Notes on Topological Dynamics

Michael Hochman*

June 15, 2022

Abstract

These notes accompany a short advanced course at the Hebrew University, given in Spring 2022. The course introduces basic examples and results in topological dynamics.

Contents

1 Dynamical systems, orbits, and subsystems

1.1 Definitions

Dynamical systems theory deals with mathematical models of change, ones that evolve in an orderly way over time. By this we mean that there is some law governing the system's evolution; the law may be concrete, as in the theory of mechanics, where it is given by an explicit differential equation, or more abstract.

Topological dynamics is one of the abstract branches of the theory. All that it assumes is that the evolution is continuous, and that the state space is compact.

Definition 1.1. A (topological)[1] **dynamical system** is a pair (X, T) where

- $X \neq \emptyset$ is a compact metric space (called the **phase space** or **state space**).
- $T : X \to X$ is a continuous map.

When T is invertible, we say that (X, T) is an **invertible dynamical system**.

We often refer to the system by X or T, depending on the context. We denote the metric generically by $d(\cdot, \cdot)$, or by $d_X(\cdot, \cdot)$ when we want to emphasize the underlying space.

Given $T : X \to X$ and $n \in \mathbb{N}$, we define the n-th iterate of T by

$$T^n = \underbrace{T \circ T \circ T \circ \ldots \circ T}_{n}$$

Also $T^0 = \mathrm{id}_X$, and if T is invertible, $T^{-n} = (T^{-1})^n = (T^n)^{-1}$. The iterates of T form a semi-group or group acting on X.

The (forward) **orbit** of $x \in X$ is the sequence

$$O_T(x) = (T^n x)_{n=0}^{\infty}$$

When T is invertible, we define the **two-sided orbit** similarly, by

$$O_T^{\pm}(x) = (T^n x)_{n=-\infty}^{\infty}$$

Although the theory for invertible and non-invertible systems is not identical, the differences are minor. We shall usually present only one of the versions and leave the other to the reader.

Such systems arise naturally in classical physics, where each point in the phase space X represents a world state; for example, to describe a system with some number of particles (atoms or planets!), one can take $X \subseteq \mathbb{R}^N$ for some large N, with each state $x \in X$ representing the locations and momenta of the objects. A differential equation then defines the evolution: for each initial $x_0 \in X$, it determines a unique curve $x(t) : \mathbb{R} \to X$ with $x(0) = x_0$ and so that $x(t)$ is the state of the system at time t. From this, we obtain a map $T : X \to X$, given by evolving the system one time step: from initial state x we set $Tx = x(1)$, the location at time $t = 1$ of the evolution $x(t)$ of the system started from x. Then $T^n x = x(n)$, so the "orbit" of x describes

[1] The word "topological" refers to the fact that X is a topological space and T is continuous. Metrizability is not needed for many of the results below but we will assume it for convenience. The metric itself will not matter very much and its role is to give meaning to convergence, continuity etc. Thus, in almost all occurrences, it can safely be replaced with any equivalent metric.

the long term behavior of the system. Under mild conditions, the map T is continuous. Also, often X can be taken to be compact; although in our example we took $X = \mathbb{R}^N$, if we restrict to configurations of energy $\leq C$ for some constant C, this is often is a closed bounded region of $\mathbb{R}^N$. Then (X,T) is a dynamical system in the sense of definition 1.1.

There are a number of reasons for adopting our more abstract viewpoint. First, one does not always know the law governing the evolution of a system. So one would like to develop tools which apply in general. Surprisingly, it turns out that there is a lot that one can say.

Second, even when the law is explicitly known, or conjectured, this information rarely allows for long-term analysis of the system. In fact, interest in the abstract dynamics began from the failure of attempts to predict the long-term behavior of the motion of the planets in the solar system. Although the law of gravity was well understood, the problem withstood exact analysis. It was Poincare who, from general considerations which ignore the precise law of gravity, succeeded in make progress on the problem.

Finally, as is often the case in mathematics, the abstract dynamical formalism turns out to apply to problems beyond the original physical ones. We shall see that, for example, certain problems in combinatorics and diophantine approximation can be stated and solved dynamically.

1.2 The behavior of orbits, and first examples

A large part of the theory of dynamical systems concerns the long-term behavior of orbits. We introduce the following definitions.

Definition 1.2. Let (X,T) be a dynamical system.

- $x \in X$ is a **fixed point** (of T) if $Tx = x$.
- $x \in X$ is **periodic** (under T) if there exists $p \geq 1$ such that $T^p x = x$. The least such p is called the period of x.

  In particular a fixed point of T^p is a periodic point of T.
- $x \in X$ is **recurrent** (under T) if there exists $n_k \to \infty$ such that $T^{n_k} x \to x$.
- $x \in X$ is **transitive** if it has a dense (forward) orbit under T, i.e. if $O_T(x)$ is dense in X. In the invertible case we will speak **two-sided transitivity** when $O_T^{\pm}(x)$ is dense.

Not all of these types of behavior are found in every system, and when they are present, they can coexist in complicated ways. This is best demonstrated with some examples.

Example 1

Let

$$X = [0,1]$$
$$Tx = x^2$$

This is a homeomorphism. Observe:

1. The points 0 and 1 are fixed points and they are the only ones.
   (a) Every $x \in (0,1)$ satisfies $x^n \to 0$.

(b) In particular, 0 and 1 are the only recurrent points. However, both of them are limits of non-recurrent points.

(c) There are no points with dense orbit.

Example 2

Let

$$X = [0,1]$$
$$Tx = 10x \bmod 1$$

This map isn't continuous, but we can make it continuous by identifying the points $0, 1$, or equivalently, taking $X = \mathbb{R}/\mathbb{Z}$. We shall very frequently make this identification: formally we work in $\mathbb{R}/\mathbb{Z}$, but continue to represent points in $\mathbb{R}/\mathbb{Z}$ by their lift to $[0, 1)$. Note that

$$T^n x = 10^n x \bmod 1$$

Also observe that T is onto but is not injective (for example, $T(0) = T(0.1) = T(0.2) = 0$). So T is not invertible

1. If $x = s/t$, $s, t \in \mathbb{N}$, and $\gcd(10, t) = 1$, then there exists n such that $10^n = 1 \bmod t$, hence

   $$T^n(s/t) = 10^n s/t = s/t$$

   and s/t is a periodic point.

   It it not hard to see that there exist points with arbitrarily high period (e.g. $1/(10^N + 1)$).

   Also, for every prime $p \neq 2, 5$, the points k/p are periodic for all $0 \leq k < p$.

   It follows easily that **every** point $x \in X$ is the limit of periodic points, i.e.., the periodic points are dense.

   In particular, the set of points with non-dense orbit is dense.

2. If x is irrational, then $T^n x \neq x$; for otherwise, $10^n x = x \bmod 1$, so there is an integer N with $10^n x = x + N$, and $x = N/(10^n - 1) \in \mathbb{Q}$, contrary to assumption.

   Thus, the non-periodic points are also dense and **every** point is the limit of non-periodic points.

3. Do there exist transitive points? (points with dense orbit?) Let us describe a specific point with this property:

   $$c = 12345678910111213141516\ldots$$

   i.e. the number whose decimal expansion is the concatenation of the decimal expansions of the integers, in the usual order. This is known as the **Champernown number**, and its orbit is dense under T. Indeed, for every sequence $a = a_1 \ldots a_N$ of decimal digits, there is an N such that a appear in position n in the expansion of x. Then $10^n c \bmod 1 = 0.a_1 a_2 \ldots a_N \ldots$. This implies that every number $a = 0.a_1 a_2 a_3 \ldots$ in $[0, 1]$ can be approximated arbitrarily well by elements of $O_T(c)$, so the orbit of x is dense.

4. In fact, if $D \subseteq [0,1]$ is the set of points with dense T-orbit, then D itself is dense. Indeed, it is not hard to see that if $x \in D$ then $O_T(x) \subseteq D$. This follows from a general fact that we will see in the exercises, but let us give a concrete argument: Fix $x \in [0,1]$. Given N consider the point
$$x_N = 0.x_1 \ldots x_N 12345678910111213\ldots$$
Clearly $x_N \to x$. On the other hand, $T^N x = c$ so for $n \geq 0$ we have $T^n c = T^{N+n}x$, and hence $O_T(c) \subseteq O_T(x_N)$, so x_N has dense orbit.

5. It is easy to find points that are not recurrent. For example if $a_1 \ldots a_N \neq 0 \ldots 0$ is a block of decimal digits then $0.a_1 \ldots a_N 0000\ldots$ is not recurrent.

The example above was an endomorphism of a compact group, a class of maps that contains many other interesting examples. For example, a $d \times d$ integer matrix A acts on the d-dimensional torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}$, by multiplication modulo 1. This action shares many features with the example above. For instance, the periodic points are dense, since any point in $\mathbb{Q}^d/\mathbb{Z}^d$ has finite orbit, and if $\det A$ is relatively prime to the denominators, then the point will be periodic. Surely, there are also differences: the matrix A in general will not expand $\mathbb{T}^d$, and even if it does the expansion rate is not constant. We also lack something analogous to decimal expansion that would give us a more concrete understanding of the dynamics and allow us to construct points with specified behavior.

We will return to such examples in Section 3, where we shall see that a symbolic representation often does exist,

Example 3

Given $\alpha \in \mathbb{R}$, consider the system

$$X = \mathbb{R}/\mathbb{Z} \cong [0,1)$$
$$R_\alpha x = x + \alpha \bmod 1$$

Note that
$$R_\alpha^n x = x + n\alpha \bmod 1$$

Then R_α is a homeomorphism of X. In fact, $\mathbb{R}/\mathbb{Z}$ is group under addition modulo 1, and R_α is the translation by the element $\alpha \in \mathbb{R}/\mathbb{Z}$.

The behavior of orbits under R_α depends strongly on the nature of α, specifically on whether α is rational.

Indeed, if $\alpha \in \mathbb{Q}$, then **every** point is periodic! Indeed if $\alpha = s/t$ for some $s \in \mathbb{Z}$, $t \in \mathbb{N}$, then

$$R_\alpha^t x = x + t\alpha = x \bmod 1$$

In this case, every point is recurrent but no point has dense orbit.

On the other hand, if $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ then **no** point is periodic, since if $R_\alpha^p x = x$ then $x + p\alpha = x \bmod 1$, meaning that for some $N \in \mathbb{N}$ we have $x + n\alpha = x + N$, and hence $x = N/n \in \mathbb{Q}$, contrary to assumption. As for density, we have the following:

Theorem 1.3. *If $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ then every point in $\mathbb{R}/\mathbb{Z}$ has dense orbit under $R_\alpha x = x + \alpha \bmod 1$.*

Proof. We parameterize $\mathbb{R}/\mathbb{Z}$ as $[0,1)$. It suffices to prove that $n\alpha$ mod 1 is dense in $[0,1)$, since then for any $x \in [0,1)$, for any $y \in [0,1)$ we have $y - x = \lim_{k\to\infty} n_k\alpha$ mod 1 for some $(n_k) \subseteq \mathbb{N}$, and then

$$R_\alpha^{n_k} x = x + n_k\alpha \to x + (y - x) = y \bmod 1$$

showing that $y \in \overline{O_{R_\alpha}(x)}$. Since y was arbitrary this shows that x has dense orbit.

Fix $N \in \mathbb{N}$ and consider the intervals $[i/N, (1+i)/N)$ for $i = 0, 1, \ldots, N-1$. By the pigeonhole principle, there exist $0 \leq \ell_1 < \ell_2 < N+1$ such that $\ell_1\alpha, \ell_2\alpha$ fall in the same interval, so $|(\ell_2 - \ell_1)\alpha| < 1/N$. Writing $\beta = (\ell_2 - \ell_1)\alpha$ it follows that every two numbers in the sequence $\beta, 2\beta, 3\beta, [1 + 1/\beta]\beta$ are within $1/N$ of each other and similarly for the endpoint. Therefore every point in $[0,1)$ is within $1/N$ of one of these points. But $k\beta = k(\ell_2 - \ell_1)\alpha \in \{n\alpha \,:\, n \in \mathbb{N}\}$. So we have shown that $\{n\alpha\}$ is dense modulo one. □

Corollary 1.4. *For every α, every $x \in \mathbb{R}/\mathbb{Z}$ is recurrent under R_α.*

Proof. For α rational all points are periodic so the conclusion is obvious. Otherwise, we have seen that $n\alpha$ is dense in $\mathbb{R}/\mathbb{Z}$ so $n_k\alpha \to 0$ for some $(n_k) \subseteq \mathbb{N}$. Clearly $n_k \to \infty$, since $n\alpha \neq 0$ mod 1 for all $n \in \mathbb{N}$. Then given $x \in \mathbb{R}/\mathbb{Z}$ we have

$$R_\alpha^{n_k} x = x + n\alpha \to x \bmod 1$$

so x is recurrent. □

This example raise the following natural question: For a set $S \subseteq \mathbb{N}$ of natural numbers and α an irrational number, when is

$$S\alpha = \{s\alpha \,:\, s \in S\}$$

dense modulo one?

We have just answered the question when $S = \mathbb{N}$. What about sequences of the form $S = \{p(n) \,:\, n \in \mathbb{N}\}$, where $p(t)$ is a polynomial? Then $(p(n)\alpha)_{n=1}^\infty$ is not the orbit of a map on $\mathbb{R}/\mathbb{Z}$, at least not in any obvious way. Nonetheless, there is a way to view it in a dynamical context. Using dynamical ideas, later on we will prove:

Theorem 1.5 (Hardy-Littlewood)**.** *If $\alpha \notin \mathbb{Q}$ and p is a polynomial with an irrational non-constant coefficient, then $(p(n))_{n=1}^\infty$ is dense modulo 1.*

Beyond polynomials, one can ask about sequences of the form $S = \{a^n \,:\, n \in \mathbb{N}\}$. We have already answered this question: For $S = \{10^n\}_{n=1}^\infty$, then there exist irrational α such that $S\alpha$ is not dense, and also irrational α for which $S\alpha$ is dense. Our argument extends to all sets of the form $\{k^n \,:\, n \in \mathbb{N}\}$ for a fixed $2 \leq k \in \mathbb{N}$.

There exist sequences lying between the polynomial and exponential. A famous example is $S = \{2^k 3^\ell \,:\, k, \ell \in \mathbb{N}\}$. One can show that, when placed in increasing order $S = \{n_1 < n_2 < \ldots\}$, one has $n_{k+1}/n_k \to 1$ (we shall prove this later). Such a sequence is called **non-lacunary** (a **lacunary** sequence is one for which $\liminf n_{k+1}/n_k > 1$). On the other hand S is not a polynomial set.

One may also notice that $S = \{2^k 3^\ell \,:\, k, \ell \in \mathbb{N}\}$ is a semigroup under multiplication, and Sx mod 1 is the orbit of the joint action of two commuting maps of $\mathbb{R}/\mathbb{Z}$, the map $x \mapsto 2x$ and $x \mapsto 3x$. Each of these maps separately has many non-dense, non-periodic, even non-recurrent orbits. Remarkably, it turns out that together they behave much more like the polynomial case:

Theorem 1.6 (Furstenberg)**.** *If α is irrational then $S\alpha$ is dense modulo one.*

We will prove this in Section **??**.

Example 4

Let Λ be a finite set which we view as an alphabet of symbols, and let $\Lambda^{\mathbb{N}}$ or $\Lambda^{\mathbb{Z}}$ denote the space of sequences (one- or two-sided) over Λ. These are compact metrizable spaces in the product topology (viewing Λ as a discrete set), and we can define the **shift map** on it by

$$(Sx)_n = x_{n+1}$$

This maps "shifts sequences to the left", and one can check that it is a continuous onto map of $\Lambda^{\mathbb{N}}$ and ; in the the latter case is also invertible. The dynamical systems $(\Lambda^{\mathbb{N}}, S)$ and $(\Lambda^{\mathbb{Z}}, S)$ are called the one-sided and two-sided **full shift**, respectively.

These systems are a rich source of examples. they also arise in applications as a natural compactification of discrete objects. We give two examples.

First, suppose we are modeling configurations of repelling particles arranged on the line. We might then denote an empty spot by 0 and an occupied one by 1. Then repulsion would mean that no two particles can be adjacent, so we are looking at words which do not contain the sub-word 11. Compactifying the space of finite words, we arrive at the set of infinite words

$$X = \{x \in \{0,1\}^{\mathbb{Z}} \;:\; x \text{ does not contains the subword } 11\}$$

This is a compact subset of the full shift $\{0,1\}^{\mathbb{Z}}$ and evidently is invariant under S. So (X, S) is a dynamical system. Similarly, repulsion could be understood in the average sense; for example, that no finite segment of a word contains more than 1/3 of its sites occupied. We would then get a different dynamical system.

Second, any subset of $A \subseteq \mathbb{Z}$ can be identified with the sequence $1_A \in \{0,1\}^{\mathbb{Z}}$. We can think of it as a point in the full shift. Often, combinatorial properties of A translate to properties of this point, or points in its orbit, or in its orbit closure. For example, if $\{n, n+1\} \not\subseteq A$ for all n, then 1_A its entire orbit are contained in the compact set X above. We shall see later that this point of view leads to deep combinatorial results.

Summary

Let us collect some of the phenomena we have seen in these examples:

- A periodic/recurrent/transitive point can be a limit of points with different properties.
- A system may, or may not, contain periodic points (it might happen that all point or no points are periodic, and when they exist, the periodic points may be dense, or not).
- A system may, or may not, contain transitive points (also, all points might be transitive, or none. When they exist they may form a dense set).

This hints at the complexity that we can expect from dynamical systems in general. But there is also order. For instance, in all the examples we saw so far, there existed recurrent points, and we shall soon see that this is a general phenomenon.

1.3 Subsystems, minimality and recurrence

So far we have considered individual orbits. We now shift perspective and consider invariant sets.

Definition 1.7. A **subsystem** of a dynamical system (X, T) is a closed, non-empty subset $Y \subseteq X$ with $TY \subseteq Y$ (if T is invertible, the condition becomes $TY = Y$). Such Y is also called an **invariant (closed) subset** of X.

Note that if $Y \subseteq X$ is a subsystem of (X, T), then $(Y, T|_Y)$ is a dynamical system (usually we just write T, instead of $T|_Y$). Also note that there are many non-closed invariant subsets, but we shall always mean closed ones unless otherwise stated.

Examples and basic properties

1. X is a subsystem of (X, T).
2. If $x_0 \in X$ is a fixed point of T then $\{x_0\}$ is a subsystem. More generally, if x_0 is periodic with period p, then $\{x_0, Tx_0, \ldots, T^{p-1}x_0\}$ is a subsystem.
3. The system $X = \{x \in \{0,1\}^{\mathbb{Z}} \,:\, x \text{ does not contain the word } 11\}$ with the shift map S, is a subsystem of the full shift.
4. Given $x \in X$, its **orbit closure** $Y = \overline{O_T(x)}$ is a subsystem, and furthermore, if x is recurrent, then T is surjective on Y.

   To see this note that the set is closed by definition, non-empty because $x \in Y$, and invariant because if $y \in Y$ then $y = \lim T^{n_K}x$ for some sequence $(n_k) \subseteq \mathbb{N}$, so by continuity of T,

   $$Ty = T(\lim_{n\to\infty} T^{n_k}x) = \lim_{n\to\infty} T^{n_k+1}x \in \overline{Y}$$

   If x is recurrent then we can assume that $n_k \geq 1$, since otherwise $y = x$, and we may by recurrence assume even that $n_k \to \infty$. Then, passing to a subsequence if necessary, we can assume that $T^{n_k-1}x \to z \in Y$, so again by continuity of T we have

   $$y = \lim_{n\to\infty} TT^{n_k}x = T\left(\lim_{n\to\infty} T^{n_k-1}x\right) = Tz$$

   which gives surjectivity of $T : \overline{O_T(x)} \to \overline{O_T(x)}$.

   Note that if x_0 is periodic with period p then its orbit $O_T(x_0) = \{x_0, Tx_0, \ldots, T^{p-1}x_0\}$ is already closed so this is the orbit closure.
5. When T is invertible, $\overline{O_T^{\pm}(x)}$ is T and T^{-1} invariant, and T is always bijective on it.
6. If $Y_i \subseteq X$ are subsystems for $i \in I$, then $Y = \bigcap_{i\in I} Y_i$ is a subsystem. It is closed as an intersection of closed sets, and if $y \in Y$ then for each i we have $y \in Y_i$ so $Ty \in Y_i$, showing that $Ty \in Y$.

Definition 1.8. A dynamical system (X, T) is **minimal** if it contains no non-trivial subsystems.

Lemma 1.9. *A system (X, T) is minimal if and only if every orbit is dense.*

Proof. For $x \in X$, since $\overline{O_T(x)} \subseteq X$ is a subsystem, minimality implies $\overline{O_T(x)} = X$ for all $x \in X$, so the orbit of x is dense.

Conversely, if every orbit is dense and $Y \subseteq X$ is a non-empty subsystem, take $y \in Y$. Then $O_T(y) \subseteq Y$ (because $TY \subseteq Y$), so $\overline{O_T(y)} \subseteq Y$ (because Y is closed), and we get $X = \overline{O_T(y)} \subseteq Y \subseteq X$, so $Y = X$. □

Theorem 1.10. *Every dynamical system has a minimal subsystem.*

Proof. Consider the family of set of subsystems of X, partially ordered by inclusion. Using Zorn's lemma, find a maximal chain $\{Y_i\}_{i\in I}$ in this family (so $\{Y_i\}$ is totally ordered by inclusion and it cannot be extended to a larger such family). Now $Y = \bigcap_{i\in I} Y_i$ is a non-empty subsystem (a subsystem because it is the intersection of subsystems; non-empty because $\{Y_i\}$, being totally ordered by inclusion, has the finite intersection property). Y is minimal because if $Y' \subseteq Y$ we a non-trivial subsystem, then $\{Y_i\}_{i\in I}$ could be enlarged by adding Y' to it, contradicting maximality. ☐

Lemma 1.11. *If (X, T) is minimal, then every point is recurrent.*

Proof. Let $x \in X$. If x is a periodic point, clearly it is recurrent.

Otherwise, $T^n x \neq x$ for all $n \geq 1$. By minimality, the orbit of $y = Tx$ is dense, so $x = \lim_{k\to\infty} T^{n_k} y$ for some $(n_k) \subseteq \mathbb{N}$. Since none of $T^{n_k} y$ is equal to x we must have $n_k \to \infty$, and then $x = \lim_{k\to\infty} T^{n_k+1} x$, showing that x is recurrent. ☐

Corollary 1.12 (Birkhoff's recurrence theorem)**.** *Every dynamical system contains recurrent points.*

We can use this theorem to give another, more abstract proof that every point in $\mathbb{R}/\mathbb{Z}$ is recurrent under R_α. For by the last result there is some $x_0 \in \mathbb{R}/\mathbb{Z}$ that is recurrent under R_α, so $R_\alpha^{n_k} x_0 \to x_0$ with $n_k \to \infty$. Notice that $R_\alpha R_\beta x = R_\beta R_\alpha x$ for all x, α, β. Thus, if x be any other point, then

$$\begin{aligned} R_\alpha^{n_k} x &= R_\alpha^{n_k} R_{x-x_0}(x_0) \\ &= R_{x-x_0} R_\alpha^{n_k} x_0 \\ &\to R_{x-x_0} x_0 \\ &= x \end{aligned}$$

1.4 Uniformly recurrent points

In this section, we ask: when is an orbit closure minimal? First, we give another characterization of minimality.

Lemma 1.13. *(X, T) is minimal if and only if for every open $U \neq \emptyset$ there is an $N \in \mathbb{N}$ such that $X = \bigcup_{n=1}^N T^{-n} U$.*

Proof. Suppose (X, T) is minimal and let $U \neq \emptyset$ be open. The orbit of any $x \in X$ is dense by minimality, so $T^n x \in U$ for some $n \geq 0$; equivalently $x \in T^{-n} U$, so $x \in \bigcup_{n=0}^\infty T^{-n} U$. Since X was arbitrary, this means that the union covers X. By compactness finitely may of the sets already cover X.

Conversely, let $x \in X$, it is enough to show that $O_T(x)$ is dense (if every orbit is dense, (X, T) is minimal). Density of the orbit means that for every $U \neq \emptyset$ open there is an n such that $T^n x \in U$. So given U we must show that $x \in \bigcup_{n=1}^\infty T^{-n} U$, which follows from the assumption that the union is all of X. ☐

Definition 1.14. A set $E \subseteq \mathbb{N}$ or $E \subseteq \mathbb{Z}$ is **syndetic** if it has bounded gaps, i.e. if there is a $C \geq 1$ such that $[a, a + C] \cap E \neq \emptyset$ for every $a \in \mathbb{R}$.

Given a dynamical system (X,T), a point $x \in X$ and a set $U \subseteq X$ we write

$$N(x,U) = \{n \in \mathbb{N} \, : \, T^n x \in U\}$$

Lemma 1.15. *If (X,T) is minimal then $N(x,U)$ is syndetic for every $x \in X$ and every open set $\emptyset \neq U \subseteq X$.*

Proof. Suppose X is minimal. Let N be such that $X \subseteq \bigcup_{i=0}^{N} T^{-i}U$. Then for every k, we know that $T^k x \in \bigcup_{i=0}^{N} T^{-i}U$, i.e. $\{i, i+1, \ldots, i+N\} \cap N(U,x) \neq \emptyset$. This shows that every interval of length N intersects $N(x,U)$, so it is syndetic. □

There is a converse, with even weaker hypotheses: If $N(x,U) \neq \emptyset$ for all x, U as in the lemma, then every x has dense orbit, so X is minimal by Lemma 1.9.

Definition 1.16. Let (X,T) be a dynamical system. A point $x \in X$ is **uniformly recurrent** if $N(x,U)$, is syndetic for every open set $U \subseteq X$ containing x.

Proposition 1.17. *The orbit closure of x is minimal if and only if it is uniformly recurrent.*

Proof. Let $Y = \overline{O_T(x)}$. Suppose a Y is minimal. If $U' \subseteq X$ is open and contains x, certainly $U = Y \cap U'$ is open in Y, and also $\{n \in \mathbb{N} \, : \, T^n x \in U\} = \{n \in \mathbb{N} \, : \, T^n x \in U'\}$ (because Y is invariant). By minimality the former set is syndetic, so the latter is also.

Conversely, suppose x is uniformly recurrent. Let $Y = \overline{O_T(x)}$, and use Theorem 1.10 to find a minimal system $Z \subseteq Y$. Fix $z \in Z$. We will show that $x \in \overline{O_T(z)}$. This implies that $x \in Z$ so $\overline{O_T(x)} = Z$ is minimal.

Let $\varepsilon > 0$. Since n is uniformly recurrent, $N(x, B_\varepsilon(x))$ is syndetic, so there exists $\ell \geq 1$ such that $N(x, B_\varepsilon(x)) \cap [i, i+\ell] \neq \emptyset$ for all i. Using continuity of T, choose $\delta > 0$ such that $d(w,w') < \delta$ implies $d(T^n w, T^n w') < \varepsilon$ for $0 \leq n \leq \ell$.

