism**, and a conjugacy from a system to itself is an **automorphism**. Automorphisms of (X, f) form a group under the operation of function composition, the **automorphism group** of (X, f). Similarly, the endomorphisms form a monoid, the **endomorphism monoid** of (X, f).

Example 2.12. Cellular automata $f : A^{\mathbb{Z}} \longrightarrow A^{\mathbb{Z}}$, defined in Example 2.7, are endomorphisms of the full shift $(A^{\mathbb{Z}}, \sigma)$, defined in Example 2.5. Reversible cellular automata are automorphisms. We'll see in Section 3 that these are the only endomorphisms and automorphisms. □

Example 2.13. Circle rotations, defined in Example 2.2, commute with each other, so ρ_α is an automorphism of $(\mathbb{T}, \rho_\beta)$, for all $\alpha, \beta \in \mathbb{R}$. If β is irrational then there are no other automorphisms. □

Dynamical systems as monoid actions

In the dynamical systems above a single function f was iterated over time. The system can also be understood as a function $X \times \mathbb{N} \longrightarrow X$ defined by $(x, k) \mapsto f^k(x)$, where we have taken the "time" k as a second argument. The possible times k form the monoid $\mathbb{N}$ under the operation of addition. We have the obvious property $f^{m+k} = f^k \circ f^m$ that links the monoid operation to composition of transformations.

We can generalize this. Let M be a monoid with identity 1_M. We use the multiplicative notation for the monoid operation. A (right) **monoid action** of M on set X is a function $f : X \times M \longrightarrow X$ that satisfies the conditions (i) and (ii) below. For every $m \in M$ we denote by $f^m : X \longrightarrow X$ the function $x \mapsto f(x, m)$, the action of m on X. It is the transformation by generalized "time" m.

(i) $f^{1_M} = \mathbf{id}_X$ (the identity of M acts as the identity function on X),

(ii) $f^{mk} = f^k \circ f^m$ for all $m, k \in M$ (the action by a product of two monoid elements is the composition of the actions by the elements).

A **dynamical system over monoid** M is a pair (X, f) where X is a compact metric space and f is a monoid action of M on X such that for all $m \in M$ the function f^m is continuous. Note that if M is a group then each f^m is bijective with the inverse function $f^{(m^{-1})}$.

Our earlier definition of dynamical systems (X, f) coincides with this definition for the additive monoid $\mathbb{N}$ if we choose as the monoid action the mapping $(x, k) \mapsto f^k(x)$. This action is uniquely determined by f. If f is a bijection then we can view (X, f) either as a dynamical system over $\mathbb{N}$ or as a dynamical system over $\mathbb{Z}$. In the latter case the action $(x, k) \mapsto f^k(x)$ for negative k uses the inverse function of f. Remark that in the new notation, function f in (X, f) is the monoid action $X \times \mathbb{N} \longrightarrow X$, while in our previous notation it denoted the action f^1 by $1 \in \mathbb{N}$ on X. When there is no risk of confusion we continue denoting f^1 by f, and in the notation (X, f) for dynamical systems over $\mathbb{N}$ and $\mathbb{Z}$, letter f stands both for the monoid action and for f^1.

More generally, if monoid M is generated by $G \subseteq M$ then an M-action f is uniquely determined by functions f^g for generators $g \in G$. Indeed, for $m = g_1 \dots g_k$ we have $f^m = f^{g_k} \circ \dots \circ f^{g_1}$.

Remark: One can define left monoid actions analogously but the property (ii) becomes $f^{mk} = f^m \circ f^k$. We choose to use right actions so that the monoid operations and function compositions are in the compatible order to conveniently define shift actions by M on configurations in A^M, as discussed in the following example.

Example 2.14. Let A be a finite alphabet and let M be a countable monoid. The **shift action** σ of M on the configuration space A^M is defined by $\sigma^m(x)_k = x_{mk}$ for all $m, k \in M$ and all $x \in A^M$. In other words, when shifting by $m \in M$ the contents of cell mk get moved to cell k. This clearly satisfies condition (i) in the definition of right actions, and condition (ii) is verified by

$$\sigma^{mn}(x)_k = x_{(mn)k} = x_{m(nk)} = \sigma^m(x)_{nk} = \sigma^n(\sigma^m(x))_k.$$

The two-sided and the one-sided full shifts of Examples 2.5 and 2.6 correspond to the cases $M = \mathbb{Z}$ and $M = \mathbb{N}$, respectively. (Although, strictly speaking, the two-sided full shift was defined as an $\mathbb{N}$-action on $A^{\mathbb{Z}}$. But from now on we take it as a $\mathbb{Z}$-action and include also the negative shifts, i.e., shifts to the right.) □

Orbit closures and subsystems

We now have a more general concept of a dynamical system. In these lectures we mostly consider $\mathbb{N}$- and $\mathbb{Z}$-actions, but the general framework has the advantage that we no longer need to give separate definitions for one-sided and two-sided concepts. For example, the **trajectory** of point $x \in X$ is the element of $t \in X^M$ defined by $t_m = f^m(x)$, and its **orbit** is the set $\mathcal{O}(x) = \{f^m(x) \mid m \in M\}$. This captures both the concept of a forward orbit (corresponding to $M = \mathbb{N}$) and two-way infinite orbit when the system is invertible (corresponding to $M = \mathbb{Z}$).

Also the concept of a homomorphism readily works in the more general setting: A **homomorphism** between two dynamical systems (X, f) and (Y, g) over the same monoid M is a continuous function $h : X \longrightarrow Y$ that satisfies $h \circ f^m = g^m \circ h$ for all $m \in M$. Again, this coincides with our previous definition (and in this case there is no difference if invertible dynamical systems are considered as $\mathbb{N}$- of $\mathbb{Z}$-actions). It is clearly enough to verify the condition $h \circ f^m = g^m \circ h$ for generators m of M.

We can define new concepts more uniformly. Let (X, f) be a dynamical system over any monoid M. Subset $Y \subseteq X$ of the phase space is **invariant** if $f^m(Y) \subseteq Y$ for all $m \in M$. If non-empty Y is closed and invariant then restricting f to $Y \times M$ gives a dynamical system $(Y, f_{|Y\times M})$, a **subsystem** of (X, f). We simply say that $Y \subseteq X$ is a subsystem when f is clear from the context. Note that here it makes a difference whether an invertible system is viewed as a dynamical system over $\mathbb{N}$ or $\mathbb{Z}$: a subsystem over $\mathbb{N}$ may be non-invariant under the $\mathbb{Z}$-action. An non-empty intersection of subsystems is again a subsystem.

The orbit $\mathcal{O}(x)$ of a point is invariant by definition. However, it is not necessarily closed. To get a subsystem, we consider its topological closure $\overline{\mathcal{O}(x)}$, the **orbit closure** of x

Lemma 2.15 *The orbit closure $\overline{\mathcal{O}(x)}$ of x is a subsystem, the smallest subsystem that contains x.*

Proof. Set $\overline{\mathcal{O}(x)}$ is closed by definition. Let us prove invariance. Let $y \in \overline{\mathcal{O}(x)}$ and $m \in M$ be arbitrary. There are $y_1, y_2, \dots \in \mathcal{O}(x)$ such that $\lim_{i\longrightarrow\infty} y_i = y$. Then each $f^m(y_i) \in \mathcal{O}(x)$ and, because f^m is continuous, $\lim_{i\longrightarrow\infty} f^m(y_i) = f^m(y)$. This means that $f^m(y) \in \overline{\mathcal{O}(x)}$. We conclude that the orbit closure is a subsystem. It is the smallest subsystem containing x: If $Y \subseteq X$ is any subsystem containing x then $\mathcal{O}(x) \subseteq Y$ by invariance of Y, and $\overline{\mathcal{O}(x)} \subseteq Y$ because Y is closed. □

By the same proof, more generally, if Y is any invariant set then its closure $\overline{Y}$ is a subsystem.

We call point $x \in X$ **transitive** if $\overline{\mathcal{O}(x)} = X$. So the trajectories of transitive points explore the whole phase space.

Example 2.16. Consider the circle rotation $(\mathbb{T}, \rho_\alpha)$ by irrational α, introduced in Example 2.2. We can view it as a system over $\mathbb{N}$ or $\mathbb{Z}$. Either way, all $x \in \mathbb{T}$ are transitive because by Example 2.2 already the forward orbits are all dense. In contrast, if α is rational then there are no transitive points because all orbits are finite. □

2.2 Mixing properties

Some dynamical systems (X, f) mix the phase space X in the sense that all parts of X contain points that evolve to the vicinity of all other points. This mimics situations such as stirring a drop of milk in a cup of coffee. We have several variants of the mixing property.

Transitivity

Dynamical system (X, f) over monoid M is (topologically) **transitive** if for all non-empty open $U, V \subseteq X$ there exists $m \in M$ and $x \in U$ such that $f^m(x) \in V$. In other words, it must be possible to get from any open set to any other open set by the dynamics. The condition $\exists x \in U : f^m(x) \in V$ can be written equivalently as $(f^m)^{-1}(V) \cap U \neq \emptyset$ or as $f^m(U) \cap V \neq \emptyset$.

Transitivity of the system is closely related to existence to transitive points, as shown by the following propositions. Recall that residual sets are countable intersections of dense open sets, and that by Proposition 1.21 they are non-empty in compact metric spaces.

Proposition 2.17 *Let M be a group. A dynamical system (X, f) over M is transitive if and only if it has a transitive point. In this case, the set of transitive points is residual.*

Proof. Let $T \subseteq X$ be the set of transitive points. We prove a cycle of three implications:

$$(X, f) \text{ is transitive} \implies T \text{ is residual} \implies T \neq \emptyset \implies (X, f) \text{ is transitive}$$

The first two implications hold on any monoid M but the proof of the third implication uses the group structure of M.

• (X, f) is transitive $\Longrightarrow T$ is residual: By Proposition 1.20 the topology has a countable base $B_1, B_2, \ldots$. For every $i = 1, 2 \ldots$, let

$$U_i = \bigcup_{m \in M} (f^m)^{-1}(B_i) = \{x \in X \mid \mathcal{O}(x) \cap B_i \neq \emptyset\}$$

be the set of points that can be mapped to B_i by the system. Set U_i is open as a union of open sets, and it is dense because by transitivity it contains an element of every non-empty open set. Let U be the intersection of all U_i. Set U is residual as a countable intersection of dense open sets. We have $U = T$: indeed, $x \in T$ if and only if $\mathcal{O}(x)$ intersects every base set B_i. But this is equivalent to $x \in U_i$ for every i, that is, equivalent to $x \in U$.

• T is residual $\Longrightarrow T \neq \emptyset$: Compact metric spaces are Baire-spaces by Proposition 1.21, so every residual set is non-empty.

• $T \neq \emptyset \Longrightarrow (X, f)$ is transitive: This is the only implication where we use the fact that M is a group. Let $x \in T$. Consider arbitrary non-empty open sets U and V. By transitivity of x, there are $m, n \in M$ such that $f^m(x) \in U$ and $f^n(x) \in V$. Let $y = f^m(x)$. Then $y \in U$ and $f^{m^{-1}n}(y) = (f^{m^{-1}n} \circ f^m)(x) = f^{mm^{-1}n}(x) = f^n(x) \in V$. □

The last implication of the proof used the fact M is a group, and the following example shows that the existence of a transitive point does not guarantee the transitivity of the system when $M = \mathbb{N}$.

Example 2.18. Let $X = \{1, \frac{1}{2}, \frac{1}{4}, \ldots\} \cup \{0\}$ under the usual metric of reals. This set is compact (similar to Example 1.13). Define $f(x) = \frac{x}{2}$. Then (X, f) is a dynamical system over $\mathbb{N}$ and $x = 1$ is a transitive point. However, the system is not transitive because no point of the open set $\{\frac{1}{2}\}$ can be mapped to the open set $\{1\}$. □

The transformation f in Example 2.18 is not surjective. In fact, over $\mathbb{N}$, transitivity is equivalent to being surjective and having a transitive point.

Proposition 2.19 *A dynamical system (X, f) over $\mathbb{N}$ is transitive if and only if it has a transitive point and $f : X \longrightarrow X$ is surjective. In this case the set of transitive points is residual.*

Proof. The first two implications in the proof of Proposition 2.17 apply to any monoid M, so transitivity implies that the set of transitive points is residual and, in particular, that there is a transitive point. Transitivity also implies surjectivity: Suppose f is not surjective. The image of a compact set under a continuous function is compact, so $f(X)$ is topologically closed. Its complement is a non-empty open set U. Note that $f^m(X) \cap U = \emptyset$ for all $m > 0$. Pick $x, y \in X$ such that $x \in U$ and $y \neq x$. (Clearly non-surjectivity implies that $|X| \geq 2$ so such y exists.) By the Hausdorff property there are disjoint open neighborhoods $V_x \subseteq U$ and V_y of x and y respectively. Now $f^m(V_y) \cap V_x = \emptyset$ for all $m > 0$ and also $f^0(V_y) = V_y$ has empty intersection with V_x. We conclude that no point of V_y can be mapped to V_x, a contradiction with the transitivity of the system.

Conversely, suppose the system is surjective and there is a transitive point x. Transitivity of f can be proved as follows. Let U, V be two arbitrary non-empty open sets. By transitivity of x there exists $n \in \mathbb{N}$ such that $f^n(x) \in U$. The set $(f^n)^{-1}(V)$ is open and, by virtue of surjectivity, non-empty. So there exists $m \in \mathbb{N}$ such that $f^m(x) \in (f^n)^{-1}(V)$, that is, $f^{n+m}(x) \in V$. Point $y = f^n(x)$ now satisfies $y \in U$ and $f^m(y) \in V$, confirming transitivity of f.

□

Remark: The second part of the proof only used commutativity of $\mathbb{N}$, so by an analogous proof we see that for any commutative monoid M a system (X, f) is transitive if it has a transitive point and if all f^m are surjective. However, the other implication fails: There are systems over $M = \mathbb{N}^2$ that are transitive but not surjective.

Mixingness

The following stronger mixing property we only define for the case $M = \mathbb{N}$. Dynamical system (X, f) over $\mathbb{N}$ is (topologically) **mixing** if for all non-empty open $U, V \subseteq X$ the transitivity condition $f^m(U) \cap V \neq \emptyset$ holds for all sufficiently large m. In other words, for all U, V there exists $n \in \mathbb{N}$ such that for all $m > n$ there is $x \in U$ such that $f^m(x) \in V$.

Clearly a mixing system is transitive. Moreover, a mixing system has the property that eventually points from all open sets get close to each other: Let U, V and W be any non-empty open sets. If f is mixing then for all sufficiently large n we have $f^n(U) \cap W \neq \emptyset$ and $f^n(V) \cap W \neq \emptyset$ simultaneously.

Example 2.20. An isometry is not mixing if X contains at least two points: Let $x \neq y$ and let $\varepsilon = \frac{1}{4}d(x, y)$. Denote $U = B_\varepsilon(x)$ and $V = B_\varepsilon(y)$. Then for all $x' \in U$, $y' \in V$ and $n \in \mathbb{N}$ we have $d(f^n(x'), f^n(y')) > 2\varepsilon$. In particular, there is no n such such that $f^n(U) \cap U \neq \emptyset$ and $f^n(V) \cap U \neq \emptyset$ simultaneously. If f were mixing such n would exist.

We conclude that a rotation ρ_α of the circle $\mathbb{T}$ is not mixing for any α because it is an isometry. □

Example 2.21. Full shifts $A^{\mathbb{N}}$ and $A^{\mathbb{Z}}$ are mixing. Let $[u]$ and $[v]$ be arbitrary cylinders, with domains D and E respectively. All large enough n have the property $D \cap (n + E) = \emptyset$ so that there exists a configuration x with pattern u in domain D and pattern v in domain $n + E$. But this means that $x \in [u]$ and $\sigma^n(x) \in [v]$. □

It is a simple observation that factor maps preserve the mixing properties:

Proposition 2.22 *Let (X, f) and (Y, g) be dynamical systems over monoid M, and let $h : X \longrightarrow Y$ be a factor map.*

(a) If (X, f) is transitive so is (Y, g).

(b) In the case $M = \mathbb{N}$, if (X, f) is mixing so is (Y, g).