Now, $z \in \overline{O_T(x)}$, so there is an n_0 with $d(T^{n_0}x, z) < \delta$. Choose $0 \leq n_1 \leq \ell$ such that $n_0 + n_1 \in N(x, B_\varepsilon(x))$, so that

$$d(T^{n_1}T^{n_0}x, x) = d(T^{n_0+n_1}x, x) < \varepsilon$$

but also, since $d(T^{n_0 x}, z) < \delta$ we have (by choice of δ):

$$d(T^{n_1}T^{n_0}x, T^{n_1}z) < \varepsilon$$

Combining these we have

$$d(T^{n_1}z, x) < 2\varepsilon$$

Summarizing, for all $\varepsilon > 0$ we found n_1 such that $d(T^{n_1}z, x) < 2\varepsilon$, so $x \in \overline{O_T(z)}$, as claimed. □

1.5 Problems

Let (X,T) be a dynamical system.

1. Show that if x has dense orbit in X and x is not an isolated point, then x is recurrent, and every $y \in O_T(x)$ has a dense orbit.
2. Show that if y has dense orbit and $y \in \overline{O_T(x)}$ then x has dense orbit.

3. If $y \in \overline{O_T(x)}$, does it follow that $x \in \overline{O_T(y)}$?
4. Let $T : X \to X$ be a continuous map of a metric space. Show that $\bigcap_{n=1}^{\infty} T^n(X)$ is a subsystem of X. (This is only interesting when T is not onto).
5. Let $D = \{z \in \mathbb{C} \ : \ |z| \leq 1\}$ denote the closed unit disc. Let $T : D \to D$ denote the map that rotates the circle of radius r by angle r: that is, $T(re^{i\theta}) = re^{i(\theta+r)}$. Describe the periodic points and minimal subsets of (D, T_α).
6. Show that if $x \in \mathbb{R}/\mathbb{Z}$ is irrational and $Tx = 10x \bmod 1$, then $\overline{O_T(x)} \setminus O_T(x)$ is infinite.
7. Let $X = \mathbb{R}^2/\mathbb{Z}^2$ denote the two-dimensional torus and let $\alpha = (\alpha_1, \alpha_2) \in X$. Show that $R_\alpha x = x + \alpha$ is minimal if and only if α_1, α_2 are rationally independent, i.e. if the only integer solution of $\alpha_1 x + \alpha_2 y = 0$ is $x = y = 0$.

Definition 1.18. Let (X, T) be a dynamical system. A set $A \subseteq X$ is **wandering** if $A \cap T^{-n} A = \emptyset$ for all $n \geq 1$.

We say that (X, T) is **non-wandering** if no non-empty open set is wandering.

8. Show that every invertible system $(X.T)$ decomposes uniquely into a disjoint union $(\bigcup_{n=-\infty}^{\infty} T^{-n}U) \cup Y$ where U is and $Y \subseteq X$ is a non-wandering subsystem.
9. Let (X, T) be a non-wandering dynamical system. Show that the set of recurrent points is a dense G_δ subset of X.

   Hint: Let $\varepsilon > 0$. Show that the set
   $$C_\varepsilon = \{x \in X \ : \ \exists n \in \mathbb{N} \ : \ d(T^n x, x) < 2\varepsilon\}$$
   contains a dense open set by showing that for every ball B of radius $\varepsilon > 0$, the set $A = B \setminus \cup_{n=1}^{\infty} T^{-n} B$ has empty interior.

2 Isometries, equicontinuity, and group translations

Our goal in this section is to study isometries, which are perhaps the simplest maps in the context of metric spaces. As a motivating example, it is good to keep in mind the rotations $R_\alpha x = x + \alpha \bmod 1$ of $\mathbb{R}/\mathbb{Z}$ that we encountered in the last chapter. In this example we have several structures simultaneously:

- R_α is an isometry in the induced metric metric from $\mathbb{R}$.
- $\mathbb{R}/\mathbb{Z}$ is an abelian group with continuous group operations, and R_α is addition by a constant group element.

And on the dynamical side, we also saw that

- Either the system is minimal (when $\alpha \notin \mathbb{Q}$), or else it is the disjoint union of periodic orbits, each of which is itself a minimal subsystem, and a coset of a finite subgroup of $\mathbb{R}/\mathbb{Z}$.

We will soon see that the coexistence of these properties is not a coincidence, and that, in fact, all isometries share them.

In this section, all dynamical systems are invertible.

2.1 Isometries and equicontinuous systems

Definition 2.1. An **isometry** f between metric spaces (X, d_X) and (Y, d_Y) is a distance-preserving function $f : X \to Y$, that is:

$$d_Y(f(x), f(x')) = d_X(x, x') \qquad \text{for all } x, x' \in X$$

A dynamical system (X, T) is **isometric** if $T : X \to X$ is an isometry.

Lemma 2.2. *An isometry $T : X \to X$ of a compact metric space is invertible.*

Proof. Since T preserves distances it is injective, so we must show that it is onto. If $TX \neq X$ then we can take $x_0 \in X \setminus TX$. Setting $\delta = d(x_0, TX)$ we note that $d(T^2x_0, Tx_0) \geq \delta$ because T is an isometry, and that $d(T^2x_0, x_0) \geq \delta$ because $T^2x_0 \in TX$. By induction, one shows that $d(T^n x_0, T^k x_0) \geq \delta$ for all $n > k$, contradicting compactness. □

The definition of an isometric system depends on the metric. Since the metric often is not specified (e.g. in product spaces, there are many natural choices), our first objective is to determine when a map is equivalent to an isometry, in the sense that, by a slight change of the metric, it becomes one. For this we require the following definition.

Definition 2.3. A dynamical system (X, T) is **equicontinuous** if it is invertible and the family of maps $\{T^n \ : \ n \in \mathbb{Z}\}$ is equicontinuous, that is: For every $\varepsilon > 0$ there exists $\delta > 0$ such that for every $x, x' \in X$ and every $n \in \mathbb{Z}$, if $d(x, x') < \delta$ then $d(T^n x, T^n x') < \varepsilon$.

This definition also uses the metric but it is a simple exercise to show that it in fact depends only on the topology, in the following sense. Recall that two metrics d_0, d_1 on the same space are **equivalent** if the identity map $X \to X$ is continuous both as a map from (X, d_0) to (X, d_1), and as a map from (X, d_1) to (X, d_0). Equivalently, for every $\varepsilon > 0$ there is a $\delta > 0$ such that $d_i(x, y) < \delta$ implies $d_{1-i}(x, y) < \varepsilon$ for $i = 1, 2$. The a family of functions is equicontinuous for some metric if and only if it is equicontinuous for every equivalent metric.

Proposition 2.4. *If a dynamical system (X, T) is isometric then it is equicontinuous. Conversely, if it is equicontinuous, then there is an equivalent metric on X that makes T into an isometry.*

Proof. All the T^n are isometries, so the first direction is trivial (in fact even the larger group of all isometries of X is equicontinuous).

Conversely, suppose that d is a metric on X and $\{T^n\}_{n\in\mathbb{Z}}$ is equicontinuous. Define metric d_n on X by

$$d_n(x,y) = \frac{1}{n}\sum_{i=0}^{n-1} d(T^i x, T^i y)$$

It is clear that these are metrics. Also, it is not hard to check that the family $\{d_n\}_{n\geq 0}$ is equicontinuous as a subset of $C(X \times X, \mathbb{R})$. Therefore by Arzela-Ascolli, there is a uniformly converging subsequence

$$d_{n_k} \to d_\infty$$

Since d_∞ is a pointwise limit of metrics, it is clearly a pseudo-metric, i.e. it is non-negative, symmetric, and satisfies the triangle inequality. The fact that it is positive follows from the fact that it is equivalent to d. Let us show, for example, that for every $\varepsilon > 0$ there is a $\delta > 0$ such that if $d_\infty(x,y) < \delta$ then $d(x,y) < \varepsilon$. Indeed, by equicontinuity of $\{T^n\}$, given $\varepsilon > 0$ there exists $\delta > 0$ such that $d(x,y) < \delta$ implies $d(T^n x, T^n y) < \varepsilon$ for all n. Now, if $d_\infty(x,y) = \lim \frac{1}{n_k}\sum_{i=1}^{n_k} d(T^i x, T^i y) < \delta$, then clearly $d(T^i x, T^i y) < \delta$ for some i, so $d(x,y) < \varepsilon$. In the other direction: Given $\varepsilon > 0$, if we choose $\delta > 0$ so that $d(x,y)$ implies $d(T^n x, T^n y)$ for all n, then $d(x,y) < \delta$ implies $d_n(x,y) < \varepsilon$ for all n and hence $d_\infty(x,y) \leq \varepsilon$.

Finally, we claim that T is an isometry with respect to d_∞. Indeed,

$$\begin{aligned}
d_\infty(Tx, Ty) &= \lim_{k\to\infty} d_{n_k}(Tx, Ty) \\
&= \lim_{k\to\infty} \frac{1}{n_k}\sum_{i=0}^{n_k-1}(T^{n_k}Tx, T^{n_k}Ty) \\
&= \lim_{k\to\infty} \frac{1}{n_k}\left(d(T^{n_k+1}x, T^{n_k+1}y) + \sum_{i=0}^{n_k-1}(T^{n_k}x, T^{n_k}y) + d(x,y)\right) \\
&= d_\infty(x,y)
\end{aligned}$$

In the last equality we used compactness of X to bound the diameter of X and thus eliminate the first and last summand in the line before last. $\square$

An alternative proof, slightly shorter, can be given using the metric $d'_\infty(x,y) = \sup_{n\in\mathbb{Z}} d(T^n x, T^n y)$.

We now turn to the dynamics.

Lemma 2.5. *If (X, T) is an invertible dynamical system and $x \in X$ is recurrent, then $\overline{O_T(x)} = \overline{O_T^{\pm}(x)}$.*

Proof. The inclusion $\subseteq$ is trivial since $O_T(x) \subseteq O_T^{\pm}(x)$.

Conversely suppose $y \in \overline{O_T^{\pm}(x)}$. Given $\varepsilon > 0$, there is an $\ell \in \mathbb{Z}$ such that $T^\ell x \in B_\varepsilon(y)$. Now let $n_k \to \infty$ such that $T^{n_k}x \to x$, which exists because x is recurrent. Then

$$T^{\ell+n_k}x = T^\ell(T^{n_k}x) \to T^\ell(x) \in B_\varepsilon(y)$$

so for all large enough k we have $T^{\ell+n_k}x \in B_\varepsilon(y)$. Since $\ell + n_k \to \infty$, we have shown that a positive iterate of T sends x to $B_\varepsilon(y)$ so, since $\varepsilon > 0$ was arbitrary, $y \in \overline{O_T(x)}$. □

Proposition 2.6. *If (X,T) is an equicontinuous system then every orbit closure is minimal, X decomposes into a disjoint union of minimal sets, and if X contains a point with a dense orbit, then it is minimal.*

Proof. The statement is clearly independent of the metric used (as long as we do not change the topology), so by the previous proposition, we can assume that T is an isometry.

Fix $x \in X$. Choose a minimal subsystem $Y \subseteq \overline{O_T(x)}$. Fix $y \in Y \subseteq \overline{O_T(x)}$, so $y = \lim T^{n_k}x$ for some $n_k \to \infty$, or equivalently, $d(T^{n_k}x, y) \to 0$. Since T is an isometry, this is the same as $d(x, T^{-n_k}y)$, so $x \in \overline{O_T^{\pm}(y)}$. But y is recurrent (Corollary 1.11), so by the previous lemma, $x \in \overline{O_T(y)} = Y$. Therefore x belongs to a minimal subsystem and in fact $\overline{O_T(x)} = Y$.

Every two minimal sets are disjoint or equal (since their intersection is a subsystem both one of them). We have just proved that every point belongs to a minimal subsystem. The fact that X decomposes into a disjoint union of minimal subsystems is now immediate.

The last statement follows: if $X = \overline{O_T(x)}$ for some x, then by the previous part of the proof, X is minimal. □

As a consequence of this theorem we have:

Corollary 2.7. *In an equicontinuous system, every point is uniformly recurrent.*

2.2 Group translations

We have already mentioned that R_α acting on $\mathbb{R}/\mathbb{Z}$ is an example of a group translation: $\mathbb{R}/\mathbb{Z}$ is a group and $R_\alpha : x \mapsto x + \alpha \bmod 1$ acts by translating by the fixed element α (if a-priori $\alpha \in \mathbb{R}$, we reduce it modulo 1, and we have the same map R_α). More generally,

Definition 2.8. A **compact metric group** is a group G together with a compact metric, with respect to which the group operations are continuous; i.e. $(h,g) \mapsto hg$ is continuous as a map $G \times G \to G$ and $g \mapsto g^{-1}$ is continuous as a map $G \to G$.

In other words, if $g_n \to g$ and $h_n \to h$ in G then $g_n h_n \to gh$ and $g_n^{-1} \to g^{-1}$.

Examples

1. $\mathbb{R}/\mathbb{Z}$ and more generally, $(\mathbb{R}/\mathbb{Z})^d$, the d-dimensional torus, are compact abelian groups.
2. The group of $n \times n$ orthogonal matrices under matrix multiplication (the metric can be taken to be the operator norm on linear transformations, or any norm on the linear space of matrices, restricted to the orthogonal ones). This space is compact because it is a closed and bounded subspaces of a finite dimensional normed space. The group operations are continuous because they are polynomial in the coordinates; inversion is continuous because it is a permutation of coordinates (transpose).
3. Any finite group.
4. A countable product of finite groups in the product topology.

Given $g \in G$, the **translation by** g is the map $R_g : G \to G$ given by $h \mapsto gh$. This notation is slightly confusing because R_g is the left translation; it is sometimes denoted L_g, and R_g the right translation $h \mapsto hg$. In non-abelian groups these are different maps. We retain R_g for consistency to remind us it is a Rotation.

Lemma 2.9. *If G is a compact group then the family of maps $\{R_g\}_{g \in G}$ is equicontinuous.*

Proof. If not, then there exists some $\varepsilon > 0$ and a sequence $\delta_n \to 0$, and elements $g_n \in G$ and $h_n, h'_n \in G$, such that $d(h_n, h'_n) < \delta_n$ but $d(g_n h_n, g_n h'_n) \geq \varepsilon$. Using compactness of G and passing to a subsequence if necessary, we can assume that $g_n \to g$, $h_n \to h$ and $h'_n \to h'$. Of course $h = h'$ because $d(h_n, h'_n) \leq \delta_n \to 0$. But by continuity of multiplication and of the metric, we get

$$0 = d(gh, gh) = d(gh, gh') = \lim d(g_n h_n, g_n h'_n) \geq \varepsilon$$

which is impossible. □

Proposition 2.10. *Let G be a compact metric group G with identity element e, and $g \in G$. Then*

1. *$E = \overline{O_{R_g}(e)}$ is a compact abelian subgroup of G*
2. *$\overline{O_{R_g}(h)} = Eh$ for every $h \in G$, and the decomposition of G into R_g-minimal sets is the decomposition into right E-cosets.*
3. *If (G, R_g) contains a transitive point then it is minimal and G is abelian.*
4. *There exists an R_g-invariant metric.*[2]

Proof. First note, that for any h, we have $R_g^n h = g^n h$.

(1) Let $E = \overline{O_{R_g}(e)}$. Then $e \in E$ and by Lemma 2.9, R_g acts equicontinuously and minimally on E. If $h_1, h_2 \in E$ then $h_1 = R_g^{n_k} e = \lim g^{n_k}$ and $h_2 = \lim R_g^{m_k} e = \lim g^{m_k}$ for some sequences $(n_k), (m_k)$, so by continuity of the product in G,

$$\begin{aligned} h_1 h_2 &= (\lim_{k\to\infty} g^{n_k})(\lim_{k\to\infty} g^{m_k}) \\ &= \lim_{k\to\infty} (g^{n_k} g^{m_k}) \\ &= \lim_{k\to\infty} g^{n_k + m_k} \end{aligned}$$

This shows that $h_1 h_2 \in E$. If we compute $h_2 h_1$, we get the same limit (since $g^{n_k+m_k} = g^{m_k+n_k}$, so $h_1 h_2 = h_2 h_1$. Finally if $R_g^{n_k} e = g^{n_k} \to h$ then

$$R_g^{-n_k} e = g^{-n_k} = (g^{n_k})^{-1} \to h^{-1}$$

so $h^{-1} \in \overline{O^{\pm}_{R_g}(e)} = E$, where the last equality is because e is recurrent and by Lemma 2.5. We have shown that E is an abelian subgroup of G. Compactness is immediate.

(2) We have $R_g^n h = (R_g^n e)h$, so $O_{R_g}(h) = (O_{R_g}(e))h$, hence $\overline{O_{R_g}(h)} = \overline{O_{R_g}(e)}h$, giving (2).

(3) follows from (1) and Proposition 2.6.

(4) follows from Proposition 2.4. □

[2]In fact, a compact group admits a metric such that R_g is an isometry for all $g \in G$ simultaneously, but this is slightly harder to prove when G is nor abelian.

Perhaps more surprisingly, the converse is also true:

Theorem 2.11. *Suppose that (X, T) is an isometric system and some point has a dense orbit. Then we can give X the structure of a compact abelian group, and $T = R_x$ for some $x \in X$.*

Proof. Consider the group Γ of isometries of X with the sup metric,

$$d(\gamma, \gamma') = \sup_{y \in X} d(\gamma(y), \gamma'(y))$$

Then (Γ, d) is a complete metric space, the group operations are continuous,[3] and d is invariant: $d(\gamma \circ \delta, \gamma' \circ \delta) = d(\gamma, \gamma')$.

Let $y_0 \in X$ have dense orbit and set $X_0 = \{T^n y_0\}_{n \in \mathbb{Z}}$. If the orbit is finite, $X = X_0$ is a finite set permuted cyclically by T, so the statement is trivial. Otherwise $y \in X_0$ uniquely determines n such that $T^n y_0 = y$ and we can define $\pi : X_0 \to \Gamma$ by $y \mapsto T^n \in \Gamma$ for this n.

We claim that π is an isometry. Fix $y, y' \in X_0$, so $y = T^n y_0$ and $y' = T^{n'} y_0$, so

$$d(\pi y, \pi y') = \sup_{z \in X} d(T^n z, T^{n'} z)$$

Given $z \in X$ there is a sequence $n_k \to \infty$ such that $T^{n_k} y_0 \to z$. But then

$$\begin{aligned}
d(T^n z, T^{n'} z) &= d(T^n(\lim T^{n_k} y_0), T^{n'}(\lim T^{n_k} y_0)) \\
&= \lim d(T^n T^{n_k} y_0, T^{n'} T^{n_k} y_0) \\
&= \lim d(T^{n_k}(T^n y_0), T^{n_k}(T^{n'} y_0)) \\
&= \lim d(T^n y_0, T^{n'} y_0) \\
&= d(T^n y_0, T^{n'} y_0) \\
&= d(y, y')
\end{aligned}$$

Thus $d(\pi y, \pi y') = d(y, y')$ and π is an isometry $X_0 \hookrightarrow \Gamma$. Furthermore, for $y = T^n y_0 \in X_0$,

$$\pi(Ty) = \pi(T T^n y_0) = T^{n+1} = R_T(T^n) = R_T \pi(y)$$

It follows that π extends uniquely to an isometry with $X \hookrightarrow \Gamma$ also satisfying $\pi(Ty) = R_T(\pi y)$. The image $\pi(X_0)$ is compact, being the continuous image of the compact set X. Since $\pi(X_0) = \{T^n\}_{n \in \mathbb{Z}} = O_{R_T}(\mathrm{id}_X)$, this is a group, and its closure is also a group G. □

2.3 Discrete spectrum and the Halmos-von Neumann theorem

Definition 2.12. Dynamical systems (X, T) and (Y, S) are **isomorphic** or **topologically conjugate** is there is a homeomorphism (continuous, 1-1 onto map) $\pi : X \to Y$ map $\pi : X \to Y$ that intertwines the action, i.e.

$$S\pi = \pi T$$

Such a map π is an **isomorphism** of the systems. An isomorphism $(X, T) \to (X, T)$ is called an **automorphism**.

[3] A word of warning: in the space of all continuous self-maps of X, the set of isometries is closed in the metric d, but the set of homeomorphisms is not. To fix this, one uses the metric $d'(\gamma, \delta) = d(\gamma, \delta) + d(\gamma^{-1}, \delta^{-1})$. But for isometries this correction is not needed.

Isomorphism preserves all topological properties of orbits, and we would like to classify systems up to this relation. Our study of equicontinuous systems does not yet achieve this goal – we have described how such systems arrise but not when they are are isomorphic. Note that isomorphism of the underlying groups is not the same as isomorphism of the systems. For example, for irrational α, β, are $(\mathbb{R}/\mathbb{Z}, R_\alpha)$ and $(\mathbb{R}/\mathbb{Z}, R_\beta)$ isomorphic? We shall now show that the answer is generally negative, and obtain a complete classification of minimal equicontinuous systems up to isomorphism.

Denote the unit circle by $S^1 = \{z \in \mathbb{C} \,:\, |z| = 1\}$.

Definition 2.13. Let (X,T) be a dynamical system. A **continuous eigenfunction with eigenvalue** $\lambda \neq 0$ is a continuous function $f : X \to S^1$, such that $f \circ T = \lambda f$. The **discrete spectrum** of (X,T) is the set $\Sigma(X,T)$ of eigenvalues.

These notions are just the usual linear algebra definitions to the linear map $g \mapsto g \circ T$ induced by T on $C(X)$. Note that $\Sigma(X,T)$ is a group under multiplication, because if f, g are continuous eigenfunctions with eigenvalues λ, ρ, respectively, then the pointwise product $f \cdot g$ is a continuous eigenfunction with eigenvalue $\lambda\rho$.

Theorem 2.14 (Halmos-von Neumann)**.** *Two equicontinuous minimal systems are isomorphic if and only if they have the same discrete spectrum.*

For the proof we require some facts the theory of topological groups. If G is a compact abelian group, then $\chi : G \to \{z \in \mathbb{C} \,:\, |z| = 1\}$ is called a **character** if it is a group homomorphism (with the multiplication operation in the range), so that $\chi(gh) = \chi(g) \cdot \chi(h)$ for all $h, g \in G$, and in addition χ is continuous. Characters do not have to exist in general groups, e.g. a simple non-abelian group will not have any. But they are abundant in compact abelian groups:

Theorem (Pontryagin)**.** *Let G be a compact metrizable abelian group. Then there are countably many characters and they separate points in $C(G)$.*

For example in $\mathbb{R}/\mathbb{Z}$, the characters are the functions $\{\varphi_n\}_{n\in\mathbb{Z}}$ given by $\varphi_n(t) = e^{2\pi i n t}$, and in $\mathbb{R}^d/\mathbb{Z}^d$ the characters are $\{\varphi_v\}_{v\in\mathbb{Z}^d}$ given by $\varphi_v(t) = e^{2\pi i \langle v,t\rangle}$, where $\langle\cdot,\cdot\rangle$ is the standard inner product.

Corollary 2.15. *If (X, R_g) is a group rotation, then every character is a continuous eigenfunction, and if R_g acts minimally then every eigenfunction is proportional to a character. In particular the continuous eigenfunctions separate points and, up to multiplication by a scalar, there are countably many of them.*

Proof. Every character $\chi : X \to \mathbb{C}$ is an eigenfunction with eigenvalue $\chi(g)$, because

$$\chi(R_g x) = \chi(gx) = \chi(g)\chi(x)$$

Conversely, suppose R_g acts minimally and let $f \in C(X)$ be an eigenfunction with eigenvalue λ. By multiplying f by a scalar we can assume that $f(e) = 1$, where $e \in X$ is the identity element.

Let $h_!, h_2 \in X$. By minimality, $g^{n_k} \to h_1$ for some $n_k \to \infty$. Therefore

$$\begin{aligned} f(h_1 h_2) &= \lim_{k\to\infty} f(g^{n_k} h_2) \\ &= f(h_2) \lim_{k\to\infty} \lambda^{n_k} \\ &= f(h_2) \lim_{k\to\infty} \lambda^{n_k} f(e) \\ &= f(h_2) \lim_{k\to\infty} f(g^{n_k} e) \\ &= f(h_1) f(h_2) \end{aligned}$$

so f is a character. □

Thus, for example, in $(\mathbb{R}/\mathbb{Z}, R_\alpha)$ the eigenfunctions are the scalar multiples of $\varphi_n(t) = e^{2\pi int}$, and the eigenvalue of φ_n is $e^{2\pi in\alpha}$, so $\Sigma(\mathbb{R}/\mathbb{Z}, R_\alpha) = \{e^{2\pi in\alpha}\}_{n\in\mathbb{Z}}$.

Proof of the Halmos-von Neumann Theorem. Since minimal equicontinuous systems are isomorphic to group translations, we may assume that the systems have the form (G, R_g) and (H, R_h) for compact abelian groups G, H.

Write $\{\lambda_i\}_{i\in I} = \Sigma(G, R_g) = \Sigma(H, R_h)$, and define

$$\Gamma = \prod_{I\in I} S^1$$

The product is countable so this is a compact metrizable group under pointwise multiplication. Let

$$\gamma = (\lambda_i)_{i\in I}$$

so R_γ acts on Γ by

$$R_\gamma(t_1, t_2, \ldots) = (\lambda_1 t_1, \lambda_2 t_2, \lambda_3 t_3, \ldots) \in \Gamma$$

This is a continuous map, and the system (Γ, L) is a group translation.

Let $f_i : G \to \mathbb{C}$ be continuous eigenfunctions associated to λ_i such that $\{f_i\}$ separate points, and define the map $\pi : G \to \Gamma$ by

$$\pi(x) = (f_i(x))_{i\in I}$$

Then π is continuous, it is injective because $\{f_i\}$ separate points, and it is an isomorphism to its image because

$$\begin{aligned} \pi(R_g h) &= (f_1(R_g h), f_2(R_g h), \ldots) \\ &= (\lambda_1 f_1(h).\lambda_2 f_2(H), \ldots) \\ &= L(f_1(h), f_2(h), \ldots) \\ &= L\pi(h) \end{aligned}$$

Similarly, let $g_i : H \to \mathbb{C}$ be corresponding eigenfunctions for λ_i and let $\sigma : H \to \prod_{i\in I} S^1$ denote the map

$$\sigma(y) = (g_i(y))_{i\in I}$$

By the same considerations, $\sigma : H \to \sigma(H)$ is an isomorphism to $(\sigma(H), L)$.