Proof. Let $U, V \subseteq Y$ be non-empty open sets. Then $h^{-1}(U)$ and $h^{-1}(V)$ are non-empty open subsets of X. If $m \in M$ is such that $f^m(h^{-1}(U)) \cap h^{-1}(V) \neq \emptyset$ then also $f^m(U) \cap V \neq \emptyset$. □

Example 2.23. The doubling map $\times_2$ on the circle $\mathbb{T}$ is mixing. By Example 2.11 it is the factor of the full shift $\{0, 1\}^{\mathbb{N}}$, and full shifts are mixing by Example 2.21. □

Minimality

A dynamical system (X, f) is **minimal** if it has no proper subsystems. Because the orbit closure $\overline{\mathcal{O}(x)}$ of every $x \in X$ is a subsystem, in minimal systems all orbits must be dense.

Proposition 2.24 *Dynamical system (X, f) over M is minimal if and only if all $x \in X$ are transitive points.*

Proof. If (X, f) is minimal then the subsystems $\overline{\mathcal{O}(x)}$ must be equal to X, that is, points x are transitive. Conversely, if (X, f) is not minimal then it has a proper subsystem $Y \subsetneq X$. But for $x \in Y$ we have $\overline{\mathcal{O}(x)} \subseteq Y$ so x is not transitive. □

A minimal system is always transitive. Indeed, if non-empty U and V are open then by minimality the orbit of every point in U intersects with V. So $f^m(U) \cap V \neq \emptyset$ for some $m \in M$, confirming transitivity.

However, a minimal system over $M = \mathbb{N}$ does not need to be mixing:

Example 2.25. Consider again the circle rotation $(\mathbb{T}, \rho_\alpha)$ by irrational α. By Example 2.16, all $x \in \mathbb{T}$ are transitive points. By Proposition 2.24 the system is minimal. However, by Example 2.20 it is not mixing. □

A factor of a minimal system is also minimal: If $h : X \longrightarrow Y$ is a factor map then every point $y \in Y$ has a pre-image $x \in X$, and for every non-empty open $U \subseteq Y$ the set

$h^{-1}(U) \subseteq X$ is non-empty and open. By minimality the orbit of x intersects $h^{-1}(U)$, so it follows that the orbit of y intersects U.

Next we show that every system has a minimal subsystem. This can be proved directly using Zorn's lemma. We present an elementary topological proof.

Theorem 2.26 *Every dynamical system (X, f) over monoid M contains a minimal subsystem.*

Proof. By Proposition 1.20 the topology has a countable base $B_1, B_2, \ldots$. For every $i = 1, 2, \ldots$, let

$$U_i = \bigcup_{m \in M} (f^m)^{-1}(B_i) = \{x \in X \mid \mathcal{O}(x) \cap B_i \neq \emptyset\}$$

be the set of points that can be mapped to B_i by the system. Set U_i is open as a union of open sets, so its complement $F_i = X \setminus U_i$ is closed. Moreover, F_i is invariant so it is a subsystem (or the empty set). Set F_i contains all those points whose orbits do not visit B_i.

Inductively we construct a sequence $X_0 \supseteq X_1 \supseteq X_2 \supseteq \ldots$ of subsystems of X as follows. It starts with $X_0 = X$. Then suppose that X_{i-1} has been defined. If $X_{i-1} \cap F_i = \emptyset$ then $X_i = X_{i-1}$, else $X_i = X_{i-1} \cap F_i$. Then X_i is a subsystem. Let

$$Y = \bigcap_{i=0}^{\infty} X_i.$$

Then Y is closed and invariant, and it follows from the compactness that $Y \neq \emptyset$. So Y is a subsystem.

Let us show that Y is minimal. Suppose on the contrary that there exist $y \in Y$ that is not transitive in Y, that is, that for some base set B_i holds $Y \cap B_i \neq \emptyset$ but $\mathcal{O}(y) \cap B_i = \emptyset$. The latter condition implies that $y \in F_i$. Now, $Y \cap F_i \neq \emptyset$ because of point y, so that $X_{i-1} \cap F_i \neq \emptyset$ and the construction assigns $X_i = X_{i-1} \cap F_i$. This means that $Y \subseteq F_i$, which contradicts $Y \cap B_i \neq \emptyset$. □

2.3 Recurrence

A periodic point $x \in X$ returns exactly back to its initial value regularly under the iteration of f. This is a very strong form of repetition. One can define a number of weaker repetition properties. The most important of them, uniformly recurrence, is closely related to minimality, and we define it for dynamical systems over arbitrary monoid M. Concepts of recurrence and quasi-periodicity we only define for systems over $\mathbb{N}$.

Recurrent points

Let (X, f) be a dynamical system over monoid $\mathbb{N}$. We say $x \in X$ is **recurrent** if it returns back to its every open neighborhood:

$$\forall_{\text{open}} U \ni x, \exists k > 0 \ : \ f^k(x) \in U.$$

In fact, a recurrent point returns to its open neighborhoods infinitely many times:

Lemma 2.27 *Point $x \in X$ is recurrent if and only if its forward trajectory has a subsequence that converges to x. In particular, a recurrent point returns to each open neighborhood of x infinitely many times.*

Proof. Let $x \in X$ be recurrent. If x is periodic then $x, x, \ldots$ is a subsequence of the trajectory. If x is not periodic we pick for every $n > 0$ the smallest $k_n > 0$ such that $f^{k_n}(x) \in B_{1/n}(x)$. Then $k_1 \leq k_2 \leq \ldots$. The set $\{k_n \mid n > 0\}$ is infinite because $f^{k_n}(x) \neq x$ for all n so that $f^{k_n}(x) \notin B_{1/m}(x)$ for all large enough m. It follows that there are $n_1 < n_2 < \ldots$ such that $k_{n_1} < k_{n_2} < \ldots$, so the subsequence $f^{k_{n_1}}(x), f^{k_{n_2}}(x), \ldots$ of the trajectory converges to x. The reverse implication is trivial, and also the last claim follows directly. □

Uniformly recurrent points

Let (X, f) be a dynamical system over an arbitrary monoid M. Point $x \in X$ is **uniformly recurrent** (or **almost periodic**) if it returns back to every open neighborhood U of x frequently in the sense that at any time the system can return to U within bounded time. More precisely, for every open neighborhood U of x there exists finite $R \subseteq M$ such that from every point of the orbit of x one can get to U by one of the transformations f^r, $r \in R$:

$$\forall_{\text{open}} U \ni x, \exists_{\text{finite}} R \subseteq M, \forall k \in M, \exists r \in R \ : \ (f^r \circ f^k)(x) \in U.$$

Over $M = \mathbb{N}$, the difference between recurrent and uniform recurrent points is that the time gaps between consecutive returns to an open neighborhood are bounded if the point is uniformly recurrent. The bound may be different for different neighborhoods. Also recurrent points return to open neighborhoods infinitely many times but the time gaps between consecutive returns do not need to be bounded.

Example 2.28. Consider the one-sided full shift $(A^{\mathbb{N}}, \sigma)$ for $A = \{0, 1\}$. If $x \in A^{\mathbb{N}}$ is any sequence that contains as finite subwords all $u \in A^*$ then x is recurrent but not uniformly recurrent. Recurrence follows from the fact that every finite word necessarily appears infinitely many times in x. But the return times to the cylinder $[x_0]$ with domain $\{0\}$ are not bounded because x contains as subword 0^n and 1^n with arbitrary large n. □

Uniformly recurrent points have the important property that all points of minimal dynamical systems are uniformly recurrent. Minimal dynamical systems are precisely the orbit closures of uniformly recurrent points.

Theorem 2.29 *The orbit closure $\overline{\mathcal{O}(x)}$ is minimal if and only if x is uniformly recurrent.*

Proof. Assume $\overline{\mathcal{O}(x)}$ is minimal. Let U be an arbitrary open neighborhood of x. All points of a minimal system are transitive (Proposition 2.24) so for every $y \in \overline{\mathcal{O}(x)}$ there exists $m \in M$ such that $f^m(y) \in U$. This means that the open sets $(f^m)^{-1}(U)$ over $m \in M$ cover $\overline{\mathcal{O}(x)}$. By compactness there is a finite subcover, say $(f^r)^{-1}(U)$ for $r \in R$. But this means that for all $y \in \mathcal{O}(x)$ there exists $r \in R$ such that $f^r(y) \in U$, confirming that x is uniformly recurrent.