It remains to show that $(\sigma(G), L) \cong (\sigma(H), L)$. Indeed, $\sigma(H), \sigma(G)$ are minimal subsets of (Γ, L), so by Proposition 2.10, they are both cosets of the compact group $\Lambda = \overline{O_L(e)}$, say $\sigma(G) = a\Lambda$ and $\sigma(G) = b\Lambda$. Therefore, $R_{ba^{-1}}$ is a homeomorphism of $\sigma(G) \to \sigma(H)$, and it commutes with L because both are group translation in an abelian group. This is the desired isomorphism of $(\sigma(G), L)$ and $(\sigma(H), L)$, and completes the proof. $\square$

Thus, if $\alpha, \beta \in \mathbb{R} \setminus \mathbb{Q}$, we have $\Sigma(\mathbb{R}/\mathbb{Z}, R_\alpha) = \{e^{2\pi i n\alpha}\}_{n\in\mathbb{Z}}$ and $\Sigma(\mathbb{R}/\mathbb{Z}, R_\beta) = \{e^{2\pi i n\beta}\}_{n\in\mathbb{Z}}$. If $\alpha \neq \pm\beta$, these sets are distinct (since $\alpha\mathbb{Z}, \beta\mathbb{Z}$ are distinct). We conclude that there are uncountably many non-isomorphic minimal rotations of $\mathbb{R}/\mathbb{Z}$.

2.4 Bohr almost periodic functions

As an application of the theory we have developed, we briefly discuss Bohr's notion of an almost-periodic function, and show how one gets a representation for all almost-periodic functions.

Let $\ell^\infty(\mathbb{Z})$ denote the vector space space of bounded functions $\mathbb{Z} \to \mathbb{C}$ with the sup norm.[4]

Definition 2.16. A sequence $a \in \ell^\infty(\mathbb{Z})$ is called **almost periodic** (in the sense of Bohr) if, for every $\varepsilon > 0$, the set

$$\{n \in \mathbb{Z} \ : \ |a_{k+n} - a_k| < \varepsilon \text{ for all } k \in \mathbb{Z}\}$$

is syndetic.

Clearly every periodic sequence is almost periodic but there non-periodic examples, e.g. write

$$e(t) = \exp(2\pi i t)$$

and let

$$a_n = e(\alpha n)$$

for some $\alpha \notin \mathbb{Q}$. Then $a = (a_n)$ is almost periodic. Indeed, given $\varepsilon > 0$ choose δ so that $|e(t) - 1| < \delta$ whenever $|t| < \varepsilon$. Suppose that $n \in \mathbb{Z}$ is such that $n\varepsilon \in (m, m + \delta)$ for some $m \in \mathbb{N}$; this set is syndetic because 0 is recurrent under R_ε in $\mathbb{R}/\mathbb{Z}$. For any such n,

$$e(\alpha n) = \exp(2\pi i \alpha n) = \exp(2\pi i (m + \delta)) = \exp(2\pi i \delta)$$

hence $|a_n - 1| < \delta$ by choice of δ. Now, for any k,

$$\begin{aligned} |a_{k+n} - a_k| &= |e(\alpha(k+n)) - e(\alpha k)| \\ &= |e(\alpha k) e(\alpha n) - e(\alpha k)| \\ &= |e(\alpha k)| \cdot |e(\alpha n) - 1| \\ &< \varepsilon \end{aligned}$$

More generally, one can show that every trigonometric polynomial is almost periodic, i.e. every $a \in \ell^\infty$ of the form $a_n = \sum_{\ell=1}^{L} c_\ell e(\alpha_\ell n)$ for $c_\ell \in \mathbb{C}$, $\alpha_\ell \in \mathbb{R}$. For this we use recurrence of translation by $(\alpha_1, \ldots, \alpha_\ell)$ in $\mathbb{R}^\ell/\mathbb{Z}^\ell$, which holds because every point in a compact group is recurrent for any translation. Finally, any uniform limit of almost periodic functions is almost periodic.

[4]Almost everything we do here transfers with minimal changes to the space of bounded continuous functions on $\mathbb{R}$, with the same definition of almost periodicity.

The question now arises whether there are any almost periodic functions besides limits of trigonometric polynomials. We shall not prove a theorem of Bohr, which states that the answer is negative.

We can define an isometry S on $\ell^\infty(\mathbb{Z})$ by shifting:

$$(Sa)_n = a_{n+1}$$

Then $a \in \ell^\infty(\mathbb{Z})$ is almost periodic if and only if it is a uniformly recurrent point with respect to the map S. Of course, $\ell^\infty(\mathbb{Z})$ is far from being compact. However,

Proposition 2.17. *If $a \in \ell^\infty(\mathbb{Z})$ is almost periodic, then $\overline{O_S(a)}$ is minimal.*

Proof. Suppose $a \in \ell^\infty(\mathbb{Z})$ is almost periodic. We first show that $O_S(a)$ is totally bounded, so that $\overline{O_S(a)}$ is compact. Indeed, let $\varepsilon > 0$. let M be such that

$$\{n \in \mathbb{Z} \mid \|S^n a - a\|_\infty < \varepsilon\}$$

intersects every interval of length M (such M exists because by assumption the set above is syndetic). We claim that

$$O_S(a) \subseteq \bigcup_{i=0}^{M-1} T^i B_\varepsilon(a)$$

Indeed consider some $S^k a$. There exists $n \in (k - M, k]$ such that $\|S^n a - a\|_\infty < \varepsilon$. Hence (because S^{k-n} is an isometry) $\left\|S^k a - S^{k-n} a\right\|_\infty < \varepsilon$, or equivalently, $S^k a \in B_\varepsilon(S^{k-n} a)$ and $0 \le k - n < M$.

Now $X = \overline{O_S(a)}$ is compact, S acts on it as an isometry, and $a \in X$ has dense orbit. So by Proposition 2.6, $(\overline{O_S(a)}, S)$ is minimal. □

Theorem 2.18. *If $a \in \ell^\infty(\mathbb{Z})$ is an almost periodic function, then a is the limit in ℓ^∞ of trigonometric polynomials.*

Proof. Let a be almost periodic and let $G = \overline{O_S(a)}$. This is a minimal set and S is an isomtery with respect to $\|\cdot\|_\infty$, so G can be give a group operation $\cdot$ under which $Sb = h \cdot g$ for some $h \in G$ and all $g \in G$.

Consider the function $\pi : \ell^\infty(\mathbb{Z}) \to \mathbb{C}$ given by $\pi(b) = b_0$. This is a continuous function and restricts to a continuous function on G, so given $\varepsilon > 0$, there exist $d_n \in \mathbb{C}$ and characters $\chi_1, \chi_2, \ldots, \chi_N \in \Sigma(G)$ such that $\left\|\pi - \sum_{n=1}^N d_n \chi_n\right\|_\infty < \varepsilon$.

Next, note that

$$\begin{aligned} a_n &= \pi(S^n a) \\ &= \pi(h^n \cdot a) \\ &= \sum_{i=1}^N d_i \chi_i(h)^n \chi_i(a) \pm \varepsilon \end{aligned}$$

Writing $c_i = d_i \chi_i(n)$ and defining α_i so that $\chi_i(a) = e(\alpha_i)$ (recall that $|\chi_i| = 1$), we have $\|a - \sum c_i e(\alpha_i)\|_\infty < \varepsilon$, which shows that a is in the closure of trigonometric the polynomials. □

2.5 Problems

In the following questions (X, d) is a compact metric space.

1. Show that the definition of equicontinuity is independent of the metric. That is, if d_1, d_2 are equivalent metrics, then equicontinuity with respect to a metric d_1 is equivalent to equicontinuity with respect to d_2.

2. Give an example of a system which is not equicontinuous.

3. Assuming that T is invertible, show that (X, T) is equicontinuous if and only if (X, T^n) is.

4. Let (X, T) be an equicontinuous system and d a metric on X. Show directly from the definition that for every $x, y \in X$ with $x \neq y$,

$$\liminf_{n \to \infty} d(T^n x, T^n y) > 0$$

5. Let $D \subseteq \mathbb{C}$ denote the closed unit disc. Let $T : D \to D$ denote the map that rotates the circle of radius r by angle r: that is, $T(re^{i\theta}) = re^{i(\theta + r)}$. Show that this system decomposes into disjoint minimal systems, but is not equicontinuous.

6. Let $Tx = 10x \bmod 1$. Show that every subsystem $([0, 1), T)$ such that $(X, T|_X)$ is equicontinuous, must be finite.

3 Expansive and symbolic systems

We now pivot to a class of systems of a completely different nature – those which are, from the metric point of view, far from equicontinuity. We will focus on symbolic systems, which besides being a rich source of examples, are important as "models" for other, less accessible dynamical systems. One particular outcome of this section is that there exist minimal systems which are not equicontinuous.

3.1 Expansive systems

Definition 3.1. A system (X, T) is **forward expansive** if there exists an $\varepsilon > 0$ such that for every $x, y \in X$ with $x \neq y$, there is an $n \in \mathbb{N}$ such that $d(T^n x, T^n y) > \varepsilon$. If T is invertible we say it is **two-sided expansive** the same holds but allowing $n \in \mathbb{Z}$. The constant ε is called the expansiveness constant.

This definition excludes equicontinuity except in the most trivial cases. Indeed, if (X, T) is equicontinuous and expansive, let $\varepsilon > 0$ be the expansivity constant. Then there is a $\delta > 0$ such that $d(x, y) < \delta$ implies $d(T^n x, T^n y) < \varepsilon$ for all n. It follows that there are no distinct pairs x, y with $d(X, y) < \delta$, so every point in X is isolated, and by compactness, X is finite.

Remarks and examples

1. The definition uses the metric, the property of expansiveness is independent of the metric, although the constant ε changes.
2. Clearly a subsystem of an expansive system is expansive (with the same constant).
3. Given an integer $a \geq 2$, the system $\mathbb{R}/\mathbb{Z} \cong [0, 1)$ with the map $x \mapsto ax \bmod 1$ is expansive.
4. If A is a $d \times d$ integer matrix acting on $\mathbb{T}^d$ by $x \mapsto Ax \bmod 1$, then A is expansive if and only if all eigenvalues are of modulus > 1. If $\det A = 1$ then A^{-1} is also integral and so the action on the torus is invertible; in this case two-sided expansiveness will hold if all eigenvalues are not of unit modulus (such a matrix is called **hyperbolic**).

Perhaps the most important example, however, is that of a symbolic system.

3.2 Symbolic systems

Let Λ be a finite set, write

$$\Lambda^* = \bigcup_{n=0}^{\infty} \Lambda^n$$

for the set of finite words in the language Λ, including the empty word. We denote one- and two-sided infinite sequences by $(x_1 x_2 \ldots)$ or $(\ldots, x_{-1} x_0 | x_1 x_2 \ldots)$, respectively, where $|$ separates the 0th and 1st coordinates.

Topology

The spaces of all such sequences are the product spaces $\Lambda^{\mathbb{N}}$ and $\Lambda^{\mathbb{Z}}$; thinking of Λ as a discrete topological space, $\Lambda^{\mathbb{N}}$ and $\Lambda^{\mathbb{Z}}$ carry the product topology, which is compact in the product topology. This is the topology generated by the **cylinder sets**, by which we mean sets $[a]_k$ defined by fixing $a \in \Lambda^*$, $k \in \mathbb{Z}$ and

$$[a]_k = \{y \,:\, a \text{ appears in } y \text{ at index } k\}$$

We abbreviate $[a] = [a]_1$. These sets are open and closed, and the product spaces are compact by Tychonoff's theorem.

Metric

One can also introduce a metric inducing this topology: for two sequences x, y, set

$$|x \wedge y| = \min\{|n| \,:\, x_n \neq y_n\}$$

and[5]

$$d(x,y) = 2^{-|x \wedge y|}$$

The closed ball of radius 2^{-N} centered at x is precisely the cylinder $[x_1, \ldots, x_N]_1$ in the one-sided case, and $[x_{-N}, \ldots, x_N]_{-N}$ in the two-sided case. Every cylinder is a finite union of cylinders of this form, so the metric induces theproduct topology.

Convergence

Convergent sequences can be described as follows. Say that x, y agree on their first N symbols if they agree at all coordinates $1 \leq i \leq N$ in the one-sided case, or $-N \leq i \leq N$ in the two-sided case. This is equivalent to $d(x,y) \leq 2^{-N}$. Then $x^{(n)} \to x$ if and only if for every N, the points $x^{(n)}, x$ agree on their first N symbols as soon as n is large enough (depending on N).

Shift map

Next, define the **shift map** S on $\Lambda^{\mathbb{N}}$ and $\Lambda^{\mathbb{Z}}$ by

$$(Sx)_i = x_{i+1}$$

Note that the shift map is continuous: if $x \neq y$, then $|Sx \wedge Sy| \geq |x \wedge y| - 1$, hence $d(Sx, Sy) \leq 2d(x,y)$. Also, S is onto, and it is bijective in the two-sided case, with inverse given by

$$(S^{-1}x)_i = x_{i-1}$$

In the one-sided case it is not injective, since $S(abbb\ldots) = S(bbb\ldots)$ for every $a \neq b$ in Λ. In fact, for every $x \in \Lambda^{\mathbb{N}}$ we have $S^{-1}x = \{(ax_1x_2\ldots) \mid a \in \Lambda\}$, and in particular, $|S^{-1}x| = \Lambda$.

[5] We remark that the choice of the base 2 in the definition of the metric is arbitrary, and if one uses any other constant greater than one an equivalent metric is obtained.

Symbolic systems

Definition 3.2. Let Λ be a finite set. The systems $(\Lambda^{\mathbb{N}}, S)$ or $(\Lambda^{\mathbb{Z}}, S)$ are called the **one-sided full shift** and the **two-sided full shift**, respectively.

A **subshift** of **symbolic system** is a subsystem of a full shift. It is one-sided or two-sided, depending on the full shift in question.

The full shift is expansive with constant $\varepsilon = 1$ in the metric we have chosen: if $x \neq y$, then $x_n \neq y_n$ for some n, and than $(S^n x)_0 = x_n \neq y_n = (S^n y)_0$, so $d(S^n x, S^n y) = 1$.

As examples of subshifts we have trivial examples: the full shift itself, and periodic orbits. In fact, the subshifts form an extremely rich class of dynamical systems, and will soon see a more interesting example.

Language of a subshift

Let us now re-interpret the dynamical properties we have encountered in the context of symbolic systems. If X is a subshift, then its **language** is

$$\begin{aligned} L(X) &= \{a \in \Lambda^* \ : \ X \cap [a] \neq \emptyset\} \\ &= \{a \in \Lambda^* \ : \ a \text{ appears in some sequence } x \in X\} \end{aligned}$$

If X is one-sided, then $L(X)$ determines X. For suppose $X \neq Y$ are subshifts. Then either there exists $x \in X \setminus Y$ or there exists $y \in Y \setminus X$. In the first case, suppose that $x_1 \dots x_n \in L(Y)$ for all n. Then there are points in Y containing $x_1 \dots x_n$, and by applying a shift we can assume the occurrence is at the first coordinate. Then these points converge to x, so $x \in Y$, contrary to assumption. Thus there must be some n with $x_1 \dots x_n \in L(X) \setminus L(Y)$, hence $L(X) \neq L(Y)$. If we had $y \in Y \setminus X$ we would argue similarly.

In the two-sided case, $L(X)$ determines the subshift X provided we assume $S(X) = X$. The argument is the same as above using $x_{-n} \dots x_n$; surjectibity is needed in order to shift occurences of this word in Y back to their original position.

We have the following dictionary between dynamical properties of x and languages:

- $x \in \Lambda^{\mathbb{Z}}$ is recurrent if and only if for every n there are infinitely many occurrences of $x_{-n} \dots x_n$ at positive coordinates in x. In the one-sided case, one demands occurrences of $x_1 \dots x_n$.

  Indeed, recurrence means that for every open set U containing x, there exist arbitrarily large n with $T^n x \in U$. Since the cylinders form a basis for the open sets, it suffices to consider cylinders containing x, and therefore cylinders of the form $U = [x_{-k} \dots x_k]_{-k}$. Then $T^n x \in U$ just means that $x_{-k} \dots x_k$ occurs in x at position n.

- If $x \in \Lambda^{\mathbb{Z}}$ then $\overline{O_S(x)}$ is the unique subshift X such that $L(X) = L(x)$ (In the one-sided case $x \in \Lambda^{\mathbb{N}}$, use one-sided orbits).

  Indeed, if $y \in \overline{O_S^{\pm}(x)}$ then y is a limit of points $S^n x$, so every word in y appears in $S^n x$ at the same position for some large n. hence in x, at another position. This gives $L(\overline{O_S^{\pm}(x)}) \subseteq L(x)$, and the reverse inclusion is obvious. Uniqueness follows since $L(X)$ determines X.

  For instance, this shows that the full-shift is transitive, since if x is a sequence that contains all finite words (e.g. a concatenation of all words) then $L(x) = L(\Lambda^{\mathbb{Z}})$.

- Notice that for any $a \in \Lambda^*$,

$$N(x,[a]) = \{n \,:\, x_n \dots x_{n+|a|-1} = a\}$$

Thus, $x \in \Lambda^{\mathbb{Z}}$ is uniformly recurrent if and only if every $a \in L(x)$ appears in x syndetically, and a subshift $X \subseteq \Lambda^{\mathbb{Z}}$ is minimal if and only if every $a \in L(X)$ appears syndetically in every $x \in X$.

Proposition 3.3. *If $X \subseteq \Lambda^{\mathbb{Z}}$ is a subshift then there exists a unique subshift $Y \subseteq \Lambda^{\mathbb{N}}$ such that $L(Y) = L(X)$.*

Conversely, for every subshift $Y \subseteq \Lambda^{\mathbb{N}}$ such that $SY = Y$, there exists a unique $X \subseteq \Lambda^{\mathbb{Z}}$ such that $L(X) = L(Y)$.

Proof. Given X. For $x \in X$ let $\pi(x) = (x_1, x_2, \dots) \in \Lambda^{\mathbb{N}}$ and

$$Y = \pi(X) = \{(x_1, x_2, \dots) \,:\, x \in X\}$$

Since π is continuous and X is compact, Y is compact. It is a subshift because if $y \in Y$ then there exists a choice of negative coordinates y_i, $i \leq 0$, such that $x = (\dots y_{-1}y_0|y_1y_2y_3\dots) \in X$. Then $Sx = (\dots y_1|y_2y_3\dots) \in X$ so $(y_2y_3\dots) = \pi(Sx) \in Y$. Finally, it is clear that any word appearing in Y appears in X, and if a word $a \in \Lambda^*$ appears in some $x \in X$ at coordinate i then it appears in $x' = S^{-|i|+1}x$ at a positive coordinate, so it appears in $\pi x' \in Y$.

In the opposite direction, let Y be given and let $\widetilde{X} = \pi^{-1}Y$. Let $X = \bigcap_{n=1}^{\infty} S^n\widetilde{X}$. One may then verify that X is closed, shift invariant and $L(X) = L(Y)$. $\square$

3.3 Minimal substitution systems

We can now give non-trivial examples of minimal subshifts that are not equicontinuous. A **substitution** is a map $\sigma : \Lambda \to \Lambda^*$. We extend σ to Λ^* pointwise: $\sigma(a_1 \dots a_n) = \sigma(a_1)\dots\sigma(a_n)$, and also to $\Lambda^{\mathbb{N}}$ using the same formula. The extension to $\Lambda^{\mathbb{N}}$ is continuous, provided $|\sigma(a)| > 1$ for all $a \in \Lambda$. A substitution is called **primitive** if $\sigma(a)$ contains all symbols of Λ for every $a \in \Lambda$.

For example, let $\Lambda = \{0, 1\}$ and let

$$\begin{aligned} \sigma(0) &= 01 \\ \sigma(1) &= 10 \end{aligned} \tag{1}$$

Then

$$\sigma(0110) = 01\ 10\ 10\ 01$$

(the spaces were inserted for emphasis only). This example is known as the Thue-Morse substitution.

Lemma 3.4. *Suppose that σ is substitution and that $\sigma(a)$ begins with a for some $a \in \Lambda$. Then $\sigma^n(a)$ is a prefix of $\sigma^{n+1}(a)$. In particular there is a point $x \in \Lambda^{\mathbb{N}}$ such that $\sigma^n(a)$ is a prefix of x for all n.*

Remark 3.5. Even if no symbol $a \in \Lambda$ exists as in the proposition, by replacing σ by σ^n (for some $n \leq |\Lambda|$) we will be able to find such a symbol.

Proof. By induction. □

The example above is of the type in the lemma with $a = 0$. Thus

$$\begin{aligned} \sigma(0) &= 01 \\ \sigma(01) &= 0110 \\ \sigma(0110) &= 01101001 \\ \sigma(01101001) &= 0110100110010110 \end{aligned}$$

etc. The limiting sequence x is called the **Thue-Morse sequence**.

Proposition 3.6. *Suppose that σ is primitive and that x is as in Lemma ??. Then x is uniformly recurrent.*

Proof. We first claim that x can be written as $x = \sigma^k(x) = \sigma^k(x_1)\sigma^k(x_2)\ldots$. Indeed $\sigma^n(a) = \sigma^k(\sigma^{n-k}(a))$, holds for all $n \geq k$, and both $\sigma^n(a)$ and $\sigma^{n-k}(a)$ are prefixes of x. Therefore σ transforms arbitrarily long prefixes of x into prefixes of x. The conclusion follows.

Taking $k = 1$, and using the fact that σ is primitive, we find that every symbol appears syndetically in x.

Now fix any $w \in L(x)$, so w it appears in $\sigma^k(a)$ for some k. Let M denote the maximal length of $\sigma^k(b)$, $b \in \Lambda$. We saw above that a appears in x syndetically, with gaps of at most N for some $N \geq 0$. Since $x = \sigma^k(x)$, the word $\sigma^k(a)$ appears in x separated by at most N words of the form $\sigma^k(b)$; so the gaps between occurrences of $\sigma^k(a)$, and hence between occurrences of w, are at most MN. □

Returning to the Thue-Morse sequence x, the orbit closure $\overline{O_S(x)}$ is an infinite minimal subshift. We now claim that it is not periodic:

Proposition 3.7. *The Thue-Morse sequence x is not periodic, and $\overline{O_S(x)}$ is infinite.*

Proof. The two statements are equivalent (since the orbit closure is minimal). We prove the first..

An easy induction shows that $\sigma^n(0)$ ends alternately in 0 (for odd n) and 1 (for even n). Also, observe that 00 appears in $\sigma^2(0)$. Because $\sigma^n(00) = \sigma^n(0)\sigma^n(0)$, the second occurrence of $\sigma^n(0)$ is preceded alternately by 0 and 1. But $\sigma^n(0)$ also begins with $\sigma^{n-1}(0)$. Thus, we find $1\sigma^k(0)$ in $\sigma^{2k}(00)$ and we find $0\sigma^k(0)$ in $\sigma^{2k+1}(00)$.

Summarizing, we can find arbitrarily long words b in x such that both $0b$ and $1b$ appear in x. Therefore, x is not periodic. □

In conclusion we have shown for σ as in (1), the subshift $X_\sigma \subseteq \{0,1\}^{\mathbb{N}}$ is infinite and minimal. The two sided subshift $X_\sigma \in \Lambda^{\mathbb{Z}}$ with the same language $L(X) = L(x)$ is also infinite and minimal. It is called the **substitution system** generated by σ.

3.4 When does the past predict the future?

For a point $x \in \Lambda^{\mathbb{Z}}$ write

$$x^- = (\ldots, x_{-2}, x_{-1}, x_0) \in \Lambda^{-\mathbb{N}\cup\{0\}}$$

and for a two-sided subshift $X \subseteq \Lambda^{\mathbb{Z}}$ write

$$X^- = \{x^- \,:\, x \in X\}$$

We say that $x \in X$ extends $y \in X^-$ if $y = x^{-1}$.

Proposition 3.8. *Let $X \subseteq \Lambda^{\mathbb{Z}}$ be a two-sided subshift X. Then every $z \in X^-$ extends to a unique $x \in X$ if and only if X is finite.*

Proof. A finite two-sided subshift is a finite union of periodic orbits. In this case the unique extension property is immediate, since a periodic sequence in $\Lambda^{-\mathbb{N}}$ extends uniquely to a periodic sequence in $\Lambda^{\mathbb{Z}}$.

For the converse, suppose every $z \in X^-$ extends uniquely to a point in X. We claim that then there is an $n \in \mathbb{N}$ such that $x_{-n}, \ldots, x_0$ determines x_1 for all $x \in X$. Otherwise, for every n there is a word $a_n \in L_n(X)$ and distinct symbols $u_n, v_n \in \Lambda$ such that $a_n u_n, a_n v_n \in L_{n+1}(X)$. Therefore, there are words $x^{(n)}, y^{(n)} \in X$ such that $x^{(n)}|_{[-n+1,1]} = a_n u_n$ and $y^{(n)}|_{[-n+1,1]} = a_n v_n$. By compactness, we can choose a subsequence $n(k)$ such that $x^{(n(k))} \to x$ and $y^{(n(k))} \to y$. Then $x^- = y^-$ but $x_1 \neq y_1$, contrary to the unique extension assumption.

Now given n as above, we claim that $|X| \leq |\Lambda|^n$. Indeed, it is enough to show that $L_N(X) \leq |\Lambda|^n$ for all N. This follows from the fact that every $a \in L_N(X)$ is determined by it initial n symbols, because once these are known, the subsequence symbols are determined one by one. □

Define X^+ analogously to X^-. Then:

Corollary 3.9. *Every $z \in X^-$ extends uniquely to $x \in X$ if and only if every $w \in X^+$ extends uniquely to $x \in X$.*

In a dynamical system (X, T) a pair of points $x, x' \in X$ is called **forward asymptotic** if $d(T^n x, T^n x') \to 0$, and **backward asymptotic** if the same holds with $n \to -\infty$. Proposition 3.8 can be rephrased as saying that in an infinite two-sided subshift, there always exist asymptotic pairs; indeed if $x^- = y^-$ then x, y are backward asymptotic and conversely if x, y are backward asymptotic then one easily checks that $x_n = y_n$ for all sufficiently negative n. Of course, we also have a result in the positive direction.

In this language, there is an analogous result for expansive systems in general:

Proposition 3.10. *If (X, T) is an invertible, expansive and infinite system, then it contains forward and backward asymptotic pairs.*

The proof appears as a guided exercise at the end of this section.

We end this section with a surprising fact: Although in general the past does not predict the future in a minimal symbolic system, there is always a large set of points where this does hold.

Proposition 3.11. *Let $X \subseteq \Lambda^{\mathbb{Z}}$ be a minimal symbolic system. Then there is a dense G_δ set $P \subseteq X$ such that for every $x \in P$, the sequence x is the only extension x^- to a point in X.*

Proof. We first prove a more modest claim: There is a dense open set $P_0 \subseteq X$, such that for $x \in P_0$ and $y = x^-$, there is a unique $a \in \Lambda$ such that ya appear in X.

Let $w \in L(X)$ be any word and write w' for the word obtained by deleting the last symbol of w. Since w appears in X with bounded gaps, there exists a maximal gap; let $wvw \in L(X)$ realize this gap, so wvw is the longest word in X such that w appears in wvw only as written.

Now let $b = b(w) = wvw'$. The only way to extend b to a word in $L(X)$ is to extend it to wvw, because any other extension leads to a point in X with a too-large gap between ws. Now set

$$P_0 = \bigcup_{w \in L(X)} [b(w)]_{-\ell(b(w))}$$

This is an open set, and it is dense because the words w' at the end of $b(w)$ is arbitrary.

Now let

$$P = \bigcap_{n=0}^{\infty} S^{-n} P_0$$

If $x \in P$ then $x \in P_0$, so x^- determines x_0. Now $x \in S^{-1}P_0$, so $Sx \in P_0$ meaning that $\ldots x_{-1}x_0x_1$ determines x_2; and so on.

Finally, since P_0 is open and dense, so are all of its shifts $S^{-n}P_0$; so P is a dense G_δ set by Baire's category theorem. □

A word of warning: a dense G_δ set is large from a topological point of view, but may be small in other ways, e.g. there can well be an invariant measure on X for which there is no set of positive measure with the property of P.

3.5 Problems

1. Show that $L \subseteq \Lambda^*$ is the language of a two-sided subshift if and only if it satisfies the following properties:

   Closed to subwords: If $a \in L$ and b is a subword of a then $b \in L$.

   Extensibility: For every $a \in L$ there exist $u, v \in \Lambda$ such that $uav \in L$.

2. Show that the condition that S is onto is necessary in Proposition 3.3.

3. Construct a word $x \in \{0,1\}^{\mathbb{N}}$ inductively. Start with the word $x_1x_2x_3 = 111$, and assuming we have defined $x_1 \ldots x_{3^n}$, define $x_{3^n+1} = \ldots = x_{2\cdot 3^n} = 0$ and $x_{2\cdot 3^n+1} \ldots x_{3^{n+1}} = x_1 \ldots x_{3^n}$.

   (a) Show that x is recurrent, but not uniformly recurrent.

   (b) Describe all the minimal subsystems of $\overline{O_S(x)}$.

4. Prove the statement in Remark 3.5.

5. Show that every two-sided expansive system (X, T) contains an asymptotic pair.

   (a) Prove that for every $\delta > 0$ there are points $x \neq y$ with $d(x, y) \geq \delta$ and $d(T^n x, T^n y) \leq \delta$ for $n \geq 1$.

   (b) Let $\varepsilon > 0$ be the expansivity constant of the system. Show that for every $0 < \delta < \varepsilon/2$ there is an $N = N(\delta)$ such that if x, y are as in (a), then there exists $0 \leq n \leq N$ such that $d(T^{-n}x, T^{-n}y) \geq 2\delta$.

   (c) Now consider points x_n, y_n as in (a) for $\delta = \varepsilon \cdot 2^{-n}$ and $n = 1, 2, 3 \ldots$. Show that for some $k(n)$ the points $T^{-k(n)}x_n$ and $T^{-k(n)}y_n$ will converge to an asymptotic pair.

4 The Enveloping semigroup

We return now to the study dynamical systems in general and minimal systems in particular. The tool we introduce here is the Ellis semigroup, which is the closure of $\{T^n\}_{n=0}^{\infty}$ in a suitable topology. We have already seen this idea at work in our study of equicontinuous systems, where we took the closure in the uniform topology, and obtained a compact group, on which translation by T mirrored the original dynamics. However, the uniform closure is compact **only** when the system is equicontinuous, so in order to go any further we must weaken the topology. This produces quite different, and often odd, results, but turns out to be a very powerful tool, which we will need later on. As an immediate application, we will derive a combinatorial result due to Furstenberg on a class of subsets of the integers with additive structure.

4.1 The enveloping semigroup

Let X be a compact metric space. The set of all functions $X \to X$ is identified as the product set X^X, which carries the product topology. A basis of open sets for this topology is given by sets of the form $U_{x_1,\ldots,x_n,\varepsilon}(f)$ where $f : X \to X$, $x_1, \ldots, x_n \in X$, $\varepsilon > 0$, and

$$U_{x_1,\ldots,x_n,\varepsilon}(f) = \{g : X \to X \ : \ d(f(x_i), g(x_i)) < \varepsilon \text{ for } i = 1, \ldots, n\}$$

This topology is not metrizable unless X is countable; but by Tychonoff's theorem, it is compact.

Definition 4.1. The **enveloping semigroup** (or Ellis semigroup) $\mathcal{E} = \mathcal{E}(X,T)$ of a dynamical system (X,T) is the closure of $\{T^n\}_{n\geq 0}$ in X^X, with the operation of composition of functions.

If $f \in \mathcal{E}$ then $f(x) \in \overline{O_T(x)}$ for every $x \in X$, since for every k there is an n_k such that $d(T^{n_k}x, f(x)) < 1/k$ and thus $T^{n_k}x \to f(x)$.

More generally, for any $x_1, \ldots, x_n \in X$, there is a sequence (n_k) such that $f(x_i) = \lim T^{n_k}x_i$.

Thus, an element in $\mathcal{E}$ corresponds to a choice of a point in the orbit closure for any initial point, in a way that the times along which the orbit converges to the point is compatible across all finite choices of initial points. Note that it is not possible to choose a single (n_k) that works for all points, and the existence of non-trivial elements in $\mathcal{E}$ generally requires the axiom of choice (for more discussion, see problems at the end of this section).

Lemma 4.2. *Let $\mathcal{E} = \mathcal{E}(X,T)$.*

1. *$T^n \in \mathcal{E}$ for every $n \geq 0$.*
2. *For every $x \in X$ the map $f \mapsto fx$ is a continuous map on $\mathcal{E}$.*
3. *For every $g \in \mathcal{E}$, the map $f \mapsto f \circ g$ is continuous on $\mathcal{E}$.*
4. *If $g \in \mathcal{E}$ is continuous, then the map $f \mapsto g \circ f$ is continuous on $\mathcal{E}$.*
5. *$\mathcal{E}$ is a semigroup (under composition).*

Proof. (1) is obvious. Also (2) is immediate from the definition of the topology.

Properties (3) and (4) hold in the larger semigroup X^X of all functions $X \to X$. Indeed, continuity of $f \mapsto f \circ g$ follows from the identity

$$h \in U_{g(x_1),\ldots,g(x_n),\varepsilon}(f) \qquad \Longleftrightarrow \qquad h \circ g \in U_{x_!,\ldots,x_n,\varepsilon}(f \circ g)$$

If g is continuous, assuming $d(y_1, y_2) < \delta$ implies $d(g(y_1), g(y_2)) < \varepsilon$, we conclude that

$$h \in U_{x_1,\ldots,x_n,\delta}(f) \qquad \Longrightarrow \qquad g \circ h \in U_{x_1,\ldots,x_n,\varepsilon}(g \circ f)$$

hence $f \mapsto g \circ f$ is continuous.

Finally we prove (5): Since $\{T^n\}_{n\geq 0} \subseteq \mathcal{E}$ and the family $\{T^n\}_{n\geq 0}$ is closed under post-composition with T, and since T is continuous, (4) implies that $\overline{\{T^n\}}_{n\geq 0}$ is also closed to post-composition with T. Thus for any $f \in \mathcal{E}$ also $T^n f \in \mathcal{E}$ for all $n \geq 0$, and so by (3), we have $\mathcal{E}f \subseteq \mathcal{E}$. □

Note that even when T is invertible, $\mathcal{E}$ can contain non-invertible maps. For example, if $x, y \in X$ are forward asymptotic, then there will exist $f \in \mathcal{E}$ such that $f(x) = f(y)$ – just take any accumulation point of $\{T^n\}$.

4.2 Ideals, idempotents and minimal points

There is a close connection between algebraic structures in $\mathcal{E}$ and the behavior of orbits in (X, T). We prove some of them below.

Definition 4.3. A subset $\emptyset \neq \alpha \subseteq \mathcal{E}$ is called an **ideal** if $\mathcal{E}\alpha \subseteq \alpha$.

The set $\mathcal{E}$ itself is a closed ideal, and if $f \in \mathcal{E}$ then $\mathcal{E}f$ is a closed ideal, since $g \mapsto gf$ is continuous. The intersection of any decreasing family of closed ideals is again a closed ideal. Therefore by Zorn's lemma, there exist closed ideals in $\mathcal{E}$ which are minimal to inclusion.

If α is a minimal closed ideal, then $\mathcal{E}\alpha$ is an ideal and $\mathcal{E}\alpha \subseteq \alpha$, so by minimality, $\mathcal{E}\alpha = \alpha$.

Proposition 4.4. *A closed subset $Y \subseteq X$ is a minimal subsystem if and only if $Y = \alpha x$ for some point $x \in X$ and some minimal closed ideal $\alpha \subseteq \mathcal{E}$.*

Proof. Suppose $Y = \alpha x$ as in the statement. Let $y, z \in Y$. Then $y = fx$ for some $f \in \alpha$. Now, $\mathcal{E}f$ is a closed ideal and $\mathcal{E}f \subseteq \mathcal{E}\alpha \subseteq \alpha$, so $\mathcal{E}f = \alpha$. Thus $z \in \mathcal{E}x = \mathcal{E}fx = \mathcal{E}y$ and we conclude that $z \in \overline{O_T(y)}$. Since $y, z \in Y$ were arbitrary, every orbit in y is dense in Y, so Y is minimal.

Conversely, suppose that Y is minimal and let $x \in Y$. Since $T^n x \in Y$ for all n, we have $\mathcal{E}x \subseteq Y$. Let α be a minimal ideal in $\mathcal{E}$. Then $\alpha x \subseteq Y$ and αx is a minimal subsystem of Y, since $T\alpha x \subseteq \alpha x \subseteq Y$. Hence $Y = \alpha x$. □

Definition 4.5. An element $f \in \mathcal{E}$ is **idempotent** if $f^2 = f$.

This means that f acts as the identity map on its image.

Proposition 4.6. *Every closed, minimal ideal $\alpha \subseteq \mathcal{E}$ contains an idempotent. In particular, idempotents exist in $\mathcal{E}$.*

Proof. Consider the family $\mathcal{S}$ of non-empty closed sub-semigroups of α (i.e. closed subsets $\sigma \subseteq \alpha$ satisfying $\sigma\sigma \subseteq \sigma$). This family is non-empty because $\alpha \in \mathcal{S}$, and any decreasing intersection in $\mathcal{S}$ is again in $\mathcal{S}$. So by Zorn's lemma, we can choose a minimal element $\sigma \in \mathcal{S}$.

Let $f \in \sigma$ and note that σf is again a closed sub-semigroup of α and $\sigma f \subseteq \sigma$, so by minimality $\sigma f = \sigma$. Let

$$\tau = \{g \in \sigma \ : \ gf = f\}$$

This is again a closed sub-semigroup of σ, and it is non-empty since $\sigma f = \sigma$. So by minimality again, $\tau = \sigma$, hence $f \in \tau$, and f is an idempotent. □

Definition 4.7. Let (X, T) be a dynamical system. Points $x, y \in X$ are **proximal** if

$$\inf_{n \geq 0} d(T^n x, T^n y) = 0$$

and they are **distal** if this fails, i.e. if

$$\inf_{n \geq 0} d(T^n x, T^n y) > 0$$

Note that if $x, y \in X$ are distal then $f(x) \neq f(y)$ for all $f \in \mathcal{E}$. Put differently, if $f \in \mathcal{E}$ and if $f(x) = f(y)$, then x, y are proximal.

Theorem 4.8. *Let (X, T) be a dynamical system. Then every point is proximal to a uniformly recurrent point.*

Observe that given $x \in X$, we know that there are uniformly recurrent points $y \in \overline{O_T(x)}$. Now, $T^{n_k} x \to y$ for some sequence n_k and if also $T^{n_k} y \to y$ then x, y would be proximal and the theorem would follow. However, while y is certainly recurrent, there is no reason it must recur along the given sequence (n_k). So the content of the theorem is that we can choose y in such a way that its recurrent pattern is similar to the pattern by which x is attracted to y. It is not clear at all why such a point should exist!

Proof. Choose a minimal ideal $\alpha \subseteq \mathcal{E}$ and an idempotent $f \in \alpha$. Let $x \in X$ and let $y = f(x)$. Then $y \in \alpha x$, so y belongs to the minimal subsystem αx, hence y is uniformly recurrent. On the other hand,

$$fy = f(fx) = f^2 x = fx$$

Therefore x, y are proximal. □

To the best of my knowledge, there is no constructive proof of the last theorem (note that we used Zorn's lemma to produce an idempotent, which is crucial to the proof).

4.3 A combinatorial application to IP-sets

In the 1970s, Furstenberg, Katznelson and Weiss uncovered deep connections between topological dynamics and certain problems in combinatorics. We shall see one such connection here.

Definition 4.9. Given a finite or infinite sequence $(n_k) \subseteq \mathbb{N}$, the set of partial sums is

$$FS(n_k) = \{n_{i_1} + \ldots + n_{i_\ell} \ : \ i_1 < i_2 < \ldots < i_\ell \ , \ \ell \in \mathbb{N}\} \tag{2}$$

an **IP-set** is any set of this form.

For example,

- If $(n_k) = (p, p, p, \ldots)$ then $FS(n_k) = p\mathbb{N}$.
- If $(n_k) = (1, 2, 4, 8, \ldots)$ then $FS(N_k) = \mathbb{N}$.
- If $(n_k) = (1, 3, 9, 27, \ldots)$ then $FS(n_k)$ consists of natural numbers whose base-3 expansion contains only the digits $0, 1$ (not the similarity with the Cantor set).

In general, $FS(n_k)$ is not a semigroup under addition, because we are not allowed to use each n_k more than once in the sum in (2). One should think of it as parallelogram with sides $n_1, n_2, \ldots$. In fact this is the source of the name: Infinite-dimensional Parallelogram.

Theorem 4.10. *Let $\mathbb{N} = A_1 \cup \ldots \cup A_r$ be a finite partition of $\mathbb{N}$. Then one of the sets A_i contains an infinite IP-set.*

Proof. In this proof we write $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ and consider the shift space $\Sigma^{\mathbb{N}_0}$,everything there is defined in the same way as before.

Consider the full shift $(\{1, \ldots, r\}^{\mathbb{N}_0}, S)$. Form the sequence $x \in \{1, \ldots, r\}^{\mathbb{N}_0}$ by setting x_0 arbitrary and

$$x_n = i \qquad \Longleftrightarrow \qquad n \in A_i$$

This point encodes the partition (A_i).

Let $y \in \{1, \ldots, r\}^{\mathbb{N}_0}$ be a uniformly recurrent point such that x, y are proximal under the shift map S. Such a y exists by Theorem 4.8.

In the symbolic setting, proximality means the following: For every k there are arbitrarily large n such that all coordinates of x, y agree on the interval $[n, n+k]$.

Let $u = y_0 \in \{1, \ldots, r\}$. We will show that A_u contains an IP-set.

To begin, using uniform recurrence of y, choose k_1 such that the occurrences of u in y occur syndetically with gaps at most k_1. Then, using proximality of x, y, choose n'_1 so that x, y acree on $[n'_1, n'_1 + k_!]$. One of these coordinates, say at index $n'_1 + j_1$, is u. So set $n_1 = n'_i + j_1$; we have found that u appears in x at this coordinate, so $n_1 \in A_u$.

Next, the word $y_0 \ldots y_{n_1}$ appears syndetically in y with gaps at most k_2. Choose $n'_2 > n_1$ so that x, y agree on the interval $[n'_2, n'_2 + (n_1 + k_2)]$. Choose $0 \leq j_2 \leq k_2$ so that $y_1 \ldots y_{n_1}$ occurs in y at $n_2 = n'_2 + j_2$, and hence also in x. Notice that the occurrence of u appears at n_1 in x, y now repeats at $n_2 + n_1$ because the entire word $y_0 \ldots y_{n_1}$ repeats starting at n_2. So now $n_1, n_1 + n_2 \in A_u$.

Proceed inductively. Assume we have found $n_1 < n_2 < \ldots n_p$ such that u appears in x and y at all indices $i \in FS(n_1, \ldots, n_p)$. The word $y_0 \ldots y_{n_1 + \ldots + n_p}$ contains all these occurrences and appears syndetically in y with gap at most k_{p+1}. Choose n'_{p+1} so that x, y agree on $[n_{p+1}, n_{p+1} + (n_1 + \ldots + n_p + k_{p+1})]$, choose $0 \leq j_{p+1} \leq k_{p+1}$ so that $y_0 \ldots y_{n_1 + \ldots + n_{p+1}}$ appears in y (and therefore x) at index $n_{p+1} = n'_{p+1} + j_{p+1}$. In particulat the occurrence at y_0 is replicated at n_{p+1} so $n_{p+1} \in A_u$. Also, for every $i \in FS(n_1, \ldots, n_p)$ the occurrence of u at index i in y repeats in x at $n_{p+1} + i$. Thus, also $n_{p+1} + i \in A_u$ for akk $i \in FS(x_1 \ldots x_p)$.

The end result is that we have constructed $n_1 < n_2 < \ldots$ such that u appears in x at every coordinate $i \in FS(n_k)$, and this is what we wanted. □

This conversion from a problem about subsets of $\mathbb{N}$ to a problem in symbolic dynamics, was introduced by Furstenberg, and is today called the **Furstenberg correspondence principle.**

The proof argument given above in the symbolic setting may be re-formulated in general terms. We leave the proof as an exercise.

Proposition 4.11. *Let (X, T) be a dynamical system, let $x \in X$ and suppose that x is proximal to a uniformly recurrent point y. Then for every open set U containing y, the set $N(x, U)$ contains an IP-set.*

To conclude, we mention the following generalization of the last theorem:

Theorem (Hindman)**.** *If $A \subseteq \mathbb{N}$ is an IP-set and $A_1 \cup \ldots \cup A_r$ is a partition of A, then one of the sets A_i contains an IP-set.*

Compare this to Ramsey's theorem about graphs, which says that in any partition of a large enough complete graph, one of the partition elements contains an induced complete graph of large size. In other words, complete graphs are "hard to destroy": by finitely partitioning a large complete graph, many smaller complete graphs necessarily survive. We have the same phenomenon above: if we finitely partition an IP-set, many smaller IP-sets will survive.

Hindman's theorem can also be proved by dynamical methods, see **??**.

5 Factors and extensions

In this section we introduce the notion of "homomorphim" for dynamical systems, which is called the factor. These provide thelanguage for various results on the structure of dynamical systems, much as homomorphism allow structure theorems in group theory, e.g. the structure theorem for nilpotent groups.

5.1 Factors

Definition 5.1. A dynamical system (Y, S) is a **factor** of a system (X, T), and (X, T) is an **extension** of (Y, S), if there is a continuous onto map $\pi : X \to Y$ that intertwines the actoin: action: $\pi T = S\pi$. We then write $\pi : (X, T) \to (Y, S)$

Examples

1. Trivial factors: Every system is a factor of itself via the identity map, and every system factors to the 1-point system.

2. An isomorphism is a factor map that is injective (note that we use compactness to deduce continuity of the inverse map).

3. When we analyzed the map $T_{10} : x \mapsto 10x \bmod 1$ on $\mathbb{R}/\mathbb{Z}$, we did so by identifying a point $x \in \mathbb{R}/\mathbb{Z}$ with its binary expansion as a number in $[0, 1)$: $x = 0.x_1x_2\ldots$. We used the fact that $T_{10}x = 0.x_2x_3\ldots$.

   Thus, in decimal representation, T_{10} acts as a shift on the space of sequences. More precisely we can form
   $$X = \{0, 1, \ldots, 9\}^{\mathbb{N}}$$
   with the product topology, and let $S : X \to X$ denote the map
   $$(Sx)_n = x_{n+1}$$
   Then S acts on sequences in "the same way" as T_{10} acts on decimal expansions. However, this is not an isomorphism of $(\mathbb{R}/\mathbb{Z}, T_{10})$ with the full shift $\{0, \ldots, 9\}^{\mathbb{N}}$ because decimal expansions are not unique $(0.500\ldots = 0.4999\ldots)$.

   However, we do have a factor map from (X, S) to $(\mathbb{R}/\mathbb{Z}, T_{10})$, given by $(x_1, x_2, \ldots) \mapsto 0.x_1x_2\ldots$.

4. The following example explains the terminology. Let $X = \mathbb{Z}/m\mathbb{Z}$ and $Y = \mathbb{Z}/k\mathbb{Z}$, both with the maps $x \mapsto x + 1$. When k is a factor of m, the map $\pi(x) = x \mod k$ is a factor map between the systems, and conversely, if $\pi : X \to Y$ is a factor map then, since $x \mapsto x + 1$ on X is injective, and since π intertwines the maps $+1$, we will have $|\pi^{-1}(y)| = |\pi^{-1}(y + 1)|$. Thus all fibers have the same size p and $m = pk$, so k is a facto of m.

5. If $f \in C(X)$ is a continuous eigenfunction of $(X.T)$ with eigenvalue $\lambda \in S^1$, then $f : (X, T) \to (S^1, R_\lambda)$ is a factor map, since
   $$f(Tx) = \lambda f(x) = R_\lambda f(x)$$

6. Given (X,T) and (Y,S), form the product $X \times Y$ and $T \times S(x,y) = (Tx, Sy)$. This is called the **product system**. Then the projections $X \times Y \to X$ and $X \times Y \to Y$ are factor maps.

Factors arise naturally in the original physical interpretation of dynamical systems, as follows. Suppose that (X,T) is a system that we are interested in; usually, we are not able to measure the world-state precisely, and all we can usually do is make a measurement of the current state and observe the outcome at different times. Mathematically, this means that there is a (continuous) function $f : X \to \mathbb{R}$, such thThese are the trivial cases.at and at each time step we observe the value of f, so if the world state is $x \in X$ our observation consists of the sequence $(f(x), f(Tx), f(T^2x), \ldots, f(T^N x))$ for some large N. Having measured the system for long enough we may (taking a leap of faith imagine that we have observed $(f(T^n x))_{n=-0}^{\infty}$ for all time. Then what we have observed is a pont in a factor of the original system; for by continuity, the values of f lie in some bounded interval $[a,b]$, and the map

$$\widetilde{f} : x \mapsto (f(T^n x))_{n=1}^{\infty} \in [a,b]^{\mathbb{N}}$$

is a factor map from (X,T) to $([a,b]^{\mathbb{N}}, S)$. We leave it as an exercise to verify that this is indeed a factor map.

It is an interesting question when such a measurment is enough to determine everything about the system. This may, or may not, depend on the function. One can also ask whether adding more measurments – equivalently, taking $f : X \to \mathbb{R}^N$ for some N – may be neough to reveal the entire system. This answer is subtle but nearly understood, more details can be found in **??**.

Lemma 5.2. *Let $\pi : (X,T) \to (Y,S)$ be a factor map.*

1. *If $Y' \subseteq Y$ is a subsystem then $\pi^{-1}(Y') \subseteq X$ is a subsystem.*
2. *If $X' \subseteq X$ is a subsystem then $\pi(X') \subseteq Y$ is a subsystem*
3. *If $x \in X$ then $\pi(O_T(x)) = O_S(\pi(x))$ and $\pi(\overline{O_T(x)}) = \overline{O_S(\pi(x))}$.*

Proof. Continuous images and pre-images of closed sets in compact spaces are closed. Also, if $TX' \subseteq X'$ then $\pi X' \supseteq \pi(TX') = S\pi(X')$, and $\pi X'$ is S-invariant; and if $SY' \subseteq Y'$ and $T\pi^{-1}Y' = \pi^{-1}SY' \subseteq \pi^{-1}Y'$ so $\pi^{-1}Y'$ is T-invariant.

For the last statement, note that if $\pi T^n x = S^n y$ which gives $\pi(O_T(x)) = O_S(\pi(x))$. The statement with closures follows from the general fact that whenever $\pi : X \to Y$ is a continuous map of compact metric spaces, for any set $A \subseteq X$ we have $\pi(\overline{A}) = \overline{\pi(A)}$. □

Many dynamical properties are preserved under factors.

Proposition 5.3. *A factor of a minimal system is minimal.*

Proof. If$(X,T) \to (Y,S)$ is a factor map then every non-trivial subsystem of Y lifts to a non-trivial subsystem of X. Thus, minimality of X implie the same for Y. □

Proposition 5.4. *A factor of an equicontinuous system is equicontinuous.*

Proof. If X is equicontinuous we may choose a T-invariant metric on X. The metric on Y given by

$$d(y,y') = \min\{d(x,x')\ :\ x \in \pi^{-1}(Y)\ ,\ x' \in \pi^{-1}(Y')\}$$

is equivalent to the original metric on Y, and with respect to it, S is an isometry. □

Dynamical properties of points transfer to factors as well:

Lemma 5.5. *If a point is transitive / recurrent / uniformly recurrent / periodic point, then its image under a factor map has the same property.*

We leave the proof as an exercise.

5.2 Factor maps between symbolic systems (the Curtis-Hedlund Theorem)

It is easy to create factors of a symbolic system using the following recipe. Let A, B be finite alphabets. Let $\pi_0 : A^{2n+1} \to B$ be a function, and define a map $\pi : A^{\mathbb{Z}} \to B^{\mathbb{Z}}$ by

$$(\pi x)_i = \pi_0(x_{i-n}, x_{i-n+1} \ldots x_i \ldots x_{i+n})$$

Such a map is called a **sliding block code** (based on π_0), because the symbols in $\pi(x)$ are obtained by "sliding a window of radius n along x" and applying π_0 towhat we see.

The map π above is continuous (the preimage of a cylinder set is a union of cylinder sets), and it commutes with the shift, i.e.. $S \circ \pi = \pi \circ S$ (Note that when $A \neq B$, the map S denotes a different map on the two sides of this expression), Letting $Y = \pi(A^{\mathbb{Z}})$, it is easy to see that Y is closed and shift invariant, and $\pi : A^{\mathbb{Z}} \to Y$ a factor map.

Example 5.6. Let $X = \{0,1\}^{\mathbb{Z}}$ and $\pi x = x_n + x_{n-1} \bmod 1$. This is a sliding block code and a factor map. (We already met this factor map in Problem **??**). .

Theorem 5.7 (Curtis-Hedlund-Lyndon)**.** *If X, Y are subshifts and $\pi : X \to Y$ a factor map, then π is given by a block code.*

Proof. Let $X \subseteq A^{\mathbb{Z}}$ and $Y \subseteq B^{\mathbb{Z}}$. Fix $i \in B$ and consider the cylinde $[i] = \{y \in Y\ :\ y_0 = i\}$. It is open and closed, so $\pi^{-1}[i] \subseteq X$ is open and closed. Since it is open, it is the union of cylinder sets; since it is closed, hence compact, it can be covered by finitely many of these cylinders. Denote them $C_{i,1}, \ldots, C_{i,n(i)}$. For any $x \in C_{i,n(i)}$ we know that $(\pi x)_0 = i$. Repeat this for every $i \in B$. We obtain finitely many cylinder sets $\{C_{i,j}\}_{i \in B, j \leq n(i)}$ such that if we know which $C_{i,j}$ x belongs to, we know $(\pi x)_0$. Since membership to $C_{i,j}$ is determined by $x_{-n}, \ldots, x_n$ for some n, there is a function $\pi : A^{2n+1} \to B$ such that $(\pi x)_0 = \pi_0(x_{-n} \ldots, x)$. Finally, since π is a factor map,

$$(\pi x)_k = \pi(S^k x)_0 = \pi_0((S^k x)_{-n} \ldots (S^k x)_n) = \pi_0(x_{k-n} \ldots x_{k+n})$$

so π is the restriction to X of the sliding block code determined by π_0. □

The same works for one-sided shifts if we define a block code for $\pi_0 : A^n \to B$ by $(\pi x)_i = \pi_0(x_i \ldots x_{i+n-1})$. The details are left as an exercise.