Conversely, assume that x is uniformly recurrent. Let $y \in \overline{\mathcal{O}(x)}$ be arbitrary. Let us show that $x \in \overline{\mathcal{O}(y)}$. It then follows that $\overline{\mathcal{O}(x)} = \overline{\mathcal{O}(y)}$ so that y is transitive in $\overline{\mathcal{O}(x)}$. This is enough to show that $\overline{\mathcal{O}(x)}$ is minimal.

So let U be an arbitrary open neighborhood of x. Because we are in a metric space, there is an open neighborhood V of x such that $V \subseteq \overline{V} \subseteq U$. (To see this, note that an ε-ball around x is contained in U and we can take as V the $\frac{1}{2}\varepsilon$-ball around x.) Because x is uniformly recurrent, there is finite $R \subseteq M$ such that

$$\mathcal{O}(x) \subseteq \bigcup_{r \in R} (f^r)^{-1}(V).$$

By taking the closure,

$$\overline{\mathcal{O}(x)} \subseteq \bigcup_{r \in R} (f^r)^{-1}(\overline{V}) \subseteq \bigcup_{r \in R} (f^r)^{-1}(U).$$

In particular, for $y \in \overline{\mathcal{O}(x)}$ we have that $f^r(y) \in U$ for some $r \in R$. Because U was arbitrary, it follows that $x \in \overline{\mathcal{O}(y)}$. □

Corollary 2.30 *Minimal systems are orbit closures of uniformly recurrent points, and all points of a minimal system are uniformly recurrent. Every dynamical system contains a uniformly recurrent point.* □

Example 2.31. It was shown in Example 2.25 that the circle rotation $(\mathbb{T}, \rho_\alpha)$ by irrational α is minimal. It follows that every $x \in \mathbb{T}$ is uniformly recurrent. By Example 2.2 the points $x \in \mathbb{T}$ are not periodic. □

Quasi-periodic points

Let (X, f) be a dynamical system over monoid $M = \mathbb{N}$ or $M = \mathbb{Z}$. A **quasi-periodic** x returns back to each neighborhood periodically:

$$\forall_{\text{open}} U \ni x, \exists p, \forall i \in M \ : \ f^{ip}(x) \in U.$$

We clearly have the following implications:

$$x \text{ periodic } \Longrightarrow x \text{ quasi-periodic } \Longrightarrow x \text{ uniformly recurrent } \Longrightarrow x \text{ recurrent}$$

Example 2.32. By Example 2.31 all $x \in \mathbb{T}$ are non-periodic but uniformly recurrent under the rotation $f = \rho_\alpha$ by an irrational α. There are no quasi-periodic points because, for every period p, the set $\{x, f^p(x), f^{2p}(x), \dots\}$ is the orbit of x under the irrational rotation by $p\alpha$, and hence dense. □

Example 2.33. Quasi-periodic elements of $A^{\mathbb{N}}$ and $A^{\mathbb{Z}}$ are called **Toeplitz-sequences**. For $M = \mathbb{N}$ or $M = \mathbb{Z}$, a configuration $x \in A^M$ is a Toeplitz-sequence if for every $m \in M$ there exists $p > 0$ such that $x_m = x_{m+ip}$ for all $i \in M$. In other words, every symbol belongs to an infinite arithmetic progression of identical symbols. Any quasi-periodic point must have this property, as seen by considering cylinders with singleton domain in the definition of quasi-periodicity. Conversely, if x is Toeplitz then every finite pattern in x must be part of an arithmetic progression of identical patterns, where as the period one can take the least common multiple of the periods for the individual symbols of the pattern, guaranteed by the Toeplitz-property. Hence the quasi-periodicity follows as cylinders are a base of the topology.

For an example of a non-periodic Toeplitz-sequence, consider $A = \{0, 1\}$ and the sequence $x \in A^{\mathbb{N}}$ where $x_k = 1$ if and only if $k + 1 = m2^n$ for some odd integers n, m. The sequence starts $010001010100\dots$. Every positive integer $k+1$ can be written uniquely as $k+1 = m2^n$ for an odd m. With period $p = 2^{n+1}$ we then have that $x_k = x_{k+ip}$ for all $i \in \mathbb{N}$, because $k + ip + 1 = m2^n + i2^{n+1} = (m + 2i)2^n$ and $m + 2i$ is odd. The sequence is clearly not periodic: For every p we have that $x_{p-1} \neq x_{2p-1}$. □

2.4 Sensitivity to initial conditions

Some dynamical systems have the property that small changes in the initial state may be amplified over time ("butterfly effect"). Also here we have several variants of the phenomenon depending on the strength of the amplification. At the opposite end of the spectrum we have equicontinuous systems that are stable under perturbations.

Equicontinuity

Recall that a family S of functions $X \longrightarrow Y$ between metric spaces is called equicontinuous at $x \in X$ if all $f \in S$ are continuous at x using a common parameter value: For every $\varepsilon > 0$ there exists $\delta > 0$ such that for all $f \in S$ holds $f(B_\delta(x)) \subseteq B_\varepsilon(f(x))$.

Let (X, f) be a dynamical system over monoid M. The system is **equicontinuous at** $x \in X$ if the family $\{f^m \mid m \in M\}$ is equicontinuous at x. In this case we say that x is an **equicontinuity point** of the system. Stated briefly, the system is equicontinuous at x iff

$$\forall \varepsilon > 0, \exists \delta > 0, \forall m \in M, \forall y \in B_\delta(x) \;:\; d(f^m(y), f^m(x)) < \varepsilon.$$

We say that a system is **equicontinuous** if it is equicontinuous at all $x \in X$, and we call it **almost equicontinuous** if the set of equicontinuity points is a residual set.

Example 2.34. The circle rotation $(\mathbb{T}, \rho_\alpha)$ is equicontinuous. In fact, any isometry is equicontinuous by definition. $\square$

Proposition 2.35 *In a transitive dynamical system (X, f) over an arbitrary monoid M, every equicontinuity point is transitive.*

Proof. Let $x \in X$ be an equicontinuity point, and let $z \in X$ and $\varepsilon > 0$ be arbitrary. By the equicontinuity of x there exists an open neighborhood U of x such that $d(f^m(y), f^m(x)) < \varepsilon/2$ for all $y \in U$ and all $m \in M$. By transitivity there is $y \in U$ and $m \in M$ such that $d(f^m(y), z) < \varepsilon/2$. Using the triangle inequality we get that $d(f^m(x), z) \leq d(f^m(y), f^m(x)) + d(f^m(y), z) < \varepsilon$. We conclude that $\mathcal{O}(x)$ is dense.

$\square$

Example 2.36. In the circle rotation $(\mathbb{T}, \rho_\alpha)$ all $x \in \mathbb{T}$ are equicontinuity points. If α is irrational then the system is transitive, and all points are transitive confirming the proposition above. If α is rational then no point is transitive: this shows that the assumption about transitivity of the system is necessary in the proposition. $\square$

Sensitivity

Dynamical system (X, f) over monoid M is called **sensitive** if there exists $\varepsilon > 0$, the **sensitivity constant**, such that arbitrarily close to any point there is another point and a time when the trajectories of the two points deviate by at least ε:

$$\exists \varepsilon > 0, \forall x \in X, \forall \delta > 0, \exists m \in M, \exists y \in B_\delta(x) \ : \ d(f^m(y), f^m(x)) \geq \varepsilon.$$

A sensitive system has no equicontinuity points. The converse is not true in general, but if a system over a group or over the monoid $\mathbb{N}$ is transitive then we have the following dichotomy.

Proposition 2.37 *Let M be a group or $M = \mathbb{N}$. Let (X, f) be a transitive dynamical system over M and let $E \subseteq X$ be the set of its equicontinuity points. Then exactly one of the two following conditions holds:*

(i) The system is sensitive and $E = \emptyset$.

(ii) The system is not sensitive and E is a residual set (and hence non-empty).

Proof. For any $\varepsilon > 0$ we define the set

$$E_\varepsilon = \{x \in X \mid \exists \delta > 0, \forall m \in M \ : \ f^m(B_\delta(x)) \subseteq B_\varepsilon(f^m(x))\}$$

of points that satisfy the equicontinuity condition for ε. This means that E is the intersection of E_ε over all $\varepsilon > 0$. Note also that (X, f) is sensitive with sensitivity constant ε if and only if $E_\varepsilon = \emptyset$.