The Curtis-Hedlund further theorem reduces the theory of symbolic systems to combinatorics: If X, Y are given and we ask whether there exists a factor map $X \to Y$, then this boils down to looking for sliding block codes.

For example here is one surprising application. In general, two systems which factor onto each other may factor in many ways; for example $(\mathbb{R}/\mathbb{Z}, R_\alpha) \to (\mathbb{R}/\mathbb{Z}, R_{2\alpha})$ vit the map $x \mapsto 2x$. But any map of the form $x \mapsto 2x + \beta$ is a factor map as well. For symbolic systems, we have:

Corollary 5.8. *If X, Y are subshifts, then there are at most countably many factor maps $\pi : X \to Y$.*

Proof. Every factor map is a sliding block code, and there are countably many sliding block codes, because each is determined by one of countably many maps π_0. ☐

Cellular automata

Sliding block codes are natural models of phisical evolution of infinite configurations, since the dynamics is "local" – each symbol in $\pi(x)$ depends only on nearby symbols in x. A sliding block code from a full shift to itself, $\pi : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$, is sometimes called a *cellular automaton*, and they are often studied as dynamical systems in themselves, i.e. one considers iterates of π (rather than the shift). The theory has a combinatorial flavor; note that there are only countably many sliding block codes (for each given pair of alphabets). There is a close connection with computation and recursion theory, and many problems about the dynamics of cellular automata are not decidable based on the sliding block code.

5.3 Inverse limits

For any finite sequence of factors $X_n \to X_{n-1} \to \ldots \to X_1$ of systems (X_i, T_i), the "top" system (X_n, T_n) contains "all the information", in some sense about all the others. The next definition gives a general construction of a system that sits "at the top" of an infinite sequence of factors of this kind.

Definition 5.9. A directed system of factors is a sequence $\{(X_n, T_n)\}_{n=1}^\infty$ of dynamical systems and factor maps $\pi_n : X_{n+1} \to X_n$, forming the sequence

$$\ldots \xrightarrow{\pi_n} X_n \xrightarrow{\pi_{n-1}} X_{n-1} \to \ldots \xrightarrow{\pi_1} X_1$$

The **inverse limit** $(X, T) = \lim_{\leftarrow}(X_n, T_n)$ is the dynamical system defined as a subset of $\prod_{n=1}^\infty X_n$ by

$$X = \{(x_n)_{n=1}^\infty \,:\, x_n \in X_n \text{ and } x_n = \pi_n x_{n+1}\} \subseteq X^{\mathbb{N}}$$

with the map T defined by

$$T(x)_n = T_n x_n$$

.

Let us verify some properties:

1. $X \neq \emptyset$; taking any $x_1 \in X_1$, we can extend x_1 to $x = (X_n) \in X$ using the fact that π_n are onto.

2. $X = \lim_{\leftarrow} X_n$ is a closed set because π_n are continuous; if $x^i \to x$ and $x^i \in X$, then $x^i_n \to x_n$ for all n, and since $\pi_n(x^i_{n+1}) = x^i_n$ for all i and n and π_n are continuous this relation is maintained in the limit, $\pi_n(x_{n+1}) = x_n$.

3. X is T invariant: if $x \in X$ then for each n we have $\pi_n x_{n+1} = x_n$, so $\pi_n T x_{n+1} = T \pi_n x_{n+1} = T x_n$, so $Tx \in X$.

4. If each T_n is onto then T is onto: if $x = (x_n) \in X$, choose any $y_1 \in T_1^{-1}(x_1)$. Since π_1 is onto, we have $\pi^{-1}(T_1^{-1}(x_1)) = T_2^{-1}(\pi^{-1} x_1)$ so we can choose $y_2 \in T_2^{-1}(x_2)$ with $\pi_1 y_2 = y_1$. continuing by induction we find $y_n \in T_n^{-1}(x_n)$ with $\pi_{n-1} y_n = y_{n-1}$. Then $y = (y_n) \in T^{-1}(x)$.

5. Let $\widetilde{\pi}_n(x) = x_n$. Then $\widetilde{\pi}_n : X \to X_n$ is a factor map and $\pi_{n-1}\widetilde{\pi}_n = \widetilde{\pi}_{n-1}$. The maps $\widetilde{\pi}_n$ are onto (this again uses the fact that π_n are onto). Furthermore if $x \neq x'$ in X then $\widetilde{\pi}_n x \neq \widetilde{\pi}_n x'$ for some n.

6. Suppose that (Y, S) is a system and $\sigma_n : Y \to X_n$ are factor maps such that $\pi_{n-1}\sigma_n = \sigma_{n-1}$. Then the map $\sigma : Y \to \prod_{n=1}^{\infty} X_n$ given by

$$\sigma(y) = (\sigma_n(y))_{n=1}^{\infty}$$

is continuous, and maps Y onto X; indeed the relation $\pi_{n-1}\sigma_n = \sigma_{n-1}$ shows that $\sigma(y) \in X$ and since σ_n are all onto, given $x \in X$, we can find points $y_n \in Y$ such that $\sigma_n y_n = x_n$. Note that for $m \leq n$,

$$\sigma_m y_n = \pi_m \pi_{m+1} \ldots \pi_{n-1} \sigma_n y_n = \pi_m \pi_{m+1} \ldots \pi_{n-1} x_n = x_m$$

Passing to a subsequence such that $y_{n_k} \to y$, we have

$$\sigma_m(y) = \lim \sigma_m(y_{n_k}) = \lim x_m = x_m$$

so $\sigma(y) = x$, showing that σ is onto X. Finally, it is not hard to check that σ is a factor map – one must verify that $\sigma S = T\sigma$. We leave this verification as an exercise.

7. If (Y, S), $\sigma_1, \sigma_2, \ldots$ and σ are as above and if $\{\sigma_n\}$ separate points (i.e. for every $y \neq y'$ in Y we have $\sigma_n(y) \neq \sigma_n(y')$ for some n), then σ is 1-1, so it is an isomorphism of (Y, S) and (X, T). This shows, that the properties of (X, T) stated in (5) characterize it up to isomorphism.

This construction has many uses. For example

Proposition 5.10. *Let (X, T) be a dynamical system with T onto. Then there exists an invertible dynamical system $(\widetilde{X}, \widetilde{T})$ and $\pi : \widetilde{X} \to X$ a factor map.*

Furthermore, we can choose (X, T) to be minimal in the following sense: If (Y, S) is any invertible dynamical system with a factor map $\sigma : Y \to X$, then there exists a factor map $\rho : Y \to \widetilde{X}$ such that $\sigma = \pi\rho$.

Definition 5.11. The system in the proposition is unique up to isomorphism and is called the **natural extension** of (X, T).

Proof. Since T is onto and trivially intertwines the T-action, it is a factor map, and we can take the inverse limit of the directed system $\ldots \xrightarrow{T} X \xrightarrow{T} X \to \ldots \xrightarrow{T} X$. Let $(\widetilde{X}, \widetilde{T})$ denote the inverse limit as constructed above and $\pi = \widetilde{\pi}_1$. A point $x \in \widetilde{X}$ has the form (x_n) with $x_n = T x_{n+1}$. Therefore,

$$(\widetilde{T}x)_{n+1} = T x_{n+1} = x_n$$

It follows that the shift map S on X is an inverse to $\widetilde{T}$. □

The existence of the natural extension allows many results to be transfered painlessly from invertible systems to non-invertible ones and vice versa. For example, suppose we had proved the existence of a forward-recurrent point in any invertible system; if now (X,T) is non-invertible, we can apply the result in its natural extension to find a forward-recurrent point there, and project it down to X.

5.4 Problems

1. Complete the proofs in this section.Show that the inverse limit of a minimal systems is minimal.
2. Show that the natural extension of an invertible system is isomorphic to the original system.
3. Let $X \subseteq \Lambda^{\mathbb{N}}$ be a subshift such that S is onto. Let $Y \subseteq \Lambda^{\mathbb{Z}}$ be the subshift satisfying $L(Y) = L(X)$. Show that (Y,S) is isomorphic to the natural extension of (X,S).
4. In this question we discuss skew-products which are rich souce of examples for extensions for a given dynamical system.

   Definition 5.12. Let (Y,S) be a dynamical system, let Z be a compact metric space, and $f : Y \to C(Z,Z)$ a continuous map. The system $X = Y \times Z$ with the map

   $$T(y,z) = (Sy, f_y z)$$

   is called a **skew product** over Y, and is denoted $Y \times_f Z$. The map f is called the **cocycle** of the skew product, (Y,S) is the **base** and and Z the **fiber**.

   Remarks and examples

   - If $X = Y \times_f Z$ is a skew-product then the projection $\pi(y,z) = y$ is a factor map $X \to Y$.
   - Start with $(Y,S) = (\mathbb{R}/\mathbb{Z}, R_\alpha)$ for some $\alpha \in \mathbb{R}$. Take $Z = \mathbb{R}/\mathbb{Z}$ and for $y \in Y = \mathbb{R}/\mathbb{Z}$ let $f_y(z) = z + y$. Then the skew-product map in $Y \times_f Z$ is

     $$(x,y) \mapsto (x+\alpha, x+y)$$

     This is a map of the 2-torus.
   - Let $Y = Z = \{1,-1\}^{\mathbb{Z}}$ and let S be the shift map. Define $f : Y \to \{S, S^{-1}\}$ by

     $$f(x) = S^{x_0} = \begin{cases} S & x_0 = 1 \\ S^{-1} & x_0 = -1 \end{cases}$$

     Then the skey product $Y \times_f Z$ can be interpreted as follows: the first coordinate contains instructions which way to shift the second coordinate each time T is applies. Thus, given $(y,z) \in Y \times_f Z$, we have

     $$T^n(y,x) = (S^n y, S^{\sum_{i=0}^{n-1} y_i} z)$$

     For a randomly chosen sequence y and any z, the central limit theorem implies that for large n the second component of $T^n(y,z)$ will be a shift of z by an order of $\sqrt{n}$.

Not every extension $\pi : X \to Y$ is a skew product over Y. In fact, it is easy to give examples where the fibers $\pi^{-1}(x)$ are not homeomorphic (all one needs is a topological example, then take the identity maps). Even if the fibers are homeomorphic, we may not have $X = Y \times Z$ for any Z.

For example, consider $(X,T) = (\mathbb{R}/\mathbb{Z}, R_\alpha)$ and $(Y,S) = (\mathbb{R}/\mathbb{Z}, R_{2\alpha})$ with $\pi : X \to Y$ given by $\pi(x) = 2x \bmod 1$. Then $\pi^{-1}(y) = \{y, y+1/2\}$ but $X \ncong Y \times \{0, 1/2\}$ as topological spaces.

Problems

(a) Let $X = Y = \{0,1\}^{\mathbb{Z}}$ and $\pi : X \to Y$ given by

$$\pi(x)_n = x_n + x_{n+1} \bmod 1$$

Show that this map is 2-to-1 everywhere, that topologically $X = Y \times \{0,1\}$. Then show that it is a skew product.

(b) Let (X,T),(Y,S) be a dynamical systems. Suppose $X = Y \times Z$ as a topological space, and that $\pi(y,z) = y$ is a factor map $X \to Y$. Show that X is a skew-product over Y.

6 Transitivity and weak mixing

We continue to study systems at the “chaotic” end of the spectrum. First, we introduce the class of transitive system, which weakens the notion of minimality. Then we discus efine mixing systems, which are systems where every open set “spreads over the entire space” not only when viewed over all time, but also at many individual (but perhaps far away) times. Finally, we show that weak mixing systems are precisely those with no equicontinuous factors.

6.1 Transitivity

In a minimal system (X,T), every non-emtpy open set covers the space under iteration by T. The following notion is slightly weaker.

Definition 6.1. A (X,T) a dynamical system is transitive if for every open set $\emptyset \neq U,V \subseteq X$ there exists $n \geq 0$ such that $U \cap T^{-n}V \neq \emptyset$.

Equivalently: $\bigcup_{n=0}^{\infty} T^{-n}U$ is dense in X for every open $\emptyset \neq U \subseteq X$.

Examples

1. Every minimal system is transitive.
2. Full shifts are transitive. Indeed, let $U,V \subseteq \Lambda^{\mathbb{N}}$. Choose words $u,v \in \Lambda^*$ such that $[u] \subseteq U$ and $[v] \subseteq V$, and let $x = uvvvv\ldots \in \Lambda^{\mathbb{N}}$. Evidently, $x \in [u] \cap S^{-|u|}[v] \subseteq U \cap S^{-|u|}V$. So the latter intersection is non-empty and the system is transitive.
3. The identity map on any space with more than one point is not transitive.
4. Translation of $\mathbb{R}/\mathbb{Z}$ by α for a rational α on is not transitive. Indeed, if $\alpha = \ell/k$ in reduced terms, then $U = (0,1/2k)$ satisfies $\bigcup_{n=0}^{\infty} R_{\alpha}^{-n} = \bigcup_{m=0}^{k-1}(m/k, m/k+1/2k)$, which is not dense.
5. Translation by (α,α) on $\mathbb{R}^2/\mathbb{Z}^2$. This is non-transitive even when α is irrational.
6. A factor of a transitive system is transitive, since if $\pi : (X,T) \to (Y,S)$ is a factor map, and $\emptyset \neq U,V \subseteq Y$ are open, then $\pi^{-1}(U), \pi^{-1}(V) \neq \emptyset$ are open in X, hence for some n we have
$$\begin{aligned}\emptyset &\neq \pi^{-1}(U) \cap T^{-n}\pi^{-1}(V)\\ &= \pi^{-1}(U) \cap \pi^{-n}(S^{-1}V)\\ &= \pi^{-1}(U \cap S^{-n}V)\end{aligned}$$
hence $U \cap S^{-n}V \neq \emptyset$.

There is a close connection between transitivity and the existence of transitive points (points with dense orbit). It will be convenient to write

$$\tau(X,T) = \{x \in X \ : \ x \text{ is a transitive point}\}$$

This set may, of course, be empty. We abbraviate it as $\tau(X)$ or $\tau(T)$ depending on the context.

Proposition 6.2. *Let (X,T) be a dynamical system. Then $\tau(X) \subseteq \bigcup_{n=1}^{\infty} T^{-n}U$ for every open set $U \subseteq X$, and if $\{V_i\}_{i=1}^{\infty}$ is a basis for the topology*[6] *and $V_i \neq \emptyset$, then*

$$\tau(X) = \bigcap_{i=1}^{\infty} \bigcup_{n=1}^{\infty} T^{-n}V_i$$

Consequently, if (X,T) is transitive, then $\tau(X,T)$ is a dense G_δ set, and in particular, transitive points exist.

Proof. Let $U \neq \emptyset$ be open. If x is transitive then by definition $O_T(x) \cap U \neq \emptyset$, i.e. $T^n x \in U$ for some $n \geq 0$. Thus, $\tau(X) \subseteq \bigcup_{n=1}^{\infty} T^{-n}U$. Moreover, given $\{V_i\}$,

$$\begin{aligned}
x \text{ is transitive} \quad &\iff \quad O_T(x) \text{ intersects every non-emtpy open set} \\
&\iff \quad O_T(x) \text{ intersects } V_i \text{ for every } i \\
&\iff \quad x \in \bigcup_{n=1}^{\infty} T^{-n}V_i \text{ for every } i \\
&\iff \quad x \in \bigcap_i \bigcup_{n=1}^{\infty} T^{-n}V_i
\end{aligned}$$

Finally, suppose that (X,T) is transitive. Fixing a countable basis $\{V_i\}$ for the topology, by the previous lemma, $\tau(X) = \bigcap_i \bigcup_{n=1}^{\infty} T^{-n}V_i$. Transitivity implies that each union in this expression is dense, and the intersection is countable, so $\tau(X)$ is a dense by Baire's theorem. $\square$

The converse of this proposition is almost true, but not quite. To see what can go wrong, consider $X = \{\frac{1}{n} : n \in \mathbb{N}\} \cup \{0\}$ with the map $T0 = 0$ and $T(1/n) = 1/(n+1)$. Then $1 \in X$ is a transitive point, but the system is not transitive: $\bigcup_{n=0}^{\infty} T^{-n}\{1/n\} = \{1/k : k \leq n\}$ is not dense in X.

The problem in this example may be identified as the presence of isolated points. Without such points, we obtain a converse to the previous proposition.

Proposition 6.3. *Let (X,T) be a dynamical system without isolated points. If there exist transitive points, then (X,T) is transitive.*

Proof. Let $x \in X$ be a transitive point. We claim that also $T^n x$ is transitive for every $n \geq 0$. Assuming this, for any open set $U \neq \emptyset$ we have $O_T(x) \subseteq \tau(X) \subseteq \bigcup_{n=0}^{\infty} T^{-n}U$, so $\bigcup_{n=0}^{\infty} T^{-n}U$ is dense, giving transitivity.

So fix $x' = T^n x$ and let $y \in X$. Let V be a neighborhood of y. Then V is infinite, because y is not isolated, so $W = V \setminus \{x, Tx, \ldots, T^n x\}$ is a non-empty open set, and there exists a $k \in \mathbb{N}$ such that $T^k x \in W$. Clearly $k > n$, so $T^{k-n}x' \in W \subseteq V$. This shows that x' has dense forward orbit. $\square$

We next discuss transtivity in invertible systems. When T is invertible we say that it is bi-transitive if for every $\emptyset \neq U, V \subseteq X$ open, there is an $n \in \mathbb{Z}$ such that $U \cap T^n V \neq \emptyset$, i.e. $\bigcup_{n \in \mathbb{Z}} T^n U$ is dense whenever $U \neq \emptyset$ is open. Recall that a point is bi-transitive if its two-sided

[6] A basis for the topology of a metric space X is a family $\{V_i\}$ of open sets such that every open set in X is a union of (some of the) V_i. When the space is compact, countable bases exist (e.g. take $V_i = B_{r_i}(x_i)$, where $\{x_i\}$ is a countable dense set and $\{r_i\}$ an enumeration of the positive rationals).

orbit is dense. We then have analogs of Lemma 6.2 and Proposition 6.3, with bi-transitivity in place of transitivity, and with the unions $\bigcup_{n=1}^{\infty} T^{-n}U$ replaced by $\bigcup_{n=-\infty}^{\infty} T^{-n}U$.

One naturally wonders about the relation between transitivity of T and of T^{-1}, and also their connection with bi-transitivity. This is addressed in the following proposition. We discuss the situation when there are isolated points in the problems section.

Proposition 6.4. *Let (X,T) be an invertible dynamical system without isolated points. Then each of the following are equivalent:*

1. *T is transitive.*
2. *T^{-1} is transitive.*
3. *T is bi-transitive.*

Proof. Let us first show the equivalence of (1) and (2). By symmetry it is enough to prove that (1) implies (2), so assume T is transitive. For every non-empty $\emptyset \neq U, V \subseteq X$, there exists an $n \geq 0$ such that $V \cap T^{-n}U \neq \emptyset$. Applying the homeomorphism $T^n = (T^{-1})^{-n}$ we get

$$(T^{-1})^{-n} V \cap U \neq \emptyset$$

Since this holds for all such U, V this shows that T^{-1} is transitive.

Each of the conditions (1) and (2) imply (3), simply because $\bigcup_{n\geq 0} T^{-n}U \subseteq \bigcup_{n\in\mathbb{Z}} T^n U$.

Now suppose that (3) holds. Let $x \in X$ be bi-transitive (there is a dense G_δ set of such points). We first claim that it is either forward or backward recurrent. Indeed, for every neighborhood U of x, for every N, the set $U \setminus \{T^i x \,:\, |x| < N\}$ is non-empty and open so there is an n with $|n| > N$ and $T^n x \in U$. It follows that $T^{n_k} x \to x$ with $|n_k| \to \infty$, and passing to a subsequence we can assume $n_k \to \infty$ or $n_k \to -\infty$.

Now apply Lemma 2.5, to conclude that either the forward or the backward orbit closure is equal to $O_T^{\pm}(x) = X$. Thus (again using the absence of transitive points), T is forward or backwards transitive, and we have proved (1) or (2) (and hence both). □

6.2 Weak mixing

As we shall see here and in the coming sections, the behavior of the self-product system $(X \times X, T \times T)$ provides a wealth of information about X. Outside of the trivial case where X is a single point, the product $X \times X$ is never minimal, because, for example, it contains the **diagonal subsystem** $\Delta = \{(x,x) \;:\; x \in X\}$, and also the **off-diagonal subsystems**, $\Delta_k = \{(x, T^k x) \,:\, x \in X\}$. On the other hand, it sometimes does happen that the self-product is transitive, and it turns out this has a close connection to (the absence of) equicontinuous behavior in the system.

Definition 6.5. A system (X,T) is **weak mixing** if $(X \times X, T \times T)$ is transitive.

Remarks and examples

1. Why "weak" mixing? Because there is another stronger notion, calles strong mixing. We will discuss it in the problems section..

2. The full shift $X = \Lambda^{\mathbb{N}}$ is weak mixing. To see this note that $X \times X$ is isomorphic to $(\Lambda \times \Lambda)^{\mathbb{N}}$, which is again a full shift, and so it is transitive.

   Nevertheless, let us examin more closely how one might construct a transitive point in $\Lambda^{\mathbb{N}} \times \Lambda^{\mathbb{N}}$. To be concrete take $\Lambda = \{0, \ldots, 9\}$. A transitive point in the product must consist of a pair of transitive points, since it must project to a transitive point on each coordinate. So let us begin with the point corresponding to the Champernown number

   $$x = (0, 1, 2, \ldots, 9, 1, 0, 1, 1, 1, 2, 1, 3, \ldots, 1, 9, 2, 0, \ldots)$$

   and let us lok for a point y such that (x, y) is transitive in the product. What this means is that in the pair (x, y), an every pair $a, b \in \Lambda^n$, we want the pair (a, b) to appear in (x, y). Now, a appears in x, and in fact it appears infinitely many times. So we can construct y as follows. Let $\Lambda^* = \{c^1, c^2, \ldots\}$ be an enumeration of all finite words. Let $\{(i_1, j_1), (i_2, j_2), \ldots\} = \mathbb{N} \times \mathbb{N}$ be an enumeration of all pairs of natural numbers. Now form y as recursively follows. At stage n of the construction, find the first occurrence of c^{i_n} in x that is to the right of all symbols in y that have already been defined, and place c^{j_n} at this position in y. After this has been done for all $n \in \mathbb{N}$, any undefined symbols in y are set to 0. Now (x, y) contains every pair (c^i, c^j), so it is transitive.

   We remark that in general, when X is weak mixing and $x \in X$ is transitive, there is no guarantee that a $y \in X$ can be found such that (x, y) is transitive in $X \times X$.

3. Weak mixing implies transitivity: indeed (X, T) is a factor of $(X \times X, T \times T)$, e.g. by projection to the first coordinate, so the image of a transitive point in the latter is a transitive point in the former.

4. Transitive systems may not be weak mixing. To see this, consider $\mathbb{Z}/p\mathbb{Z}$ with $Tx = x + 1 \bmod p$ and $p > 1$. Then $X \times X$ breaks into p periodic the points, namely the cosets of the subgroup generated by $(1, 1)$. These are distinct minimal sets so there are no transitive points in $X \times X$.

5. The factor of a weak mixing system is weak mixing. Indeed, if $\pi : (X, T) \to (Y, S)$ is a factor map, then $\pi \times \pi : X \times X \to Y \times Y$ is a factor map of the product system, so transitivity of $X \times X$ implies the same for $Y \times Y$.

For $U, V \subseteq X$ let

$$N(U, V) = \{n \geq 0 \ : \ U \cap T^{-n}V \neq \emptyset\} \tag{3}$$

Transitivity of (X, T) just means that $N(U, V) \neq \emptyset$ for every open sets $U, V \neq \emptyset$. Weak mixing can be similarly characterized:

Lemma 6.6. *For all open sets $\emptyset \neq U, U', V, V' \subseteq X$,*

$$N(U, V) \cap N(U', V') = N(U \times U', V \times V')$$

where the right hand side is taken in $(X \times X, T \times T)$.

In particular, (X, T) is weak mixing if and only if $N(U, V) \cap N(U', V') \neq \emptyset$ for all U, V, U', V' as above.

Proof. The first statement is clear form the identity

$$(U \cap T^{-n}U') \times (V \cap T^{-n}V') = (U \times V) \cap (T \times T)^{-n}(U' \times V')$$

Since the product of open sets form a basis for the topology of $X \times X$, the condition in the last statement is equivalent to $N(\widetilde{U}, \widetilde{V}) \neq \emptyset$ for all $\emptyset \neq \widetilde{U}, \widetilde{V} \subseteq X \times X$, which is transitivity of $X \times X$, i.e. weak mixing of X. □

One can view the fact that every two sets $N(U, V)$ intersect is a measure of their largeness. In fact, not only do they intersect, but the intersection is large in the same sense:

Lemma 6.7. *Suppose that (X, T) is weak mixing. Then for all open sets $\emptyset \neq U, U', V, V' \subseteq X$, there exist open sets $\emptyset \neq W, W' \subseteq X$ such that $N(W, W') \subseteq N(U, V) \cap N(U', V')$.*

Proof. Choose k such that $W = U \cap T^{-k}U' \neq \emptyset$ and $W' = V \cap T^{-k}V' \neq \emptyset$; these exist because by the previous lemma, $N(U, U') \cap N(V, V') \neq \emptyset$.

Now given n such that $W \cap T^{-n}W' \neq \emptyset$, we have

$$\begin{aligned} U \cap T^{-n}V &\supseteq W \cap T^{-n}W' \qquad \text{because } U \in W\,,\ V \supseteq W' \\ &\neq \emptyset \end{aligned}$$

hence $n \in N(U, V)$. Also,

$$\begin{aligned} T^{-k}(U' \cap T^{-n}V') &= T^{-k}U' \cap T^{-n}(T^{-k}V') \\ &\supseteq W \cap T^{-n}W' \qquad \text{because } U' \supseteq T^{-k}W\,,\ V' \supseteq T^{-k}W' \\ &\neq \emptyset \end{aligned}$$

so $n \in N(U', V')$. □

Proposition 6.8. *If (X, T) is weak mixing then $(X \times X, T \times T)$ is weak mixing and more generally all cartesian powers $(X^{\times\alpha}, T^{\times\alpha})$ are weak mixing.*

Proof. By the second part of Lemma 6.6, weak mixing of $X \times X$ means that any two sets of the form $N(\widetilde{U}, \widetilde{V})$ are non-empty, for open sets $\emptyset \neq \widetilde{U}, \widetilde{V} \subseteq X \times X$. By the first part of the same lemma, $N(\widetilde{U}, \widetilde{V})$ is itself an intersection of two sets of the form $N(U, V)$ with $\emptyset \neq U, V \subseteq X$ open.

Thus, weak mixing for $X \times X$ would follow from the non-trivial intersection of every fours sets of the form $N(U, V)$ with $\emptyset \neq U, V \subseteq X$.

But the previous lemma says that when (X, T) is weak mixing, the intersection of two such sets contains a set of the same form; so by induction any finite intersection of such sets contains a set of the same form, and so it is non-empty. □

One may ask, what do the sets $N(U, V)$ look like in a weak mixing system?