1° Assume that some E_ε is empty. Then $E = \emptyset$ and the system (X, f) is sensitive with sensitivity constant ε, so condition (i) holds.

2° Assume then that $E_\varepsilon \neq \emptyset$ for all $\varepsilon > 0$. Then the system does not have a sensitivity constant and it is hence not sensitive. It remains to prove that E is a residual set, that is, a countable intersection of dense open sets.

We start by proving backward invariance: that $(f^k)^{-1}(E_\varepsilon) \subseteq E_\varepsilon$, for all $k \in M$. This is the only part of the proof where we use the fact that M is a group or $M = \mathbb{N}$. So let $f^k(x) \in E_\varepsilon$. Then there exists $\delta > 0$ such that $f^m(B_\delta(f^k(x))) \subseteq B_\varepsilon(f^m(f^k(x)))$ for all $m \in M$. Because of continuity of f^k, there exists $\delta' > 0$ such that $f^k(B_{\delta'}(x)) \subseteq B_\delta(f^k(x))$, so that

$$f^{km}(B_{\delta'}(x)) \subseteq f^m(B_\delta(f^k(x))) \subseteq B_\varepsilon(f^{km}(x))$$

holds for all $m \in M$. This proves that $f^n(B_{\delta'}(x)) \subseteq B_\varepsilon(f^n(x))$ for all $n \in kM$. If M is a group then $kM = M$ so we have $x \in E_\varepsilon$. If $M = \mathbb{N}$ then kM contains all $n \in \mathbb{N}$ except $0, 1, \ldots, k-1$. But by using continuity of f^n, choosing δ' sufficiently small guarantees $f^n(B_{\delta'}(x)) \subseteq B_\varepsilon(f^n(x))$ also for these finitely many values of n. Hence $x \in E_\varepsilon$ also in this case.

Let us next prove that there is an open set U_ε between E_ε and $E_{2\varepsilon}$. Consider any $x \in E_\varepsilon$. When $f^m(B_\delta(x)) \subseteq B_\varepsilon(f^m(x))$ then for any $y \in B_{\delta/2}(x)$ we have that

$$f^m(B_{\delta/2}(y)) \subseteq f^m(B_\delta(x)) \subseteq B_\varepsilon(f^m(x)) \subseteq B_{2\varepsilon}(f^m(y)).$$

This means that every $x \in E_\varepsilon$ has an open neighborhood $B_{\delta/2}(x)$ that is contained in $E_{2\varepsilon}$. The union of these open neighborhoods over all $x \in E_\varepsilon$ is an open set U_ε that satisfies

$$E_\varepsilon \subseteq U_\varepsilon \subseteq E_{2\varepsilon},$$

as required.

It is now clear that E is the countable intersection of open sets $U_{1/n}$, over $n = 1, 2, \ldots$. It suffices to show that each U_ε is dense. So let $U \neq \emptyset$ be an open set. By transitivity, there is $m \in M$ such that $U \cap (f^m)^{-1}(U_{\varepsilon/2}) \neq \emptyset$. On the other hand,

$$(f^m)^{-1}(U_{\varepsilon/2}) \subseteq (f^m)^{-1}(E_\varepsilon) \subseteq E_\varepsilon \subseteq U_\varepsilon$$

so that $U \cap U_\varepsilon \neq \emptyset$. In other words U_ε is dense. □

Expansivity

A very strong form of sensitivity to initial conditions is expansivity. It requires that there exists $\varepsilon > 0$, the **expansivity constant**, such that the trajectories of any distinct points eventually deviate by at least ε:

$$\exists \varepsilon > 0, \forall x, y \in X \;:\; x \neq y \Longrightarrow (\exists m \in M \;:\; d(f^m(y), f^m(x)) \geq \varepsilon).$$

If space X has no isolated points then expansivity implies sensitivity, with the same constant ε. Indeed, every open neighborhood of every point x contains another point $y \neq x$, which by expansivity deviates from the trajectory of x under the dynamics.

Example 2.38. Consider the full shift (A^M, σ) over monoid M. If $x, y \in A^M$ and $x \neq y$ then $x_m \neq y_m$ for some $m \in M$. But then $\sigma^m(x)$ and $\sigma^m(y)$ differ at cell 1_M. This means the system is expansive, where we can choose expansivity constant $\varepsilon > 0$ so that any two configurations differing at cell 1_M have distance at least ε.

Clearly any subsystem of an expansive system is also expansive. We see that all subsystems of full shifts (called subshifts) are also expansive. These are the main systems studied by symbolic dynamics so that we always have this strong form of sensitivity. □

2.5 Chaos

We have learned about three types of properties of dynamical systems: mixing properties (transitivity, mixingness, minimality, transitive points), regularity properties (periodicity, quasi-periodicity, recurrence, uniform recurrence), and stability vs. sensitivity properties (equicontinuity, sensitivity, equicontinuity points, expansivity).

In a chaotic dynamical system there is regular and transitive behavior densely everywhere in the phase space, with sensitivity so that small changes in the initial state may change regular behavior into transitive behavior or vice versa. Different precise definitions of chaos have been proposed. One of the classical definitions is the following. A dynamical system (X, f) over the monoid $\mathbb{N}$ is called **Devaney chaotic** if

1. it is transitive,
2. it is sensitive, and
3. periodic points are dense in X.

It turns out that if X is an infinite set then the second condition is implied by the other two: an infinite transitive system with a dense set of periodic points is automatically sensitive to initial conditions. We obtain this as a corollary to the following result that is stated for systems over arbitrary monoids:

Lemma 2.39 *Suppose that a dynamical system (X, f) over a monoid M satisfies the following properties:*

(a) The system is transitive.

(b) There exist orbits with a positive distance from each other:

$$\exists a > 0 \ \exists u, v \in X \ \forall n, k \in M \ : \ d(f^n(u), f^k(v)) > a.$$

(c) The set of uniformly recurrent points is dense in X.

Then the system is sensitive.

Proof. Let $a > 0$ and $u, v \in X$ be as in (b). To prove that $\varepsilon = a/8$ works as a sensitivity constant, let $x \in X$ and let $\delta > 0$ be arbitrary, as in the definition of sensitivity. We may assume, without loss of generality, that $\delta < a/8$.

By (b) and the triangle equality, either $d(x, f^n(u)) > a/2$ for all $n \in M$ or $d(x, f^k(v)) > a/2$ for all $k \in M$. By symmetry we may assume the former. Let $U = B_\delta(x)$ denote the open ball of radius δ around x. By the condition (c) there exists a uniformly recurrent $y \in U$. So for some finite $R \subset M$ we have that whenever $k \in M$ there exists $r \in R$ such that $(f^r \circ f^k)(y) \in U$. The set

$$V = \bigcap_{r \in R} (f^r)^{-1}(B_\delta(f^r(u)))$$

is open as a finite intersection of open sets. It is also non-empty because $u \in V$. By transitivity of the system, there exists $z \in U$ and $k \in M$ such that $f^k(z) \in V$. As noted above, there exists $r \in R$ such that $(f^r \circ f^k)(y) \in U$. On the other hand, because $f^k(z) \in V$ and $r \in R$, we have that $(f^r \circ f^k)(z) \in B_\delta(f^r(u))$. Thus, by the triangle inequality,

$$\begin{array}{rcl} d((f^r \circ f^k)(y), (f^r \circ f^k)(z)) & \geq & d(x, f^r(u)) - d(x, (f^r \circ f^k)(y)) - d((f^r \circ f^k)(z), f^r(u)) \\ & > & a/2 - \delta - \delta \\ & > & a/2 - a/8 - a/8 \\ & = & a/4. \end{array}$$

Hence, either $d((f^r \circ f^k)(y), (f^r \circ f^k)(x)) > a/8$ or $d((f^r \circ f^k)(z), (f^r \circ f^k)(x)) > a/8$. Thus either y or z confirms sensitivity. □

Corollary 2.40 *Let (X, f) be a dynamical system over the monoid $\mathbb{N}$. Assume that X is infinite, the system is transitive, and the periodic points are dense. Then the system is sensitive.*

Proof. Let u be one of the temporally periodic points. Its orbit is finite and therefore the complement of the orbit is open. Because X is infinite the complement of the orbit is also non-empty. By the density of the periodic points there exists a periodic point v in the complement. The finite, periodic orbits of u and v are disjoint and therefore they have a positive distance from each other. Thus the condition (b) of the previous lemma is satisfied. Because periodic points are automatically uniformly recurrent, also (c) is satisfied. It follows from the lemma that the system is sensitive. □

Example 2.41. The doubling map $\times_2 : \mathbb{T} \longrightarrow \mathbb{T}$ of Example 2.3 is Devaney chaotic. Indeed, as noted in Example 2.3, rational numbers with odd denominators are periodic, so the periodic points are densely in $\mathbb{T}$. By Example 2.23 the doubling map is mixing and hence transitive. Sensitivity now follows from the corollary above. □

3 Symbolic dynamics

In symbolic dynamics the objects of study are the full shift dynamical systems (A^M, σ) over monoids M, introduced in Example 2.14, their subsystems and their endomorphisms. Subsystems are called **subshifts**. By the definition of a subsystem, a subshift is a non-empty topologically closed subset $S \subseteq A^M$ such that $\sigma^m(S) \subseteq S$ for all $m \in M$. If M is a group then from $\sigma^{m^{-1}}(S) \subseteq S$ we get that $S = \sigma^m(\sigma^{m^{-1}}(S)) \subseteq \sigma^m(S)$, so that for a subshift S over a group we have that $\sigma^m(S) = S$ for all $m \in M$.

In Section 3.1 we give basic definitions that apply to all monoids M. After that, in the rest of these notes, the focus is on the one-dimensional cases $M = \mathbb{Z}$ and $M = \mathbb{N}$. These we first saw in Examples 2.5 and 2.6. Their subsystems are two-sided and one-sided subshifts, respectively.

3.1 Subshifts: basic definitions

A subshift S is fully characterized by the finite patterns that occur in its configurations. Recall that a finite pattern $w \in A^D$ over alphabet A and monoid M assigns letters to cells in a finite domain $D \subseteq M$, and that the corresponding cylinder $[w]$ contains all configurations $x \in A^M$ such that $x_{|D} = w$.

We define

$$L_S = \{w \in A^D \mid D \subseteq M \text{ finite}, [w] \cap S \neq \emptyset\}$$

and call L_S the **language** of S. It consists of all those finite patterns that occur in some element of S.

Proposition 3.1 *If $S \neq T$ are distinct subshifts then $L_S \neq L_T$.*

Proof. This follows directly from the fact that for any subshift $S \subseteq A^M$,

$$A^M \setminus S = \bigcup_{w \notin L_S} [w].$$

Indeed, using the openness of $A^M \setminus S$ and the fact that cylinders are a base, we have that $x \notin S$ if and only if $x \in [w]$ for some $w \notin L_S$. □

Note that there are sets of finite patterns that are not a language of any subshift. Another, concrete way to describe subshifts is in terms of **forbidden patterns**. We say that configuration $x \in A^M$ **avoids** pattern w if $\sigma^m(x) \notin [w]$ holds for all $m \in M$, that is, no translate of x has pattern w at D.

Let P be a set of finite patterns. The set of configurations that avoid all patterns in P is denoted by

$$X_P = \{x \in A^M \mid \forall m \in M, \forall w \in P \;:\; \sigma^m(x) \notin [w]\ \}.$$

It is shown in the following proposition that, unless empty, set X_P is a subshift, and that every subshift arises this way. Note that every set P then defines some subshift X_P, but there may be several different sets that define the same subshift.

Proposition 3.2 *Set $S \subseteq A^M$ is a subshift (or the empty set) if and only if $S = X_P$ for a set P of finite patterns.*

Proof. The complement of X_P is

$$A^M \setminus X_P = \bigcup_{m\in M} \bigcup_{w\in P} (\sigma^m)^{-1}([w]),$$

so it is open as a union of open sets $(\sigma^m)^{-1}([w])$. We conclude that X_P is topologically closed. Let us prove that $\sigma^m(X_P) \subseteq X_P$: If x is such that $\sigma^m(x) \notin X_P$ for some $m \in M$, then there are $k \in M$ and $w \in P$ such that $\sigma^k(\sigma^m(x)) \in [w]$. But then $\sigma^{mk}(x) \in [w]$, proving that $x \notin X_P$. We conclude that $\sigma^m(X_P) \subseteq X_P$, and therefore X_P is a subshift (or the empty set).

Conversely, let S be any subshift (or the empty set). Let P be the complement of L_S, so that $w \in P$ if and only if $[w] \cap S = \emptyset$. Let us show that $S = X_P$:

"$\subseteq$" If $x \notin X_P$ then $\sigma^m(x) \in [w]$ for some $m \in M$ and $w \in P$. But if $x \in S$ then also $\sigma^m(x) \in S$. In this case $[w] \cap S \neq \emptyset$, a contradiction with $w \in P$.

"$\supseteq$" If $x \notin S$ then, by the openness of $A^M \setminus S$, there exists a cylinder $[w]$ such that $x \in [w]$ and $[w] \cap S = \emptyset$. By the definition of P we have $w \in P$. We conclude that x contains a forbidden pattern w, hence $w \notin X_P$.

□

Corollary 3.3 *For every subshift S we have $S = X_P$ where P is the complement of L_S.*

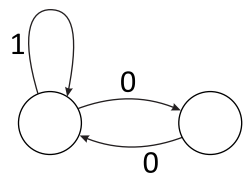
We mostly use forbidden patterns to define subshifts. This formalism also provides a fruitful approach to classifying subshifts.

Subshifts of finite type

The simplest subshifts are those defined by forbidding a finite number of patterns: Subshift S is called a **subshift of finite type (SFT)** if $S = X_P$ for a finite set P. Note that the same subshift may be defined using different sets of forbidden patterns. So if a subshift is defined by an infinite set of forbidden patterns this, naturally, does not imply that the subshift would not be of finite type: there maybe another finite set that defines it.

Example 3.4. Let $M = \mathbb{Z}$ and $A = \{0, 1\}$. Consider the SFT S_{gm} defined by forbidding word 11, that is, $S = X_P$ for $P = \{11\}$. This is the **golden mean** shift. □

Example 3.5. Consider the subshift $S_{even} \subseteq \{0, 1\}^{\mathbb{Z}}$ over $M = \mathbb{Z}$, defined by forbidding words $10^{2n+1}1$ for all $n \in \mathbb{N}$. This is called the **even shift** because any two consecutive occurrences of 1's must have an even number of 0's between them. The subshift is defined by forbidding a regular language. Then also the language of the subshifts – in this case the set of finite words that appear in the subshift – is regular, recognized by the finite automaton



where all states are considered initial and final. Such subshifts obtained by forbidding a regular language are called **sofic**. These will be discussed later in Section 3.2. The even shift is not of finite type: suppose $S_{even} = X_P$ for a finite set P of finite patterns. Let n be such that the domains of all patterns in P fit in $D = \{-n, \ldots, n-1\}$. Configurations $x_1 = ^\infty 10^{2n}1^\infty$ and $x_2 = ^\infty 10^{2n+1}1^\infty$ contain identical subwords of length $2n$. Because $x_2 \notin S_{even}$ there is an occurrence of a forbidden pattern in x_2, which means that there is the same forbidden pattern in x_1. But this contradicts the fact that $x_1 \in S_{even}$. □

Example 3.6. Once again, let $M = \mathbb{Z}$ and $A = \{0, 1\}$. Now we forbid all patterns that contain both symbols 0 and 1. This only allows two configurations $^\infty 0^\infty$ and $^\infty 1^\infty$. Although we forbid infinitely many patterns this subshift is of finite type: the same subshift is defined by forbidding two patterns 01 and 10.

□

Homomorphisms

Let A and B be alphabets, and let $S \subseteq A^M$ be a subshift. Let $D \subseteq M$ be finite (the neighborhood) and let $\phi : (L_S \cap A^D) \longrightarrow B$ be a function that assigns a symbol of B to each D-pattern that appears in some element of S. The local rule ϕ determines the **block map** $f : S \longrightarrow B^M$ by $f(x)_m = \phi(\sigma^m(x)_{|D})$ for all $x \in S$ and all $m \in M$. Function f applies ϕ at each cell on the D-pattern around the cell.