Definition 6.9. A set $E \subseteq \mathbb{N}$ is **thick** if it contains intevals of arbitrary length, i.e.: for every $n \geq 1$ there exists i such that $\{i, i+1, \ldots, i+n-1\} \subseteq E$.

Proposition 6.10. *If (X, T) is weak mixing then $N(U, V)$ is thick for all open sets $\emptyset \neq U, V \subseteq X$.*

Proof. Suppose that (X, T) is weak mixing. Fix n and open sets $U, V \neq \emptyset$ and consider the sets $\widetilde{U} = U^{\times n}$ and $\widetilde{V} = V \times T^{-1}V \times \ldots \times T^{-(n-1)}V$ in $X^{\times n}$. By weak mixing of the latter system, there exists $i \geq 0$ such that $U \cap (T^{\times n})^{-i}\widetilde{V} \neq \emptyset$. This implies that $\{i, i+1, \ldots, i+n-1\} \subseteq N(U, V)$, so $N(U, V)$ is thick. □

Here is one application:

Proposition 6.11. *If (X, T) is weak mixing and (Y, S) is minimal, than $X \times Y$ is transitive.*

Proof. Write $Z = X \times Y$. It suffices to show that for every pair of open sets $\emptyset \neq W, W' \subseteq Z$ in some basis for the topology of Z, we have $W \cap (T \times S)^{-n}W' \neq \emptyset$ for some $n \geq 0$.

Consider $W = U \times V$ and $W' = U' \times V'$ for open sets $U, U' \subseteq X$ and $V, V' \subseteq Y$. Then

$$N(W, W') = N(U, U') \cap N(V, V')$$

Since$Y$ is minimal, $N(V, V')$ is syndetic; i.e. it has gaps bounded by some $\ell > 0$. Since (X, T) is weak mixing, $N(U, U')$ is thick, so it contains some interval of length $> \ell$. Thus the intersection above non-empty, as desired. □

6.3 Equicontinuous factors

The purpose of this section is to connect weak mixing with equicontinuity. The easy direction of this relation is the observation is that equicontinuity is an obstruction to weak mixing:

Proposition 6.12. *If (X, T) is equicontinuous and X contains more than one point, then it is not weak mixing.*

Proof. By Theorem 2.4 we may assume that T is an isometry. Now if $(x, y) \in X \times X$, then

$$O_{T \times T}(x, y) \subseteq \{(x', y') \in X \times X \ : \ d(x', y') = d(x, y)\}$$

This is a closed proper subset of $X \times X$ so (x, y) is not transitive. □

Corollary 6.13. *If (X, T) is weak mixing then it has no non-trivial equicontinuous factors.*

Proof. Immediate from the last proposition and the fact that a factor of a weak mixing system is weak mixing. □

It turns out that, for minimal systems, the converse is also true: Any minimal system that is not weak mixing necessarily admits an equicontinuous factor. We prepare the ground for the proof in the next sections.

The weak-* topology on $\mathcal{P}(X)$

(This section is more detailed than what was presented in class).

We first recall some basic tools for studying measures on compact spaces.

Let X be a compact metric space and let $\mathcal{M}(X)$ denote the linear space of signed (finite) Borel measures, and $\mathcal{P}(X) = \mathcal{P}(X, \mathcal{B}) \subseteq \mathcal{M}(X)$ the convex space of Borel probability measures. Two measures $\mu, \nu \in \mathcal{M}(X)$ are equal if and only if $\int f d\mu = \int f d\nu$ for all $f \in C(X)$, and the maps $\mu \mapsto \int f d\mu$, $f \in C(X)$, separate points.

Definition 6.14. Let X be a compact metric space. The weak-* topology on $\mathcal{M}(X)$ (or $\mathcal{P}(X)$) is the weakest topology that make the maps $\mu \mapsto \int f\,d\mu$ continuous for all $f \in C(X)$. In particular,

$$\mu_n \to \mu \text{ if and only if } \int f d\mu_n \to \int f d\mu \text{ for all } f \in C(X)$$

Lemma 6.15. *Let X be a compact metric space and $\mathcal{F} \subseteq C(X)$ a dense set of funtions. Suppose that $\mu_n \in \mathcal{P}(X)$ and $\lim \int f d\mu_n$ exists for all $f \in \mathcal{F}$. Then there exists $\mu \in \mathcal{P}(X)$ such that $\mu_n \to \mu$, that is, $\int f d\mu = \lim \int f d\mu_n$ for all $f \in C(X)$.*

Proof. Let $V = \mathrm{span}\mathcal{F}$. By assumption $\lim \int f d\mu_n$ exists for all $f \in \mathcal{F}$, and hence for all $f \in V$, since integrals and limits are finitely additive. For $f \in V$ denote the limit by $\Lambda(f)$. This is positive, linear, bounded function on V and so extends to such a function on $C(X) = \overline{V}$, which we denote also by Λ. By the Riesz representation theorem there exists $\mu \in \mathcal{P}(X)$ such that $\Lambda(f) = \int f d\mu$ for all $f \in C(X)$. We now claim $\int f d\mu_n \to \int f d\mu$ for all $f \in C(X)$. We already know this for , and V is dense. Fixing any $f \in C(X)$ let $\varepsilon > 0$ and $g \in (V)$ with $\|f - g\|_\infty < \varepsilon$. We get

$$\begin{aligned} |\int f d\mu_n - \int f d\mu| &< |\int f d\mu_n - \int g d\mu_n| + |\int g d\mu_n - \int g d\mu| + |\int g d\mu - \int f d\mu| \\ &< \varepsilon + |\int g d\mu_n - \int g d\mu| + \varepsilon \\ &\to 2\varepsilon \qquad \text{as } n \to \infty \end{aligned}$$

Since ε was arbitrary, we get $\int f d\mu_n \to \int f d\mu$. □

Proposition 6.16. *The weak-* topology is metrizable and compact.*

Proof sketch. Using the Stone-Weierstrass theorem shoose a dense sequence $\{f_i\}_{i=1}^{\infty} \subseteq C(X)$. Define a metric on $\mathcal{P}(X)$ by

$$d(\mu, \nu) = \sum_{i=1}^{\infty} 2^{-i} \left| \int f_i d\mu - \int f_i d\nu \right|$$

One shows that this metric is compatible with the topology. Next, if $\mu_n \in \mathcal{P}(X)$ is a sequence of measures, a diagonal argument can be used to show that there is a subsequence $\mu_{n(k)}$ such that for every i, the limit $\lim \int f_i d\mu_{n(k)}$ exists. The previous lemma now shows that $\mu_{n(k)} \to \mu$ for some measure μ. This proves sequential compactness, which, by metrizability, is compactness. □

Let (X, T) be a topological dynamical system. Then we also get an induced map $T : \mathcal{P}(X) \to \mathcal{P}(X)$, given by $\mu \mapsto \mu \circ T^{-1}$.

Lemma 6.17. *Let (X, T) be a topological dynamical system. Then the induced map $T : \mathcal{P}(X) \to \mathcal{P}(X)$ is continuous.*

Proof. If $\mu_n \to \mu$ then for $f \in C(X)$,

$$\int f\, dT\mu_n = \int f \circ T\, d\mu_n \to \int f \circ T\, d\mu = \int f\, dT\mu$$

This shows that $T\mu_n \to T\mu$, so T is continuous. □

Invariant measures

In a dynamical system (X, T), a Borel probability measure $\mu \in \mathcal{P}(X)$ is T-invariant if $\mu(A) = \mu(T^{-1}A)$ for all Borel sets $A \in \mathcal{B}$. Our goal is to show that such measures exist.

For $x \in X$ let

$$\mu_{x,N} = \frac{1}{N}\sum_{n=0}^{N-1}\delta_{x,N}$$

This is a probability measure and we note that

$$\int f d\mu_{x,N} = \frac{1}{N}\sum_{n=0}^{N-1} f(T^n x)$$

for all $f \in C(X)$.

Proposition 6.18. *Every topological dynamical system (X, T) admits invariant measures.*

Proof. Let $x \in X$ be an arbitrary initial point and let $\mu_N = \mu_{x,N}$ be as above. Passing to a subsequence $N(k) \to \infty$ we can assume by compactness that $\mu_{N(k)} \to \mu \in \mathcal{P}(X)$. We show that $\int f\,d\mu = \int f \circ T\,d\mu$ for all $f \in C(X)$:

$$\begin{aligned}
\int f\,d\mu - \int f \circ T\,d\mu &= \lim_{k\to\infty}\int (f - f\circ T)\,d\mu_{N(l)} \\
&= \lim_{k\to\infty}\frac{1}{N(k)}\sum_{n=0}^{N(k)-1}\int (f - f\circ T)(T^n x) \\
&= \lim_{k\to\infty}\frac{1}{N(k)}\left(f(T^{N(k)-1}x) - f(x)\right) \\
&= 0
\end{aligned}$$

because f is bounded.

Next, if $U \subseteq X$ is a bounded open set let $f_n \in C(X)$ be functions increasing to 1_U. Then $f_n \circ T$ increase to $!_U \circ T = 1_{T^{-1}U}$. Therefore

$$\begin{aligned}
\mu(U) &= \int 1_U d\mu \\
&= \lim_{n\to\infty}\int f_n f\mu \qquad \text{by monotone convergence} \\
&= \lim_{n\to\infty}\int f_n \circ T d\mu \qquad \text{because } f_n \in C(X) \\
&= \int 1_{T^{-1}U} d\mu \qquad \text{by monotone convergence} \\
&= \mu(T^{-1}U)
\end{aligned}$$

Passing to complements we have the same for closed sets. Finally, to go from open and closed sets to all Borel sets one uses regularity of the measure. We leave the details as an exercise. □

This proves that invariant measures always exist. Now we specialize to the case where (X, T) is minimal. Let $U \neq \emptyset$ be open, and choose $V \neq \emptyset$ open with $\overline{V} \subseteq U$. By minimality, $T^n x \in V$ on a syndetic set of n, so there is an $\ell > 0$ such that the gap between consecutive visits of the orbit to V in at most ℓ. It follows, that, up to an $O(\ell)$ error, from time 0 to $N-1$, the number of visits to V is N/ℓ. Chose an $f \in C(X)$ with $1_{\overline{V}} \leq f \leq 1_U$. Then

$$\mu(U) \geq \int f d\mu = \lim_{k\to\infty} \int f d\mu_{N_k} \geq \lim \mu_N(U) = \frac{1}{\ell} + O_\ell(\frac{1}{N})$$

so $\mu(U) \geq 1/\ell$. In particular, $\mu(U) > 0$.

We have proved:

Theorem 6.19 (Bogolyubov)**.** *Let (X, T) be a dynamical system and $x \in X$. Then there exists a T-invariant Borel probability measure μ on X. If T is minimal, then $\mu(U) > 0$ for every open set $U \neq \emptyset$.*

Existence of equicontinuous factors

We are now ready for the main result of this section:

Theorem 6.20. *If (X, T) is invertible, minimal and not weakly mixing then it admits an equicontinuous factor.*

Lemma 6.21. *Let $g : X \to \mathbb{R}$ be a lower semi-continuous function, i.e., for all $x \in X$ and $x_n \to x$, we have $g(x) \geq \limsup g(x_n)$. If (X, T) is minimal and g is invariant, then g is constant.*

Proof. A standard argument shows that g achieves a maximal value at some point x_0. For any $x \in X$, Choose n_k so that $T^{n_k} x_0 \to x$. Then

$$\begin{aligned} g(x_0) &= g(T^{n_k} x_0) && \text{by invariance of } g \\ &= \limsup_{k\to\infty} g(T^{n_k} x_0) \\ &\leq g(x) \end{aligned}$$

Since $g(x_0)$ is the maximal value of g, we have $g(x) = g(x_0)$, and since x was arbitrary, g is constant.

We turn to the proof of the theorem.

If X is finite then, being minimal, it consists of a single periodic orbit, so T itself is equicontinuous. Thus we may assume X is infinite, and in particular, X has no isolated points. Also $X \times X$ has no isolated points.

Since X is not weak mixing, $X \times X$ is not transitive, and hence, by Proposition **??**, is is not bi-transitive. Thus there exists an open set $\emptyset \neq U \subseteq X \times X$ such that $V = \bigcup_{n=-\infty}^{\infty} T^n U$ is not dense. Write $Y = \overline{V}$ and $W = (X \times X) \setminus Y$. Then Y is subsystem of $X \times X$ with interior V and complement W.

Let μ be an invariant probability measure on X giving positive mass to every non-empty open set.

For $A \subseteq X \times X$ we write

$$A_x = \{y \in Y \ : \ (x, y) \in A\}$$

This is essentially the intersection $\pi^{-1}(x) \cap Y$ of A with the fiber $\pi^{-1}(x)$ over x, where we identify the fiber with X in the natural way. Now define $f : X \to \mathcal{B}$ denote the map

$$f(x) = 1_{Y_x} \in L^1(\mu)$$

On should think of this as mapping x to Y_x, with the pseudo-metric on sets given by $d(E, F) = \mu(E \triangle F)$. This is the same as $\|1_E - 1_F\|_1$ but the L^1 space is convenient because it takes care of identifying sets which differ only in a nullset.

We define the action of T on the image of f by

$$T(1_{Y_x}) = 1_{TY_x}$$

Our aim is to show that f is a factor map from X to its image, giving a non-trivial isometric factor. Thus, we must prove three things:

1. The action of T on the image of f is isometric: $d(f(Tx), f(Ty)) = d(f(x), f(y))$.
2. Equivariance (i.e. that $f(Tx) = Tf(x)$).
3. Non-triviality of f (i.e. that the image of f consists of more than one point).
4. Continuity of f.

Equivariance : First notice that

$$\begin{aligned} Y_{Tx} &= \{y \ : \ (Tx.y) \in Y\} \\ &= \{y \ : \ (Tx, y) \in TY\} \qquad \text{because } Y = TY \\ &= \{y \ : \ (x, T^{-1}y) \in Y\} \\ &= \{Tz \ : \ (x, z) \in Y\} \\ &= T\{z \ : \ (x, z) \in Y\} \\ &= TY_x \end{aligned}$$

Isometry on the image :

$$\begin{aligned} d(f(Tx), f(Tx')) &= \mu(Y_{Tx} \triangle Y_{Tx'}) \\ &= \mu(TY_x \triangle TY_{x'}) \\ &= \mu(T(Y_x \triangle Y_{x'})) \\ &= \mu(Y_x \triangle Y_{x'}) \\ &= d(f(x), f(x')) \end{aligned}$$

Non-triviality : The projection π is an open map, so $\pi V, \pi W \subseteq X$ are open sets. Since they are invariant and X is minimal, they are also dense, so their intersection is non-trivial.

Fix $x \in \pi V \cap \pi W$ and denote $Z = f(x)$. Observe that $V_x \subseteq X$ and $W_x \subseteq X \setminus Z$. By our choice of x we have $V_x, W_x \neq \emptyset$, so by minimality of X, we can choose n so that $T^{-n}V_x \cap W_x \neq \emptyset$. We then have

$$T^{-n}V_x \cap W_x = T^{-n}V_x \setminus Z \subseteq Z \triangle T^{-n}Z$$

and conclude that $Z \triangle T^{-n}Z$ has non-empty interior, and hence $\mu(Z \triangle T^{-n}Z) > 0$; consequently $d(Z, T^{-n}Z) > 0$. This shows that the image of f does consists of more than one point.

Continuity of f : We first claim that, given $x_0 \in X$ and an open neighborhood U of Y_{x_0}, there exists $\delta > 0$ so that if $d(x, x_0) < \delta$ then $Y_x \subseteq U$. For if not, then there is a sequence $x_n \to x_0$ in X, and a point $y_n \in Y_{x_n} \setminus U$. Passing to a subsequence we may assume that $y_n \to y_0 \in X$, so $y_0 \notin U$. Now $(x_n, y_n) \to (x_0, y_0)$ and $(x_n, y_n) \in Y$, so $(x_0, y_0) \in Y$; but this means that $y_0 \in Y_{x_0} \subseteq U$, contradicting $y_0 \notin U$.

Fix $x_0 \in X$ and let $\varepsilon > 0$. By regularity of the measure μ, we can choose a neighborhood U of Y_{x_0} so that

$$\mu(U) < \mu(Y_{x_0}) + \varepsilon$$

Let $\delta > 0$ be as in the previous paragraph for U. Now if $d(x, x_0) < \delta$ then $Y_x \subseteq U$, so

$$\mu(Y_x) \le \mu(U) < \mu(Y_{x_0}) + \varepsilon$$

Setting $g(x) = \mu(Y_x)$, this shows that g is lower semi-continuous. Also, since $Y_{Tx} = TY_x$, we have

$$\begin{aligned} g(Tx) &= \mu(Y_{Tx}) \\ &= \mu(TY_x) \\ &= \mu(Y_x) && \text{by invariance of } \mu \end{aligned}$$

so g is invariant. By the lemm preceding the theorem, g is constant, i.e. $\mu(Y_x)$ is independent of x.

Finally, if $d(x, x_0) < \delta$, then $Y_x \subseteq U$, so

$$\begin{aligned} \mu(Y_x \setminus Y_{x_0}) &\le \mu(U \setminus Y_{x_0}) \\ &= \mu(U) - \mu(Y_{x_0}) \\ &< \varepsilon \end{aligned}$$

and similarly $\mu(Y_{x_0} \setminus Y_x) < \varepsilon$; so

$$\begin{aligned} d(f(x), f(x_0)) &= \mu(Y_x \triangle Y_{x_0}) \\ &= \mu(Y_{x_0} \setminus Y_x) + \mu(Y_x \setminus Y_{x_0}) \\ &< 2\varepsilon \end{aligned}$$

and we have shown that f is continuous. □

Corollary 6.22. *If (X, T) is a minimal system that is not weak mixing, then it admits non-trivial continuous eigenvalues.*

Proof. By the previous theorem, X has a non-trivial equicontinuous factor, which is a minimal group rotation. This factor admits a non-trivial eigenfunction (see Section **??**), and this eigenfunction lifts to X by pre-composition with the factor map. □

7 Distal systems

Equicontinuous systems, when transitive, arise from a group translation; weak mixing systems, which are precisely those which cannot factor onto non-trivial equicontinuous ones. From the point of view of orbits the difference between these systems is also evident. In an equicontinuous system, pairs of orbits cannot come close to each other, whereas in weak mixing system there exist pairs of orbits which come close to each other and close to any other pair.

The question arrises what other kinds of behavior of pairs can occur. In this section and the next, we examine distal systems, which are an important generalization of equicontinuity. We shall eventually see that although they fall strictly betwen the two extremes above, minimal distal systems nevertheless enjoy an explicit and elegant description as (possibly infinite) "towers" build from isometric components.

7.1 Distal systems

Definition 7.1. A dynamical ystem (X,T) is *distal* if it contains no proximal pairs, i.e., for every distinct pair $x,y \in X$, there exists a $\delta = \delta(x,y) > 0$ such that $d(T^n x, T^n y) \geq \delta$ for all $n \geq 0$ (i.e. $\inf_{n \geq 0} d(T^n x, T^n y) > 0$).

Although the metric appears in the definition, the property does not depend on the metric and remains valid if the metric is replaced by an equivalent one.

Remarks and examples

1. Any isometric system is distal (take $\delta = d(x,y)$). Thus, any equicontinuous system is distal.
2. A weak mixing system (X,T) with more than one point is not distal. Indeed, there exists a transitive point $(x,y) \in X \times X$. We cannot have $x = y$ because then the orbit is contained in the closed diagonal set $\Delta = \{(t,t) \ : \ t \in X\}$, and this is not all of $X \times X$. On the other hand, by transsitivity the orbit of (x,y) comes arbitrarily close to (x,x), so $\inf_{n\in\mathbb{N}} d(T^n x, T^n y) = 0$. This shows that the system (X,T) is not distal.
3. If a symbolic system $X \subseteq \Lambda^{\mathbb{Z}}$ is distal, then it is finite. This follows bacuse if X is infinite then by Proposition **??** it contains a pair $x \neq x'$ such that $x^+ = (x')^+$, so therefore $d(T^n x, d^n x') \to 0$, and X is not distal.

Theorem 7.2. *The following are equivalent for a dynamical system (X,T):*

1. *(X,T) is distal.*
2. *Every point in $X \times X$ is uniformly recurrent (equivalently, $X \times X$ decomposes into minimal sets).*
3. *Every point in the infinite product $(X^{\times\omega}, T^{\times\omega})$ is uniformly recurrnt.*

Proof. We first prove equivalence of (1) and (2).

For the first direction, we will show that if (X,T) is distal, then it decomposes into minimal subsystems. Since distality of X implies distality of $X \times X$, it follows that the product decomposes in the same way.

So suppose that X is distal and let $x \in X$. By Theorem **??**, x is proximal to a uniformly recurrnt y. But by distality there are no non-trivial proximal pairs, so $x = y$, hence $\overline{O_T(x)}$ is minimal. This shows that every point in X belongs to a minimal subsystem, and the claim is proved.

In the other direction, suppose that $X \times X$ decomposes as above, and let $x, y \in X$ with $x \neq y$. Write $z = (x, y) \in X \times X$. Then $Z = \overline{O_{T\times T}(z)}$ is minimal. Now, $\Delta = \{(u, u) \,:\, u \in X\}$ is a subsystem of $X \times X$ so $Z \cap \Delta$ is a closed $T \times T$ invariant set. If it were not empty, then by minimality, $Z \cap \Delta = Z$, i.e. $Z \subseteq \Delta$. But then $z = (x, y) \in \Delta$, contrary to assumption. So $Z \cap \Delta = \emptyset$, and it follows that $\inf d(T^n x, T^n y) > 0$, so x, y are distal.

With regard to (3), note that $X \times X$ is a factor of $X^{\times\omega}$ so (3) implies (2). On the other hand, distality of X implies distality of $X^{\times\omega}$, so (1) and (2) imply that every point in $X^{\times\omega} \times X^{\times\omega}$ is uniformly recurrent, but this implies the same for $X^{\times\omega}$ (e.g. since it is a factor of $X^{\times\omega} \times X^{\times\omega}$). □

Corollary 7.3. *If (X, T) is distal then all of its points are uniformly recurrent.*

Proof. X is a factor of $X \times X$. Apply the previous theorem. □

Corollary 7.4. *If (X, T) is distal, then all of its factors are distal.*

Proof. Suppose $\pi : X \to Y$ is a factor. Then there is an induced factor map $\varphi : X \times X \to Y \times Y$. Since X is distal, $X \times X$ is a union of minimal systems, and their images under φ are minimal, also $Y \times Y$ is a union of minimal systems, which by the theorem again, implies that Y is distal. □

The condition that (X, T) decompose into minimal subsystems is not strong enough to ensure distality. As an example, we construct an infinite subshift which is the union of minimal subshifts; we already noted that infinite subshifts are never distal.

Consider an infinite minimal subshift $X \subseteq \{0, 1\}^{\mathbb{Z}}$, e.g. the Thue-Morse subshift. We shall construct a sequence of periodic points $x^1, x^2, \ldots \in \{0, 1\}^{\mathbb{Z}}$, with disjoint orbits $C^i = O_S(x^i)$, such that $Y = X \cup \bigcup_{i=1}^{\infty} C^i$ is closed. Then Y decomposes into disjoint minimal subsystems ($C^i \cap C^j = \emptyset$ for $i \neq j$ by construction, and $C^i \cap X = \emptyset$ because it is the intersection of two distinct minimal subshifts), and Y is not distal, because no infinite subshift is distal.

In order to construct x^i proceed as follows. Let $a_n \in L_n(X)$ have ength n (so a_n is a word that appears in some point in X) and for each n, choose b_n so that

$$a_n b_n a_n \in L(x)$$

One can always find such a word b_n because, since X is minimal, a_n appears in every point syndetically; if a_n appears in $x \in X$ at index i and at index $j > i$ then we can take $b_n = x|_{[i+|a_n|, j-1]}$.

Now take $x^n = \ldots a_n b_n a_n b_n a_n b_n \ldots$ and let C^n be the (finite) orbit of x^n. Ley $Y = X \cup \bigcup_{n=1}^{\infty} C^n$. This is clearly a union of minimal systems, and in particular is invariant..

We claim that Y is also closed. Indeed, since X is already closed, we only need to show that if $y = \lim y^n$ with $y^n \in C_{k(n)}$ then $y \in Y$. So fix such a sequence y^n. If $k(n)$ is bounded, we will have $y \in C_k$ for some k, so we may assume $k(n) \to \infty$. Now, $L_n(x^n) \subseteq L(X)$ and therefore also every word of length at most n in x^n belongs to $L(X)$; so $L_k(x^n) \subseteq L(X)$ for all $k \leq n$. Given a subword of length k of y, the word must appear y^n for arbitrarily large n, and hence in x^n for arbitrarily large n; in particular, for some n with $k(n) > k$ (because $k(n) \to \infty$), and so it belongs to $L(X)$. Since all finite subwords of y belong to $L(X)$ if follows that $y \in X$, as claimed.

We conclude with another characterization of distality:

Theorem 7.5. *Let (X, T) be a dynamical system with enveloping semigroup $\mathcal{E}$. Then X is distal if and only if $\mathcal{E}$ is a group (i.e. all elements $f \in \mathcal{E}$ are invertible and $f^{-1} \in \mathcal{E}$).*

Proof. Clearly if x, y are proximal then there exists $f \in \mathcal{E}$ with $f(x) = f(y)$, so $\mathcal{E}$ is not a group.

Now suppose that X is distal. Let $f \in \mathcal{E}$; we want to show that $\mathrm{id}_X \in \mathcal{E}f = \overline{\{T^n f\}}$.

Let $x_1, \ldots, x_n \in X$ and $\varepsilon > 0$. Consider $x = (x_1, \ldots, x_n) \in X^{\times n}$ and let $S = T^{\times n}$. Since X is distal, every point in $X^{\times n}$ is uniformly recurrent. Letting $U = \times_{i=1}^{n} B_\varepsilon(x_i)$, there is a k such that for every n there is a $0 \leq j(n) \leq k$ with $S^{j(n)}(S^n x) \in U$.

Now, for each j let $I(j) \subseteq \mathbb{N}$ denote those n such that $j(n) = j$. Then for some j_0, we have $f \in \overline{\{T^i\}}_{i \in I(j_0)}$. Thus $T^{j_0} f(x_i) \in \overline{B_\varepsilon(x_i)}$ for $i = 1, \ldots, n$. In other words, $T^{j_0} f \in U_{x_1, \ldots, x_n, \varepsilon}(\mathrm{id}_X)$. It follows that $\mathrm{id}_X \in \overline{\{T^n\}_{n \geq 0} f} = \mathcal{E}f$. □

Note that while $\mathcal{E}$ is a group when X is distal, it is still generally not a topological group.

7.2 Skew-products on the torus

Our first task is to provide interesting examples of distal systems which are not equicontinuous. Non-minimal examples are not hard to find. For instance, the unit disk in $\mathbb{C}$ with the map $re^{i\theta} \mapsto re^{i(\theta + r)}$ decomposes this way but is not distal. We leave the verificatoin as an exercise.