Note the similarity of the definition to cellular automata, discussed in Example 2.7. Analogously to Example 2.7, it is easy to see that f is a homomorphism $(S, \sigma) \longrightarrow (B^M, \sigma)$. To see continuity, consider any converging sequence $x^{(1)}, x^{(2)}, \ldots$ of configurations, with limit $\lim_{i\to\infty} x^{(i)} = x$. By convergence, for every $m \in M$ and all sufficiently large n we have $x^{(n)}_{|mD} = x_{|mD}$, so that also $f(x^{(n)})_m = f(x)_m$. This means that the sequence $f(x^{(1)}), f(x^{(2)}), \ldots$ converges to $f(x)$, and by Proposition 1.25 function f is continuous. To see commutation with σ^m for all $m \in M$, one calculates for any $x \in A^M$ and $k \in M$ that

$$f(\sigma^m(x))_k = \phi(\sigma^k(\sigma^m(x))_{|D}) = \phi(\sigma^{mk}(x)_{|D}) = f(x)_{mk} = \sigma^m(f(x))_k,$$

so that $f \circ \sigma^m = \sigma^m \circ f$.

In fact, block maps are precisely the homomorphisms between subshifts:

Theorem 3.7 (Curtis-Hedlund-Lyndon) *Let $S \subseteq A^M$ be a subshift. A function $f : S \longrightarrow B^M$ is a homomorphism from (S, σ) to (B^M, σ) if and only if it is a block map.*

Proof. We saw above that block maps are homomorphisms so it remains to show that every homomorphism is a block map. Let $f : S \longrightarrow B^M$ be any homomorphism. It is uniformly continuous as a continuous function between compact metric spaces (homework). It follows that there is a finite $D \subseteq M$ such that $x_{|D} = y_{|D} \Longrightarrow f(x)_{1_M} = f(y)_{1_M}$ holds for $x, y \in S$. Define $\phi : (L_S \cap A^D) \longrightarrow B$ so that $\phi(x_{|D}) = f(x)_{1_M}$ for all $x \in S$. By commutation of f with σ we then have for all $x \in S$ and all $m \in M$ that

$$f(x)_m = \sigma^m(f(x))_{1_M} = f(\sigma^m(x))_{1_M} = \phi(\sigma^m(x)_{|D}),$$

which confirms that f is the block map determined by ϕ.

□

Example 3.8. Let S_{gm} be the golden mean subshift of Example 3.4. Let $D = \{0,1\} \subseteq \mathbb{Z}$, and define local function $\phi : (L_{S_{gm}} \cap \{0,1\}^D) \longrightarrow \{0,1\}$ by

$$\begin{aligned} 00 &\mapsto 1, \\ 01 &\mapsto 0, \\ 10 &\mapsto 0. \end{aligned}$$

Let $f : S_{gm} \longrightarrow \{0,1\}^{\mathbb{Z}}$ be the corresponding block map. Then in $f(S_{gm})$ consecutive 1's must be separated by an even number of 0's. Indeed, a pre-image of any word in 10^*1 must have form $00(10)^n0$ for some n, and this word is mapped to $10^{2n}1$. So $f(S_{gm})$ is a subset of S_{even}, the even subshift of Example 3.5. But in fact every $x \in S_{even}$ has a pre-image so that f is a factor map from the golden mean subshift onto the even subshift. □

Example 3.9. By Theorem 3.7, the endomorphisms of a full shift are precisely its cellular automata, and its automorphisms are precisely the reversible cellular automata. This holds over any monoid M. More generally, the endomorphisms of a subshift S are called **cellular automata on subshift** S. By Theorem 3.7 they are all block maps. □

Now we can prove that a subshift that is conjugate to an SFT is also an SFT. We first collect in a lemma simple properties that we combine to obtain this result.

Lemma 3.10 *Following three properties hold:*

(a) Let $f : A^M \longrightarrow A^M$ be an endomorphism of a full shift. Its fixed point set $\{x \in A^M \mid f(x) = x\}$ is an SFT (or empty).

(b) Let $f : A^M \longrightarrow B^M$ be a homomorphism between full shifts, and let $S \subseteq B^M$ be an SFT. Then $f^{-1}(S)$ is an SFT (or empty).

(c) A finite intersection of SFTs is an SFT (or empty).

Proof. (a) By Theorem 3.7, homomorphism $f : A^M \longrightarrow A^M$ is a block map, determined by a finite neighborhood $D \subseteq M$ and a local rule $\phi : A^D \longrightarrow A$. Without loss of generality we may assume that $1_M \in D$. Define $P = \{w \in A^D \mid \phi(w) \neq w_{1_M}\}$. Then $f(x) = x$ if and only if $x \in X_P$:

$$\begin{aligned} x \notin X_P &\iff \exists m \in M \ : \ \sigma^m(x)_{|D} \in P \\ &\iff \exists m \in M \ : \ \phi(\sigma^m(x)_{|D}) \neq \sigma^m(x)_{1_M} \\ &\iff \exists m \in M \ : \ f(x)_m \neq x_m \\ &\iff f(x) \neq x. \end{aligned}$$

Because P is finite, the fixed point set X_P is an SFT or empty.

(b) By Theorem 3.7, homomorphism $f : A^M \longrightarrow B^M$ is a block map, determined by a finite neighborhood $D \subseteq M$ and a local rule $\phi : A^D \longrightarrow B$. Because $S \subseteq B^M$ is an SFT, there is a finite set P of patterns such that $S = X_P$. Without loss of generality we may assume that all patterns in P have the same finite domain $E \subseteq M$. Let $F = ED$, and define $g : A^F \longrightarrow B^E$ in such a way that for all $x \in A^M$

$$g(x_{|F}) = f(x)_{|E}.$$

This can be done because $x_{|F} = y_{|F} \Longrightarrow f(x)_{|E} = f(y)_{|E}$.

Let $P' = \{w \in A^F \mid g(w) \in P\}$. Then $X_{P'} = f^{-1}(S)$: We have $x \notin X_{P'}$ if and only if for some $m \in M$ we have $g(\sigma^m(x)_{|F}) \in P$. And we have $f(x) \notin S$ if and only if for some $m \in M$ we have $\sigma^m(f(x))_{|E} \in P$. But these two conditions are equivalent because

$$\sigma^m(f(x))_{|E} = f(\sigma^m(x))_{|E} = g(\sigma^m(x)_{|F}).$$

(c) Clearly $X_P \cap X_Q = X_{P \cup Q}$. We see that a non-empty intersection of two SFTs is an SFT. □

Proposition 3.11 *Let $S \subseteq A^M$ and $T \subseteq B^M$ be conjugate subshifts. If S is of finite type so is T.*

Proof. Let $f : S \longrightarrow T$ be a conjugacy. By Theorem 3.7 it is a block map, defined by a finite neighborhood $D \subseteq M$ and a local rule $\phi : (L_S \cap A^D) \longrightarrow B$. We may, arbitrarily, extend ϕ to D-patterns outside of L_S. We get a block map $\hat{f} : A^M \longrightarrow B^M$ such that $\hat{f}_{|S} = f$.

Let $g : T \longrightarrow S$ be the inverse of f. It is also a conjugacy so, as above, there is a block map $\hat{g} : B^M \longrightarrow A^M$ such that $\hat{g}_{|T} = g$.

By Lemma 3.10(a), the set $F = \{x \in B^M \mid \hat{f}(\hat{g}(x)) = x\}$ of fixed points of $\hat{f} \circ \hat{g}$ is an SFT. By Lemma 3.10(b), the set $P = \hat{g}^{-1}(S) = \{x \in B^M \mid \hat{g}(x) \in S\}$ is an SFT. By Lemma 3.10(c), the intersection $F \cap P$ is an SFT. But $F \cap P = T$: If $x \in T$ then $\hat{g}(x) = g(x) \in S$ and $\hat{f}(\hat{g}(x)) = \hat{f}(g(x)) = f(g(x)) = x$. Conversely, if $x \in F \cap P$ then $\hat{g}(x) \in S$ so that $x = \hat{f}(\hat{g}(x)) = f(\hat{g}(x)) \in T$. □

Sofic shifts

Sofic shifts are factors of SFTs: A subshift $S \subseteq B^M$ is **sofic** if there exists a subshift of finite type $S_0 \subseteq A^M$ and a homomorphism $h : S_0 \longrightarrow B^M$ such that $S = h(S_0)$. Subshift S_0 is an **SFT cover** of S. We prove in Section 3.2 that in the one-dimensional case $M = \mathbb{Z}$ this definition coincides with the property of being determined by forbidding a regular language, as discussed in Example 3.5.