We restrict ourselves to a single example which is typical of a class of constructions of distal systems on tori. Let $X = \mathbb{T}^2$, let $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ and let $T : X \to X$ be

$$T(x, y) = (x + \alpha, y + x)$$

It is easy to check that this map is continuous and bijective. Notice that this is a skew product in the sense of Definition 5.12, with base $(\mathbb{T}, R_\alpha)$ and cocycle $r_x(y) = y + x$. Note that by induction,

$$\begin{aligned} T^n(x, y) &= (x + n\alpha, y + \sum_{k=1}^{n-1} (x + k\alpha)) \\ &= (x + n\alpha, y + (n-1)x + \frac{n(n-1)}{2}\alpha) \qquad (4) \end{aligned}$$

Claim 7.6. (X, T) is distal.

Proof. Let us use the metric

$$d((y, z), (y', z')) = \max\{d(y, y'), d(z, z')\}$$

for $\mathbb{T}^2$. Fix $z = (x, y)$ and $z' = (x', y')$ in $\mathbb{T}^2$, and suppose $z \neq z'$. We must show that $\rho = \inf_{n \geq 0} d(T^n z, T^n z') > 0$. If $x \neq x'$, then, since R_α is an isometry of $\mathbb{T}$, we have $d(R_\alpha^n x, R_\alpha^n x') = d(x, x')$ for all n, so $\rho \geq d(x, x') > 0$. Otherwise $x = x'$, and then by (4), for every $n \geq 0$ we have

$$d(T^n z, T^n z') = d(y, y')$$

and again $\rho > 0$. □

Claim 7.7. (X, T) is not equicontinuous.

Proof. Let $0 < \delta < 1/4$ and consider the points $z = (x, y)$ and $z' = (x + \delta, y)$ for any $x, y \in \mathbb{T}$. Then

$$d(z, z') = \delta$$

but by (4), as long as $n < 1 + 1/4\delta$ we have

$$\begin{aligned} d(T^n z, T^n z') &= \delta + d((n-1)x, (n-1)(x+\delta)) \\ &= \delta + (n-1)\delta \\ &= n\delta \end{aligned}$$

In particular for every $\delta > 0$ there exist points z, z' which are δ-close and an n such that $d(T^n z, T^n z') > 1/10$. Therefore T is not equicontinuous. □

Proposition 7.8. *(X, T) is minimal.*

Proof. Let $Y \subseteq X$ be a minimal subset and let $\pi : \mathbb{T}^2 \to \mathbb{T}$ denote projection to the first coordinate. Then $\pi(Y) \subseteq \mathbb{T}$ is a subsystem of $(\mathbb{T}, R_\alpha)$, so by minimality of the latter $\pi(Y) = \mathbb{T}$.

Let $F_\beta(x, y) = (x, y + \beta)$. Then F_β is an automorphism of X (it is clearly a continuous bijection, and one checks that $TF_\beta = F_\beta T$). Therefore, $F_\beta Y$ is a minimal subset of X for every β. It follows that $Y \cap F_\beta Y$ is either empty, or all of Y. Let

$$E = \{\beta \in \mathbb{T} \,:\, F_\beta Y = Y\}$$

This is a closed subgroup of $(\mathbb{T}, +)$. Therefore, E is either finite, or all of $\mathbb{T}$. Note that $\pi^{-1}(x)$ is a coset (translate) of E for every $x \in \mathbb{T}$.

Thus, if E is all of $\mathbb{T}$, then $\pi^{-1}(x) \cap Y = \mathbb{T}$ for all $x \in \mathbb{T}$, so $X = \mathbb{T}^2$.

It remains to rule out the possibility that E is finite. If $|E| = 1$, then $\pi^{-1}(x) \cap Y$ is a singleton for all $x \in \mathbb{T}$. This means that Y is the graph of a function and since it is closed, the function is continuous. This is impossible by the previous Lemma.

Finally, if $1 < |E| < \infty$, then E consists of rational numbres, and there is a p such that $pq = 1$ for all $q \in E$. Consider the map $\varphi : \mathbb{T}^2 \to \mathbb{T}^2$ given by $\varphi : (x, y) \mapsto (x, py)$, and let $S(x, y) = (x + \alpha, y + px)$. Then φ is a factor map and $Y' = \varphi(X)$ is a minimal subset of $(\mathbb{T}^2, S)$. Evidently, the group $E' = \{\theta \in \mathbb{T} \,:\, F_\beta Y' = Y'\}$ is just

$$E' = pE = \{e\}$$

This means that over every point $\alpha \in \mathbb{T}$ there is a unique points in Y; that is, Y' is a graph. It is furthermore a graph of a continuous function, because the graph is closed. The proof is concluded by showing that this is impossible: □

Lemma 7.9. *For every $p \geq 1$ and every continuous function $g : \mathbb{T} \to \mathbb{T}$, the graph Z of g is not S-invariant.*

Proof. Suppose there is a S-invariant graph of a continuous function g.

Let $0 < \delta < 1/8p$ be small enough that $d(g(x), g(x + \delta)) < 1/8$ for every $x \in \mathbb{T}$.

Let $x \in X$, and let $z = (x, g(x))$ and $z' = (x + \delta, g(x + \delta))$. By a similar calculation as in Claim 7.7, for $n \leq 1/4\delta$, we have

$$d(S^n z, S^n z') = \max\{\delta, d(g(x), d(x + \delta)) + np\delta\}$$

Write $S^n z = (x_n, y_n)$ and $S^n z' = (x'_n, y'_n)$. Since the orbits of z, z' remain inside the graph of g, we have $y_n = g(x_n)$ and $y'_n = g(x'_n)$. Also, $x_n = x + \alpha n$ and $x'_n = x' + \alpha n + \delta$ so

$$d(x_n, x'_n) = \delta$$

Thus for $n = [1/4\delta]$,

$$\begin{aligned} d(g(x_n), g(x'_n) &\geq d(y_n, y'_n) \\ &= np\delta + d(g(x), g(x+\delta)) \\ &\geq 1/8 \end{aligned}$$

Since we can find such points for arbitrarily small δ. This contradicts continuity of g. □

A theorem of Hardy and Littlewood

We have obtained the example we were after, but now digress to show an application to diophantine approximation. We have seen that if α is irrational, then $\mathbb{N}\beta$ is dense modulo one (this is minimality of the rotation R_α). What about thinner sets of multiples, e.g. $\{n^2\beta\}_{n\in\mathbb{N}}$? This is no longer the orbit of a point in a dynamical system, but it turns out that density does hold as long as β is irrational:

Theorem 7.10 (Hardy-Littlewood)**.** *Let p be any real polynomial with at least one non-constant coefficient irrational. Then $\{p(n)\}_{n\in\mathbb{N}}$ is dense modulo one.*

Let us demonstrate the special case of $n^2\beta$. Observe that given β, let $\alpha = 2\beta$, and consider the orbit of $(\beta, 0)$ in the system (X, T) above. By (4),

$$\begin{aligned} T^n(\beta, \beta) &= (\beta + n\alpha, \beta + (n-1)\beta + \frac{n(n-1)}{2}\alpha) \\ &= (\beta + n\alpha, \beta(1 + (n-1) + n(n-1))) \\ &= (\beta + n\alpha, n^2\beta) \end{aligned}$$

Since (X, T) is minimal, the orbit is dense, and, in particular, its projection to the second coordinate is dense in $\mathbb{T}$. This is precisely density modulo one of $n^2\beta$.

A similar proof can be given for any quadratic polynomial – by choosing α and the initial point x, y, one can ensure $T^n(x, y) = (R^n_\alpha x, p(n))$.

In order to deal with polynomials of degree d, one needs to prove minimality of the map on $\mathbb{T}^d$ given by

$$T_d(x_1, \ldots, x_d) = (x_1 + \alpha, x_2 + x_1, x_3 + x_2, \ldots, x_d + x_{d-1})$$

Then by a suitable choice of α and the initial point $(X_1, \ldots, x_d)$ the last coordinate reduces again to $p(x)$. In order to prove minimality, one can proceed by induction; but the induction step is not quite like the base step that we did above, because for $d = 1$ the map is an isometry, while for larger d it is not, as we have seen. Nevertheless a argument similar to our proof for $d = 2$ can be done. We leave the details to the exercises.

7.3 Isometric Extensions

In this section we present a structure theorem for minimal distal systems. Roughly speaking, we will see that they qre built up from isometric components.

Definition 7.11. A factor $\pi : (X,T) \to (Y,S)$ is called an **isometric extension** if there exists a continuous function $d : X \times X \to [0,\infty)$ such that

1. The restriction of d to each fiber $\pi^{-1}(y)$ is a metric.
2. If $x_1, x_2 \in \pi^{-1}(y)$ then $d(Tx_1, Tx_2) = d(x_1, x_2)$.

Note that continuity and compactness of the fibers $\pi^{-1}(y)$ imply that the restruction of d to a fiber is equivalent to the original metric on the fiber.

Examples

1. An isometric map is an isometric extension of the 1-point system, and vice verse – an isometric extension of the 1-point system is isometric.
2. The map $T(x,y) = (x+\alpha, y+x)$ on $\mathbb{T}^2$ is an isometric extension of $(\mathbb{T}, R_\alpha)$. Here the fibers are copies of $\mathbb{T}$, which we endow with the usual metric, and the fiber $\{x\} \times \mathbb{T}$ is mapped to $\{x+\alpha\} \times \mathbb{T}$ by translation by x.

   This example shows that an isometric extension of an isometric extension need not be an isometric extension itself. To see this, consider $(\mathbb{T}, R_\alpha)$ as an isometric extension of the one-point system.
3. More generally, if (Y,S) is any system and Z is a metric space, given a continuous map $f : Y \to \mathrm{isom}(Z)$, $y \mapsto f_y$, we can form the skew-product $X = Y \times Z$, $T(y,z) = (Sy, f_y(z))$, and this is an isometric extension.

Not every isometric extension arises as a skew product like the last example. Indeed, the map $(\mathbb{T}, R_\alpha) \to (\mathbb{T}, R_{2\alpha})$, $x \mapsto 2x$, is an isometric extension because each fiber contains two points, and the map between them is an isometry in the usual metric on $\mathbb{T}$. But $\mathbb{T}$ is not a product of itself with two points.

Proposition 7.12. *If (Y,S) is distal and (X,T) is an isometric extension of Y then X is distal.*

Proof. The proof is imilar to Lemma **??**. Given $x_1, x_2 \in X$ and $\pi : X \to Y$, either $\pi(x_1) \neq \pi(x_2)$, in which case by distality of Y we have $\inf_n d_Y(S^n x_1, S^n x_2) > 0$ and so also for the orbits in X, or else x_1, x_2 are in the same fiber of $\pi^{-1}(y_0)$, in which case for a function $\widetilde{d}$ as in the definition of an isometric extension, $\widetilde{d}(T^n x_1, T^n x_2) = \widetilde{d}(x_1, x_2) > 0$ and so $T^n x_1, T^n x_2$ cannot accumulate on any diagonal point (x,x), giving distality again. □

Proposition 7.13. *Suppose that $\ldots (X_n, T_n) \xrightarrow{\pi_n} (X_{n-1}, T_{n-1}) \to \ldots \to (X_0, T_0)$ is a directed system of factors with (X_0, T_0) distal and all extensions isometric. Let (X,T) denote the inverse limit. Then (X,T) is distal.*

Proof. Using the previous proposition we see by induction that all (X_n, T_n) are distal. Now if $x_1, x_2 \in X$, then there is some n such that the images of x_1, x_2 in (X_n, T_n) are distinct, and therefore $\inf_k d(T_n^k x_1, T_n^k x_2) > 0$. This immediately implies the same in X. □

Now suppose (X,T) is a minimal distal system. Being distal, it is not weak mixing, so by Theorem 6.20 it has a non-trivial equicontinuous, and hence distal, factor $X \to X_0$. Now, X may be an isometric extension of X_0, but if not, we would like to find an isometric extension of X_0 sitting between them, $X \to X_1 \to X_0$. If we can carry out this basic step, then we can iterate it, eventually exhausting X.

In order to carry out this plan we need a construction called the **relative product over a factor**. Let $\pi : (X,T) \to (Y,S)$ be a factor map and form the set

$$X \times_Y X = \{(x_1, x_2) \in X \times X \ : \ \pi(x_1) = \pi(x_2)\}$$

Evidently, this set is closed and $T \times T$-invariant, so it is a subsystem of $X \times X$. There is a natural factor $\widetilde{\pi} : X \times_Y X \to Y$ given by $\widetilde{\pi}(x_1, x_2) = \pi(x_1) = \pi(x_2)$, and notice that $\widetilde{\pi}^{-1}(y) = \pi^{-1}(y) \times \pi^{-1}(y)$. So over every point in Y we have the full product of the fiber $\pi^{-1}(y)$.

If $X \to Y$ is a factor map, we say that X is **weak mixing relative to** Y $X \times_Y X$ is transitive. Note that weak mixing is weak mixing relative to the trivial one-point factor (since then the relative product is the full product). Observe that a distal system cannot be weak mixing over any non-trivial factor.

7.4 The Distal Structure Theorem

The following theorem is an exact analog of Theorem 6.20 but "relative to a factor".

Theorem 7.14. *Let $\pi : (X,T) \to (Y,S)$ be a factor map between minimal systems. Suppose that X is not weak mixing relative to Y. Then there exists an intermediate factor $X \xrightarrow{\pi_2} Z \xrightarrow{\pi_1} Y$ such that $\pi = \pi_1\pi_2$ and such that $Z \to Y$ is a non-trivial isometric extension.*

We will not prove this, but note that when Y is the one point system this is precisely Theorem 6.20.

The last theorem is the inductive step in the following deep characterization of distal systems:

Theorem 7.15 (Furstenberg)**.** *Let X be a minimal distal system. Then there is a countable ordinal α, a family of systems $(X_i)_{i<\alpha}$ and factor maps $\pi_i : X_{i+1} \to X_i$ for $i < \alpha$, such that*

1. *Each factor map π_i is an isometric extension,*
2. *If $\beta \leq \alpha$ is a limit ordinal then X_β is the inverse limit of the directed system of factors $(X_i)_{i<\beta}$,*
3. *$X_\alpha = X$.*

8 Topological Entropy

We now take a large swing, from systems with some algebraic structure to "large", "chaotic" systems. The primary example to have in miond are full shifts and shifts of finite type. Our aim is to develop invariants that can distinguish between them.

Consider $\{0,1\}^{\mathbb{Z}}$ and $\{0,1,2\}^{\mathbb{Z}}$. Certainly the former feels "smaller" than the latter.One can make several observations:

- As topological spaces, the two are homeomorphic.

If one asks about embedding the systems one in the other, we have

- $\{0,1\}^{\mathbb{Z}}$ factors injectively **into** $\{0,1,2\}^{\mathbb{Z}}$ (in fact, since $\{0,1\} \subseteq \{0,1,2\}$, it is even a subsystem).
- $\{0,1,2\}^{\mathbb{Z}}$ cannot factor injectively **into** $\{0,1\}^{\mathbb{Z}}$ because it has more fixed points.

However, what about factoring **onto** one another?

- $\{0,1,2\}^{\mathbb{Z}}$ factors **onto** $\{0,1\}^{\mathbb{Z}}$ (by the sliding block code taking $0 \to 0$ and $1,2 \to 1$).with our present knowledge, we are unable to answer the following question:
- Can $\{0,1\}^{\mathbb{Z}}$ factor onto $\{0,1,2\}^{\mathbb{Z}}$ ..... ???

With our present knowledge we cannot answer this question.

One can also note that the obstruction to embedding these spaces into each other involves periodic points. What happens when there are none? For example

- When can (X,T) factor injectively into $\{0,1\}^{\mathbb{Z}}$ (assuming there is no topological obstruction, i.e. X is totally disconnected)?

Topological entropy, which we study in this section, provides an invariant that can answer this question in the negative.

8.1 Definition (via covers)

Let (X,T) be a topological dynamical system.

Definition 8.1. .

1. An *open cover* of X is a collection of open sets whose union is X.
2. If $\mathcal{U},\mathcal{V}$ are open covers of X their *join* is

$$\mathcal{U} \vee \mathcal{V} = \{U \cap V \ : \ U \in \mathcal{U}\,,\, V \in \mathcal{V}\}$$

it is also an open cover of X. More generally if $\mathcal{U}_i$ are open covers then

$$\bigvee_{i=1}^{n} \mathcal{U}_i = \{U_1 \cap \ldots \cap U_N \ : \ U_1 \in \mathcal{U}_1, \ldots, U_n \in \mathcal{U}_n\}$$

and this is again an open cover.

3. An open cover $\mathcal{U}$ refines an open cover $\mathcal{V}$ if every $U \in \mathcal{U}$ is a subset of some $V \in \mathcal{V}$

4. If $T : X \to X$ is a continuous map then $T^{-1}\mathcal{U} = \{T^{-1}U : U \in \mathcal{U}\}$ is an open cover.

Lemma 8.2. .

1. $T^{-1}(\mathcal{U} \vee \mathcal{V}) = T^{-1}(\mathcal{U}) \vee T^{-1}(\mathcal{V})$.
2. *If* $\mathcal{U}$ *refines* $\mathcal{V}$ *then* $T^{-1}(\mathcal{U})$ refines $T^{-1}(\mathcal{V})$.

This is an exercise.

Definition 8.3. For an open cover $\mathcal{U}$ we denote

$$N(\mathcal{U}) = \min\{|\mathcal{V}| : \mathcal{V} \subseteq \mathcal{U} \text{ is an open cover}\}$$

and

$$H(\mathcal{U}) = \log N(\mathcal{U})$$

Remark 8.4. By compactness, every open cover has a finite sub-cover, so $N(\mathcal{U}) \in \mathbb{N}$.

Lemma 8.5. .

1. $N(\mathcal{U}) \geq 1$ *and* $H(\mathcal{U}) \geq 0$*, with equality if and only if* $X \in \mathcal{U}$
2. $\mathcal{U}$ *refines* $\mathcal{V}$ *implies* $N(\mathcal{U}) \geq N(\mathcal{V})$ *and* $H(\mathcal{U}) \geq H(\mathcal{V})$.
3. $H(\mathcal{U} \vee \mathcal{V}) \leq H(\mathcal{U}) + H(\mathcal{V})$.
4. $H(T^{-1}\mathcal{U}) \leq H(\mathcal{U})$ *and if* T *is onto then equality.*

Theorem 8.6. $\lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U})$ *exists for every open cover* $\mathcal{U}$ *of* X*, and the limit is equal to the infimum of the sequence.*

Proof. Write $a_n = H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U})$. Then

$$\begin{aligned} a_{m+n} &= H(\bigvee_{i=0}^{(m+n)-1} T^{-i}\mathcal{U}) \\ &= H(\bigvee_{i=0}^{m-1} T^{-i}\mathcal{U} \vee \bigvee_{i=m}^{(m+n)-1} T^{-i}\mathcal{U})) \\ &\leq H(\bigvee_{i=0}^{m-1} T^{-i}\mathcal{U}) + H(\bigvee_{i=m}^{(m+n)-1} T^{-i}\mathcal{U})) \\ &\leq H(\bigvee_{i=0}^{m-1} T^{-i}\mathcal{U}) + H(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U})) \\ &= a_m + a_n \end{aligned}$$

and the claim follows from sub-additivity, using the following: □

Lemma 8.7 (Fekete's lemma)**.** *Let* $(a_n)_{n=1}^{\infty}$ *be a sequence satisying*

$$a_{m+n} \leq a_m + a_n$$

Then a_n/n *converges and* $\lim \frac{1}{n}a_n = \inf_n \frac{1}{n}a_n$.

Proof. We prove this in case the sequence is bounded below (this is the case in our application to entropy). When it is not bounded the proof is similar.

Let

$$\alpha = \inf_{n \in \mathbb{N}} \frac{1}{n}a_n$$

Let $\varepsilon > 0$ and let n_0 be such that $a_{n_0}/n_0 < \alpha + \varepsilon$. For any $n \geq n_0$ write $n = kn_0 + r$ with $0 \leq r < n_0$. Then

$$\begin{aligned} a_n &\leq a_{n-n_0} + a_{n_0} \\ &\leq a_{n-2n_0} + 2a_{n_0} \\ &\ldots \\ &\leq a_r + ka_{n_0} \\ &= a_r + kn_0 \cdot \frac{1}{n_0}a_{n_0} \end{aligned}$$

Writing $c = \max\{a_0, \ldots, a_{n_0-1}\}$, noting that $k \leq n/n_0$, and using $a_{n_0}/n_0 < \alpha + \varepsilon$ we conclude that

$$a_n < c + n(\alpha + \varepsilon)$$

dividing by n we have

$$\frac{1}{n}a_n \leq \alpha + \varepsilon + \frac{c}{n}$$

so $\limsup \frac{1}{n}a_n \leq \alpha + \varepsilon$ and since $\varepsilon > 0$ is arbitrary, $\limsup \frac{1}{n}a_n \leq \alpha$. Of course $\liminf \frac{1}{n}a_n \geq \alpha$ since α is the infimum of the sequence, and we conclude that $\lim \frac{1}{n}a_n = \alpha$. □

Definition 8.8. The *topological entropy* of (X, T) and an open cover $\mathcal{U}$ is

$$h_{top}(T, \mathcal{U}) = \lim_{n\to\infty} \frac{1}{n} H(\bigvee_{i=-0}^{n-1} T^{-i}\mathcal{U})$$

Proposition 8.9. .

1. $0 \leq h_{top}(T, \mathcal{U}) \leq H(\mathcal{U})$.
2. *If* $\mathcal{U}$ *refines* $\mathcal{V}$ *then* $h_{top}(T, \mathcal{U}) \geq h_{top}(T, \mathcal{V})$.

Proof. (1) follows from the fact that the limit in the definition of entropy is the infimum of a sequence of which $H(\mathcal{U})$ is the first term.

(2) follows from the fact the if $\mathcal{U}$ refines $\mathcal{V}$ then $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}$ refines $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{V}$. □

Definition 8.10. The *topological entropy* of (X, T) is

$$h_{top}(T) = \sup\{h_{top}(T, \mathcal{U}) \quad : \quad \mathcal{U} \text{ is an open cover of } X\}$$

Sometimes we write $h_{top}(X)$ instead of $h_{top}(T)$ when T is fixed, and the set X varies (e.g. when T is the shift and we vary over sub-systems).

Remark 8.11. We can take the sup over finite sub-covers.

Proposition 8.12. .

1. $h_{top}(T) \geq 0$ (obvious).
2. If T is invertible then $h_{top}(T) = h_{top}(T^{-1})$ (this is an exercise!)

Theorem 8.13. *If $Y \subseteq X$ is a subsystem then $h_{top}(T) \geq h_{top}(T|_Y)$.*

Proof. Let $\mathcal{U}$ be an open cover of Y. Each $U \in \mathcal{U}$ is relatively open in Y so there exists an open set $U' \subseteq X$ such that $U = U' \cap Y$. Let $\mathcal{U}' = \{U'\}_{U \in \mathcal{U}} \cup \{X \setminus Y\}$. This is an open cover of X. Note that $T^{-k}\mathcal{U}' = \{T^{-k}U\}_{U \in \mathcal{U}} \cup \{T^{-k}(X \setminus Y)\}$, and $T^{-k}(X \setminus Y) \subseteq X \setminus Y$ because by assumption $TY \subseteq Y$. Thus $T^{-k}(X \setminus Y) \cap Y = \emptyset$, so if $V \in \bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}'$ intersects Y non-trivially, it must be of the form $\bigcap_{i=0}^{n-1} T^{-i}U'_i$ (otherwise there is a set of the form $T^{-i}(X \setminus Y)$ in the intersection which is impossible). Thus if $\{V'_j\} \subseteq \bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}'$ is a subcover then $\{V_j \cap Y\}$ is an open cover of Y and consists of sets of the form $(\bigcap_{i=0}^{n-1} T^{-i}U'_i) \cap Y = (\bigcap_{i=0}^{n-1} T^{-i}U_i) \cap Y$. Thus the number of V'_j must be at least as large as $N(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U})$, which implies that $h_{top}(T) \geq h_{top}(T, \mathcal{U}') \geq h_{top}(T|_Y, \mathcal{U})$. The claim follows. □

Theorem 8.14. *If (Y, S) is a factor of (X, T) then $h_{top}(T) \geq h_{top}(S)$.*

Proof. Let $\pi : X \to Y$ be a factor map. If $\mathcal{U}$ is an open cover of Y then $\pi^{-1}\mathcal{U} = \{\pi^{-1}U : U \in \mathcal{U}\}$ is an open cover of X and $N(\pi^{-1}\mathcal{U}) = N(\mathcal{U})$. Also $\pi^{-1}(\bigvee_{i=0}^{n-1} S^{-i}\mathcal{U}) = \bigvee_{i=0}^{n-1} T^{-i}\pi^{-1}\mathcal{U}$. Combining these two facts we find that $h_{top}(T, \pi^{-1}\mathcal{U}) = h_{top}(S, \mathcal{U})$. This shows that

$$h_{top}(T) = \sup_{\mathcal{V}} h_{top}(T, \mathcal{V}) \geq \sup_{\mathcal{U}} h_{top}(S, \mathcal{U}) = h_{top}(S)$$ □

Corollary 8.15. *Isomorphic systems have the same topological entropy.*

8.2 Expansive systems

Recall: (X, T) is (forward) expansive if there is an $\varepsilon > 0$ such that for every $x, y \in X$ with $X \neq y$ there is an $n \in \mathbb{N}$ such that $d(T^n x, T^n y) > \varepsilon$. It is two-sided expansive if T is invertible and the same holds but allowing $n \in \mathbb{Z}$. The constant ε is called the expansiveness constant.

Lemma 8.16. *If ε is as in the definition of expansiveness, then for every $\delta > 0$ there is an $N = N(\delta)$ such that if $x, y \in X$ and $d(x, y) \geq \delta$ then there is an $n \in \{0, 1, \ldots, N-1\}$ with $d(T^n x, T^n y) > \varepsilon$.*

Proof. If not then there is some $\delta > 0$ such that for every N there is a pair $x_N, y_N \in X$ with $d(x_N, y_N) \geq \delta$ and $d(T^n x_k, T^n y_k) \leq \varepsilon$ for all $0 \leq n < N$. Passing to subsequence we can assume that $x_{N_k} \to x$ and $y_{N_k} \to y$. Evidently $d(x, y) \geq \delta$, so $x \neq y$, but for every n we have $n < N_k$ for all large k and by continuity of T, $d(T^n x, T^n y) = \lim d(T^n x_{N_k}, T^n y_{N_k}) \leq \varepsilon$. This contradicts expansiveness. □

Lemma 8.17. *For any cover $\mathcal{U}$ and any N, $h_{top}(T,\mathcal{U}) = h_{top}(T, \bigvee_{i=0}^{N-1} T^{-i}\mathcal{U})$.*

Proof. Since $\bigvee_{i=0}^{N-1} T^{-i}\mathcal{U}$ refines $\mathcal{U}$ we certainly have $\leq$. For the other direction write $\mathcal{V} = \bigvee_{i=0}^{N-1} T^{-i}\mathcal{U}$ and notice that

$$\bigvee_{i=0}^{M-1} T^{-i}\mathcal{V} = \bigvee_{i=0}^{(N+M)-1} T^{-i}\mathcal{U}$$

hence

$$\begin{aligned} h_{top}(T,\mathcal{V}) &= \limsup \frac{1}{n} \log N(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{V}) \\ &= \limsup \frac{1}{n} \log N(\bigvee_{i=0}^{n+N-1} T^{-i}\mathcal{U}) \\ &= h_{top}(T,\mathcal{U}) \quad \square \end{aligned}$$

Proposition 8.18. *If (X,T) is expansive with expansive constant ε, and $\mathcal{U}$ is a cover of X by sets of diameter $\leq \varepsilon$, then $h_{top}(T) = h_{top}(T,\mathcal{U})$.*

Proof. It suffices to show that for every open cover $\mathcal{V}$ we have $h_{top}(T,\mathcal{U}) \geq h_{top}(T,\mathcal{V})$.