Intuitively, an SFT cover of a sofic shift provides "hidden information" that is erased by the factor map. Namely, consider $S' \subseteq A^M \times B^M$ such that $(x, y) \in S'$ if and only if $x \in S_0$ and $y = h(x)$. Set S' can be viewed as a subshift over alphabet $A \times B$ by identifying the elements of $(A \times B)^M$ and $A^M \times B^M$ in the natural way. Subshift S' is conjugate to S_0 by the map $x \mapsto (x, h(x))$. The projection $\pi : S' \longrightarrow S$ that erases the first track, defined by $(x, y) \mapsto y$, is a factor map. So S' is an SFT cover of S where the hidden information is the content of the first track.

Example 3.12. By Example 3.8 the even subshift is a factor of the golden mean SFT and hence it is sofic. By Example 3.5 it is defined by forbidding words of a regular language. The golden mean subshift is an SFT cover of the even shift.

Directly from the definition of sofic shifts we obtain the following:

Proposition 3.13 *Let $S \subseteq A^M$ be a sofic shift and $T \subseteq B^M$ a subshift. If T is a factor of S then T is sofic. In particular, if T is conjugate to S then T is sofic.*

Proof. Let S_0 be an SFT cover of S and $h : S_0 \longrightarrow S$ a factor map. If $f : S \longrightarrow T$ is a factor map then $f \circ h : S_0 \longrightarrow T$ is also a factor map, proving that T is sofic with SFT cover S_0. □

Effective subshifts

In the following we assume that M is a finitely generated monoid with decidable word problem. This means that there exists an algorithm to determine if two given products $g_1 \dots g_n$ and $g'_1 \dots g'_k$ of generators are equal. It is easy to see that decidability of the word problem is independent of the finite generator set used.

A finite pattern is given as a finite set of pairs (m, a) where $m \in M$ (given as a product of generators) and $a \in A$. The pattern assigns symbol a in cell m. Decidability of the word problem is used to make sure that a given finite pattern is consistent, that is, that it does not try to assign two different symbols in the same monoid element. One can then also algorithmically determine if two given finite patterns are identical. Moreover, there is a Gödel numbering of finite subsets of the monoid: There is a list $D_1, D_2 \dots$ of finite $D_i \subseteq M$ such that every finite subset of M appears exactly once in the list, and there are algorithms to produce D_i for any given number i, and to produce i such that $D_i = D$ for any given finite $D \subseteq M$. To get such an enumeration one can produce, in some fixed order, for every

$n = 1, 2, 3, \ldots$ the (finite) set of all finite subsets of M whose elements are a product of at most n generators, and add to the list only those sets that are not yet there.

A set P of finite patterns is called recursive if there is an algorithm to decide if a given pattern w is in P, and P is called recursively enumerable if there is a semi-algorithm to recognize elements of P. Equivalently to the semi-algorithm, elements of a recursively enumerable set are listed by an enumeration algorithm that produces for any given i the i'th pattern in the list.

A subshift X_P defined by a recursively enumerable set P of forbidden patterns is called **effective**. In fact, effective subshifts are equivalently defined by a recursive set of forbidden patterns, as shown by Proposition 3.15 below. Its proof is based on the following observation:

Lemma 3.14 *Fix a Gödel numbering $D_1, D_2 \ldots$ of all finite subsets of M. Let **Alg** be an algorithm that produces for any given positive integer n a finite pattern w_n with domain D_{i_n}. If $i_1 \leq i_2 \leq i_3 \leq \ldots$ then $P = \{w_n \mid n \in \mathbb{N}\}$ is recursive.*

Proof. Finite sets are recursive. Let us then assume that P is infinite. An algorithm to test if a given finite pattern w is in P works as follows: First compute i such that the domain of w is D_i. Using algorithm **Alg** enumerate $w_1, w_2, \ldots$ until

(a) we find $w_n = w$, or

(b) we find w_n whose domain D_{i_n} has $i_n > i$, i.e., whose domain comes later in the Gödel numbering than the domain of w.

In case (a) we have $w \in P$, and in case (b) we can conclude that $w \notin P$ because all patterns that come later in the list have domains D_j with $j > i$. Note that eventually either (a) or (b) must happen because P is infinite and therefore contains patterns with a domain D_j for some $j > i$. □

Let us call a list $w_1, w_2, \ldots$ of finite patterns monotonic if it has the property of the lemma: $i_1 \leq i_2 \leq i_3 \leq \ldots$ where D_{i_n} is the domain of pattern w_n.

Proposition 3.15 *Let M be a finitely generated monoid with decidable word problem, and let $S \subseteq A^M$ be a subshift. The following are equivalent:*

(a) $S = X_P$ for a recursively enumerable set P of finite patterns, that is, S is effective,

(b) $S = X_P$ for a recursive set P of finite patterns,

(c) The complement of L_S is recursively enumerable.

Proof. The implication $(b) \Longrightarrow (a)$ is trivial because every recursive set is recursively enumerable, and the implication $(c) \Longrightarrow (a)$ follows directly from Corollary 3.3.

Let us prove next that (a) implies (b). First we fix a Gödel numbering $D_1, D_2 \ldots$ of all finite subsets of the monoid. Let $S = X_P$ for a recursively enumerable set P of finite patterns, and let $w_1, w_2, \dots$ be a list of the elements of P produced by an enumeration algorithm **Alg**. The idea of the proof is to "pad" the patterns w_i into bigger domains to obtain a monotonic enumeration of bigger patterns $w'_1, w'_2, \dots$ such that $P' = \{w'_i \mid i = 1, 2, 3 \dots\}$ defines the same subshift $S = X_{P'}$ as P. The new list is produced by the following algorithm:

```
n ⟵ 0
for i=1,2,3,... do
begin
  w ⟵ w_i
  D ⟵ domain of w_i
  repeat
    n⟵ n+1
  until D ⊆ D_n
  output all e ∈ A^{D_n} such that e_{|D} = w
end
```

Let P' be the set of patterns produced by the algorithm. It is clear – because n never decreases – that patterns are produced in monotonic order with respect to Gödel numbering $D_1, D_2 \dots$. By Lemma 3.14 we have that P' is a recursive set. Moreover:

- For every $e \in P'$ with domain D', there exists $w \in P$ with domain $D \subseteq D'$ such that $w = e_{|D}$. It follows that $X_P \subseteq X_{P'}$.
- For every $w \in P$ with domain D, there exists domain D' such that $D \subseteq D'$ and all patterns $e \in A^{D'}$ with $e_{|D} = w$ are in P'. It follows that $X_{P'} \subseteq X_P$.

We conclude that $X_{P'} = X_P$.

Let us prove that (a) implies (c). Let $S = X_P$ for a recursively enumerable set P of finite patterns. For a finite pattern $w \in A^D$ and for $m \in M$, the translate $w^{(m)} \in A^{mD}$ of pattern w by m is defined by $(w^{(m)})_{mn} = w_n$ for all $n \in D$. Then $\sigma^m(x) \in [w]$ if and only if $x \in [w^{(m)}]$. The set $P' = \{w^{(m)} \mid w \in P, m \in M\}$ of all translates of patterns in P is also recursively enumerable and $X_P = X_{P'}$. We have that

$$A^M \setminus S = \bigcup_{u \in P'} [u]$$

because $x \in S$ if and only if $x \notin [u]$ for all $u \in P'$.

A finite pattern $w \in A^D$ is in the complement of L_S if and only if $[w] \cap S = \emptyset$. This is equivalent to

$$[w] \subseteq \bigcup_{u \in P'} [u].$$

By compactness of $[w]$, this is further equivalent to the existence of a finite $F \subseteq P'$ such that

$$[w] \subseteq \bigcup_{u \in F} [u].$$

This last condition can be tested algorithmically when w and F are given: One generates all patterns $e \in A^E$ where E is the union of the domains of w and the patterns in F, and verifies that for each such e, either $e_{|D} \neq w$ or $e_{|D(u)} = u$ for some $u \in F$ with domain $D(u)$. Because all finite subsets of a recursively enumerable P' can be recursively enumerated, we have a semi-algorithm for the complement of L_S. □

3.2 One-dimensional symbolic dynamics

See the last homework set for some results.