Let δ be a Lebesgue covering number of $\mathcal{V}$, so for every $x \in X$ we have $\overline{B_\delta}(x) \subseteq V$ for some $V \in \mathcal{V}$.

Let $N = N(\delta)$ be as in the lemma and $\mathcal{U}' = N(\bigvee_{i=0}^{N-1} T^{-i}\mathcal{U})$. We claim that every element of $\mathcal{U}'$ has diameter $\leq \delta$. Indeed, if $d(x,y) \geq \delta$ then there is some $0 \leq n < N$ with $d(T^n x, T^n y) > \varepsilon$, and hence $T^n x, T^n y$ cannot both belong to the same element of $\mathcal{U}$, hence x, y do not belong to the same element of $T^{-n}\mathcal{U}$. This shows that every x, y which belong to the same element of $\mathcal{U}'$ satisfy $d(x,y) < \delta$ as claimed.

It follows that $\mathcal{U}'$ refines $\mathcal{V}$, hence $h_{top}(\mathcal{U}') \geq h_{top}(\mathcal{V})$. But $h_{top}(\mathcal{U}) = h_{top}(\mathcal{U}')$ be the previous lemma and the proposition follows. $\square$

Corollary 8.19. *An expansive map has finite topological entropy.*

Example

Let $X = A^{\mathbb{N}}$ for a finite set A and T the shift. Then $h_{top}(T) = \log |A|$.

Indeed, define the metric by

$$d(x,y) = 2^{-n} \quad \text{where} \quad n = \min\{i \in \mathbb{N} \,:\, x_i \neq y_i\}$$

Note that if $x_1 \neq y_1$ then $d(x,y) \geq \frac{1}{2}$. Since $x \neq y$ implies that $x_n \neq y_n$ for some n, and $(T^n x)_1 = x_n \neq y_n = (T^n y)_1$, we have $d(T^n x, T^n y) \geq \frac{1}{2}$, so T is expansive with constant $\frac{1}{2}$. Also note that if $x_1 = y_1$ then $d(x,y) \leq \frac{1}{4}$, so the cylinder sets

$$[a] = \{x \in X \,:\, x_1 = a\}$$

are open (and closed) sets of diameter $\frac{1}{4}$. By the proposition, $h_{top}(T) = h_{top}(T,\mathcal{U})$ for the partition $\mathcal{U} = \{[a] \,:\, a \in A\}$. Finally, $\bigvee_{i=1}^{n} T^{-i}\mathcal{U}$ is the partition of X according tot he initial n-segments of sequences $x \in X$ and consists of $|A|^n$ pairwise disjoint sets, so it has no proper subcovers and $N(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}) = |A|^n$. Thus $h_{top}(T,\mathcal{U}) = \log |A|$, as claimed.

Corollary 8.20. *Let A, B be finite sets. If $|B| > |A|$ then there is no factor map from $A^{\mathbb{Z}} \to B^{\mathbb{Z}}$.*

Corollary 8.21. *Let A be a finite set. If (X,T) is a system and $h_{top}(T) > \log|A|$ then there is no injective factor map $X \to A^{\mathbb{Z}}$.*

Proof. If $\pi : X \to A^{\mathbb{Z}}$ is an injective factor map let $Y = \pi(X)$. This is a subsystem of $A^{\mathbb{Z}}$ and is isomorphic to X via π, so

$$h_{top}(X) = h_{top}(Y) \leq h_{top}(A^{\mathbb{Z}}) = \log|A|$$

□

Example 8.22. Let A be finite and $X \subseteq A^{\mathbb{Z}}$a sybsystem. Let

$$L_n(X) = \#\{w \in A^n : w \text{ appears in } X\}$$

Then $h_{top}(T|_X) = \lim_{n\to\infty} \frac{1}{n} \log L_n(X)$.

Indeed, T is expansive with the same constant as before so for the partition $\mathcal{U}$ into cylinders $[a] \cap X$, $a \in A$, we have again $h_{top}(T|_X) = \lim \frac{1}{m} \log N(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U})$. But $N(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}) = L_n(X)$ and the claim follows.

8.3 Spanning and separating sets

Definition 8.23. Let (X,d) be a compact metric space and $\varepsilon > 0$.

1. The ε-covering number $cov(X,d,\varepsilon)$ is the minimal number of points in an ε-dense set, i.e.

$$cov(X,d,\varepsilon) = \min\{n : \exists x_1,\ldots,x_n \in X \text{ s.t. } X = \bigcup_{i=1}^{n} \overline{B_\varepsilon}(x_i\})$$

2. The ε-separation number, $sep(X,d,\varepsilon)$, is the maximal number of ε-separated points, i.e.

$$sep(X,d,\varepsilon) = \max\{n : \exists y_1,\ldots,y_n \in X \,\text{s.t.}\; d(y_i,y_j) > \varepsilon \text{ for all } i \neq j\}$$

Remarks

1. By compactness, both numbers are finite.
2. If $\varepsilon' < \varepsilon$ then $cov(X,d,\varepsilon') \geq cov(X,d,\varepsilon)$ and $sep(X,d,\varepsilon') \geq sep(X,d,\varepsilon)$.

Lemma 8.24. $cov(X,d,\varepsilon) \leq sep(X,d,\varepsilon) \leq cov(X,d,\varepsilon/2)$

Proof. Suppose that $x_1,\ldots,x_n$ is a maximal ε-separated set, so $n = sep(X,d,\varepsilon)$. If $X \not\subseteq \bigcup \overline{B_\varepsilon}(x_i)$ there is an $x \in X$ such that $d(x,x_i) \geq > \varepsilon$ for all i and then $x_1,\ldots,x_n,x$ would also be ε-separated, contradicting maximality. Hence $X = \bigcup B_\varepsilon(x_i)$ and $cov(X,d,\varepsilon) \leq n = sep(X,d,\varepsilon)$.

On the other hand if $X = \bigcup_{i=1}^{m} \overline{B_{\varepsilon/2}}(y_i)$ then for any ε-separated set $x_1,\ldots,x_n$, no two of the points x_i are in the same ball $\overline{B_\varepsilon}(y_j)$, but each x_i is in at least one such ball, hence $n \leq m$. It follows that $cov(X,d,\varepsilon/2) \geq sep(X,d,\varepsilon)$. □

8.4 Bowen's definition of entropy

Definition 8.25. If (X,T) is a topological dynamical system, d a metric on X, then we define

$$d_n(x,y) = \max_{0\leq i\leq n-1} d(T^n x, T^n y)$$

This is a new metric on X.

Observe that the ε-ball around x in d_n is $\bigcap_{i=0}^{n-1} B_\varepsilon(T^n x)$.

Definition 8.26. For $\varepsilon > 0$, let

$$\begin{aligned} h_{sep}(T,d,\varepsilon) &= \limsup_{n\to\infty} \frac{1}{n} \log sep(X, d_n, \varepsilon) \\ h_{cov}(T,d,\varepsilon) &= \limsup_{n\to\infty} \frac{1}{n} \log cov(X, d_n, \varepsilon) \end{aligned}$$

Also let

$$\begin{aligned} h_{sep}(T,d) &= \lim_{\varepsilon\to 0} h_{sep}(T,\varepsilon) \\ &= \sup_{\varepsilon\to 0} h_{sep}(T,\varepsilon) \end{aligned}$$

and

$$\begin{aligned} h_{cov}(T,d) &= \lim_{\varepsilon\to 0} h_{cov}(T,\varepsilon) \\ &= \sup_{\varepsilon\to 0} h_{cov}(T,\varepsilon) \end{aligned}$$

Note: Since $cov(X,d,\varepsilon/2) \geq sep(X,d,\varepsilon) \geq cov(X,d,\varepsilon)$ we have

$$h_{cov}(T,d,\varepsilon) \leq h_{sep}(T,d,\varepsilon) \leq h_{cov}(T,d,\varepsilon/2)$$

so

$$h_{sep}(T,d) = h_{cov}(T,d)$$

Lemma 8.27. *$h_{sep}(T), h_{cov}(T)$ are independent of the metric (depend only on the topology).*

Proof. Let d, d' be two metrics compatible with the topology on X. For every $\varepsilon > 0$ there is an $\varepsilon' > 0$ such that if $d'(x,y) < \varepsilon'$ then $d(x,y) < \varepsilon$. Thus $B'_{\varepsilon'}(x) \subseteq B_\varepsilon(x)$, where B' denotes the ball with respect to d'. It follows that $cov(X,d',\varepsilon') \geq cov(X,d,\varepsilon)$ and $cov(X,d'_n,\varepsilon') \geq cov(X,d_n,\varepsilon)$. Hence $h_{cov}(T,d',\varepsilon') \geq h_{cov}(T,d,\varepsilon)$. Hence

$$h_{cov}(T,d') = \sup_{\varepsilon'} h_{cov}(T,d,\varepsilon') \geq \sup_{\varepsilon} h_{cov}(T,d,\varepsilon) = h_{cov}(T,d)$$

The other inequality is symmetric. The claim about h_{sep} follows from the fact that it is the same as h_{cov}. □

In view of the last lemma, from now on we drop the metric from the notation and write $h_{cov}(T), h_{sep}(T)$.

Example 8.28. If T is an isometry, then $d_n = d$. Hence $cov(X,d_n,\varepsilon) = cov(X,d,\varepsilon)$ is independent of n and $\frac{1}{n} cov(X,d_n,\varepsilon) \to 0$. Taking ε also, we have $h_{cov}(T) = 0$.

8.5 Equivalence of the definitions

For an open cover $\mathcal{U}$ write

$$\operatorname{diam}\mathcal{U} = \max\{\operatorname{diam} U \,:\, U \in \mathcal{U}\}$$

.

Proposition 8.29. *Let* $\mathcal{U}_n$ *be open covers with* $\operatorname{diam}\mathcal{U}_n \to 0$*. Then*

$$h_{top}(T) = \lim_{n\to\infty} h_{top}(T, \mathcal{U}_n)$$

Proof. First, for any open cover $\mathcal{V}$, let δ be a Lebesgue number for $\mathcal{V}$. Then for large enough n we have that $\operatorname{diam}\mathcal{U}_n < \delta$ so $\mathcal{U}_n$ refines $\mathcal{V}$ and $h_{top}(T,\mathcal{U}_n) \geq h_{top}(\mathcal{V})$. In particular, taking $\mathcal{V} = \mathcal{U}_{n_0}$, this shows that $\lim h_{top}(T,\mathcal{U}_n)$ exists, and that the limit is at least as large as $\sup_{\mathcal{V}} h_{top}(T,\mathcal{V})$. Since it also does not exceed this supremum and the supremum is equal by definition to $h_{top}(T)$, we are done. $\square$

Proposition 8.30. *If* $\mathcal{U}$ *is an open cover with Lebesgue number* δ *then*

$$N(\bigvee_{i=1}^{n-1} T^{-i}\mathcal{U}) \leq cov(X, d_n, \delta) \leq sep(X, d_n, \delta)$$

Proof. We have already seen the right inequality. For the left one, notice that in the metric d_n the open cover $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}$ has Lebesgue number δ. Therefore if $\mathcal{U}_n$ is an optimal cover of (X, d_n) by $\delta/2$ balls, then its elements have diameter δ and it refines $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}$. Thus $N(\bigvee_{i=1}^{n-1} T^{-i}\mathcal{U}) \leq N(\mathcal{U}_n) = cov(X, d_n, \delta/2)$. $\square$

Proposition 8.31. *If* $\mathcal{U}$ *is an open cover with* $\operatorname{diam}\mathcal{U} \leq \varepsilon$*, then*

$$cov(X, d_n, \varepsilon) \leq sep(X, d_n, \varepsilon) \leq N(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U})$$

Proof. The left inequality was already proved. For the right one, note that if $x_1, \ldots, x_m$ is ε-separated in d_n then for each x_i, x_j there is some $0 \leq k \leq n-1$ such that $d(T^k x_i, T^k x_j) > \varepsilon$. This means that $T^k x_i, T^k x_j$ do not lie in a common element of $\mathcal{U}$, equivalently x_i, x_j do not lie in a common element of $T^{-k}\mathcal{U}$, so they do not lie in a common element of $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}$. This means that a sub-cover of $\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}$ must contain at least m sets. Taking a maximal separated set, with $m = sep(X, d_n, \varepsilon)$, we find that $N(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{U}) \geq sep(X, d_n, \varepsilon)$. $\square$

Theorem 8.32. $h_{top}(T) = h_{sep}(T) = h_{cov}(T)$.

Proof. Let $\mathcal{U}_n$ be open covers with $\operatorname{diam}\mathcal{U}_n < 1/n$, so

$$h_{top}(T) \;=\; \lim_{n\to\infty} h_{top}(T, \mathcal{U}_n)$$

Now for each n, by the previous proposition with $\varepsilon = 1/n$,

$$\begin{aligned} h_{top}(T,\mathcal{U}_n) &= \limsup_{N\to\infty} \frac{1}{N} \log N(\bigvee_{i=0}^{N-1} T^{-i}\mathcal{U}_n) \\ &\geq \limsup_{N\to\infty} \frac{1}{N} \log cov(X, d_n, 1/n) \\ &= h_{cov}(T, d, 1/n) \end{aligned}$$

so taking $n \to \infty$ we conclude

$$h_{top}(T) \geq h_{cov}(T)$$

On the other hand let δ_n be the Lebesgue covering number of $\mathcal{U}_n$ and note that $\delta_n \leq \operatorname{diam} \mathcal{U}_n \to 0$. Then by the other proposition,

$$\begin{aligned} h_{top}(T, \mathcal{U}_n) &= \limsup_{N \to \infty} \frac{1}{N} \log N(\bigvee_{i=0}^{N-1} T^{-i} \mathcal{U}_n) \\ &\leq \limsup_{N \to \infty} \frac{1}{N} \log cov(X, d_n, \delta_n/2) \\ &= h_{cov}(T, d, \delta_n/2) \end{aligned}$$

again taking $n \to \infty$ we obtain

$$h_{top}(T) \leq h_{cov}(T)$$

as claimed. □

9 Shifts of finite type and the Krieger embedding theorem

A **shift of finite type** is a subshift defined as follows. Let $G = (V, E)$ be a directed graph. Let $X = X_G$ denote the set of directed vertex paths; i.e.

$$X = \{(x_n) \in V^{\mathbb{Z}} \,:\, (x_n, x_{n+1}) \in E\}$$

(one can also define the one-sided version but we don't). It is easy to verify that

1. X is shift invariant: if $x \in X$ then adjacent coordinates of x are edges, and this property is shift invariant (clearly holds for Sx).
2. X is closed. Indeed, $x \in X$ if and only if $S^n x \notin [uv]$ for every $uv \in E^c$, so $X = V^{\mathbb{Z}} \setminus \bigcup_{uv \notin E} \bigcup_{n \in \mathbb{Z}} S^{-n}[uv]$. This is the complement of an open set, so X is closed.
3. X is non-empty if and only if there are arbitrarily long paths in G, if and only if there is a cycle (closed path) in G.

For example, the system X from the previous example is a shift of finite type. Indeed, we can take $V = \{0, 1\}$ and $E = \{(0, 0), (0, 1), (1, 0)\}$. Then paths through $G = (V, E)$ are precisely sequences of $0, 1$ with no two consecutive 1s.

9.1 Strong irreducibility

Let $G = (V, E)$ and X_G be as above – the set of directed bi-infinite paths in G.

For $u, v \in V$, we write $u \xrightarrow{n} v$ if there exists a path $u = u_0 u_1 \dots u_{n-1} = V$ of length n starting at u and ending at v (so always $u \xrightarrow{0} u$). Connectedness means that $\forall u, v \in V\ \exists n \in \mathbb{N}\ u \xrightarrow{n} v$. The following is stronger:

Definition 9.1. A directed graph $G = (V, E)$ is **strongly irreducible** if there exists an $N = N_G$ such that $u \xrightarrow{n} v$ for all $u, v \in V$ and all $n \geq N$.

Let $A = (a_{i,j})_{i,j \in V}$ be the adjacency matrix of G, so $a_{i,j} = 1$ if $(u, v) \in E$ and $a_{i,j} = 0$ otherwise. Then $(A^n)_{i,j}$ is the number of paths of length n from i to j. We say that a matrix is positive if all its entries are positive. Then strong irreducibility says that there exists N such that $A^n > 0$ for $n \geq N$. A weaker condition is that $A^N > 0$ for some N. The weaker condition implies the stronger one:

Lemma 9.2. *G is strongly irreducible if and only if there exists an N such that $u \xrightarrow{N} v$ for all $u, v \in V$.*

Proof. One direction is clear. Suppose then that $u \xrightarrow{N} v$ for all $u, v \in V$ and some N. This just says that $A^N > 0$. Now, for $k \geq 0$ we have $A^{N+k} = A^N A^k$. The matrix A^k has non-negative entries and each column has at least one non-zero entry, because by connectedness every v certainly has at least one length-k path ending at it. Every row of A^N is positive, thus for every $u, v \in V$ the entry $(A^N A^k)_{u,v}$ is the inner product of a positive vector and a non-zero non-negative vector, so $A^{N+k} = A^N A^k > 0$. Since $k \geq 0$ was arbitrary, we have strong ireducibility. □

Example 9.3. A cyclic graph ($V = \mathbb{Z}/m\mathbb{Z}$ and $E = \{(n, n+1) \bmod 1\}$) is not strongly irreducible, since any path from v to v is of length km for some $k \in \mathbb{N}$.

Example 9.4. If G is connected and there exists $u_0 \in V$ with a loop ($(u_0, u_0) \in E$), then G is strongly irreducible. Indeed, by conectedness there is an N_0 such that for every $v \in V$ there $m(v), n(v)$ such that $v \xrightarrow{m(v)} u_0$ and $u_0 \xrightarrow{n(v)} v$. Then for any v, w, we can form a path

$$v \xrightarrow{m(v)} u_0 \xrightarrow{2N_0 - m(v) - n(w)} u_0 \xrightarrow{n(w)} w$$

where in the middle segment we traverse the loop $2N_0 - m(v) - n(w)$ times. This is a path of length $N = 2N_0$, and v, w were arbitrary, so by the lemma, G is strongly irreducible.

Fact 9.5. *One can show that the first example is, in a way, the only obstruction to stong irreducibility: if G is connected but not strongly irreducible, then there exists an $m \geq 2$ and a partition $V = V_0 \cup \ldots \cup V_{m-1}$ such that if $u \in V_i$ and $(u, v) \in E$ then $v \in V_{i+1 \bmod m}$. That is, G "factors" onto a cycle.*

One can further show that X_G factors onto a cycle (as a dynamical system) if and only if G is strongly irreducible.

9.2 Entropy of shifts of finite type

Theorem 9.6. *Suppose that $G = (V, E)$ is strongly irreducible, $|V| \geq 2$, and let $X = X_G$. Then $h_{top}(X) > 0$.*

Proof. Let $u_0 \in V$ and let $a = u_0 u_1 \ldots u_{k-1}$ be a path from u_0 to itself (by a cycle we mean that $(u_{k-1}, u_0) \in E$), By strong irreducibility, there is an $0 \leq i < k$ such that u_i has more than one edge going out of it; otherwise, the only path through u_0 is the cycle a, and by connectedness this is the whole graph, contradicting strong irreducibility. Let $b = u'_0 \ldots u'_{m-1}$ be another simple cycle starting at u_0, such that $u_0 = u'_0, \ldots, u_i = u'_i$ and $u_{i+1} \neq u'_{i+1}$ (we can choose such u'_{i+1} by by choice of i, and then return to u_0 by conectedness).

Now it is simple to see that each concatenation $c_1 c_2 \ldots c_n$ with $c_j \in \{a, b\}$ is a legal path in G, and hence a word appearing in X_G.

Assume, without loss of generality, that $k \geq m$. Then for any ℓ we can write $\ell = \ell' k + n$ where $n \leq k$, and note that $\ell' = \ell/k + o(1)$ as $\ell \to \infty$. Every concatenation $c_1 \ldots c_{\ell'}$ as above can be extended to a path of length ℓ in G, and we conclude from the discussion that

$$|L_\ell(X_G)| \geq 2^{\ell'} = 2^{\ell/k + o(1)}$$

Thus

$$h_{top}(X_G) = \lim_{\ell \to \infty} \frac{1}{\ell} \log |L_\ell(X_G)| \geq \frac{1}{k} > 0$$

∎

We can be more precise. For a matrix A let $\lambda_1, \ldots, \lambda_{|V|}$ denote the eigenvalues of the adjacency matrix A of G (listed with multiplicity) and let

$$\lambda_{\max} = \lambda_{\max}(A) = \max_i |\lambda_i|$$

.

Lemma 9.7. *For any $|V| \times |V|$ matrix A, we have $\lim_{n\to\infty} \frac{1}{n} \log \|A^n\| = \lambda_{\max}(A)$ (it does not matter which norm we take since all norms on $M_{|V|}(\mathbb{R})$ are equivalent).*

Proof. First suppose that A is in Jordan form, so it has λ_is on the diagonal, some 1s on the diagonal above the main one, and the rest is zeros. Each Joedan block of dimension $\ell \times \ell$ has the form

$$\lambda I + N$$

for $\lambda \in \{\lambda_i\}$ and N a matrix with 1 above the diagonal and all other entries 0, so that $N^{\ell+1} = 0$. Then A^n is a block matrix and the corrsponding block is (using commutation of I and N):

$$\begin{aligned}
(\lambda I + N)^n &= \sum_{k=0}^{n} \binom{n}{k} \lambda^{n-k} N^k \\
&= \sum_{k=0}^{\ell} \binom{n}{k} \lambda^{n-k} N^k \\
&= \lambda^n p(n)
\end{aligned}$$

where $p(n)$ is a non-zero matrix whose entries are polynomial polynomial in n and whose constants depend on λ and ℓ. Since the entries of the matrix grow as λ^n up to a polynomial correction (and other entries are 0), it follows that

$$\lim_{n\to\infty} \frac{1}{n} \log \|(\lambda I + N)^n\|_\infty = \log \lambda$$

Since these blocks make up A^n, we find that

$$\lim_{n\to\infty} \frac{1}{n} \log \|A^n\|_\infty = \log \lambda_{\max}$$

In general let $C = B^{-1}AB$ be the Jordan for of A, with B a complex matrix. Define a norm on $M_{|V|}(\mathbb{R})$ by $\|U\| = \left\|B^{-1}UB\right\|_\infty$. Then by the above $\frac{1}{n} \log \|A^n\| \to \lambda_{\max}$. As noted in the statement, it does not matter which norm we use, so this proves the claim. □

Theorem 9.8. *If G is strongly irreducible and $X = X_G$ then $h_{top}(X) = \log \lambda_{\max}$.*

Proof. Let A be the adjacency matrix of G. Then the number of words of length n in X is just the numberof paths of length n in G, which is the sum of entries of A^n. . This sum is just $\|A^n\|_1$, so by the lemma (using the fact that the entropy of G ispositive to deduce $\lim \frac{1}{n} \log \|A\|^n > 0$),

$$\frac{1}{n} \log |L_n(X)| = \frac{1}{n} \log \|A^n\|_1 \to \log \lambda_{\max}$$

as claimed. □

9.3 Subsystems of shifts of finite type

A shift of finite type certianly has subsystems: for example take any cycle, it gives a periodic point in X_G. It is elementay to show that

Proposition 9.9. *If G is connected then the periodic points are dense in X_G and X_G is transitive.*

We now consider the subsystems of shifts of finite type, and specifically their entropy (which is a measure of how "large" they are). First, every subsystem has less-or-equal entropy than the original. For shifts of finite type, equality is not an option.

We first need a combinatorial fact. Let $h : [0, 1] \to \mathbb{R}$ denote the function

$$h(t) = -t \log t - (1 - t) \log(1 - t)$$

Then $h(t) \geq 0$ for $t \in (0, 1)$ and we extend it to the endpoints $0, 1$ by continuity. Then $h(0) = h(1) = 0$.

For $\varepsilon > 0$ and a finite set I write $J \sim \varepsilon I$ if $J \subseteq I$ and $\varepsilon|I| \leq |J| < \varepsilon|I| + 1$.

Lemma 9.10. *Let I be a finite set and $0 < \varepsilon < 1/2$. Then*

$$\#\{J \subseteq I \,:\, J \sim \varepsilon I\} = 2^{h(\varepsilon)n+o(n)}$$

Proof. By Stirling's formula, $k! = (\frac{n}{e})^n e^{o(n)}$. Let $n = |I|$ and $k = \lceil \varepsilon n \rceil$. Then

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
&= \frac{(n/e)^n}{(k/e)^k((n-k)/e)^{n-k}} e^{o(n)-o(n-k)-o(k)} \\
&= 2^{n \log n - k \log k - (n-k)\log(n-k) - o(n)} \\
&= 2^{-k \log(k/n) - (n-k)\log((n-k)/n) - o(n)} \\
&= 2^{nh(\varepsilon)+o(n)}
\end{aligned}$$

using the fact that $k/n = \varepsilon n + o(1)$ and $(n - k)/n = (1 - \varepsilon)n + o(1)$. □

Theorem 9.11. *Let $X = X_G$ be a shift of finite type with G strongly irreducible. Let $Y \subseteq X$ be a subsystem. If $Y \neq X$ then $h_{top}(Y) < h_{top}(X)$.*

Note that in general a subsystem can have the same entropy as the super-system. For example let X' be an identical copy of X, and $Z = X \cup X'$. Then $L_n(Z) = 2L_n(X)$, so $h_{top}(Z) = h_{top}(X)$, but $X \subseteq Z$ has the same entropy and is a proper subsystem.

Proof. Suppose that $Y \subseteq X$ is given. Then there is some cylinder in X that is disjoint from Y (since $X \setminus Y$ is a non-trivial open set). In other words, there is a word $a \in V^*$ that appears in X but not in Y.

We will show that $h_{top}(X) > h_{top}(Y) + \delta$, where $\delta > 0$ depends only on the length of a and the length $N = N_G$ in the definition of storng irreducibility.

Let $|a| = k$ and assume that $N > 10k$ (if not, increase N as necessary to achieve this).

Fix a large n and let $I = \{N, 2N, \ldots, (k-1)N\} \subseteq \{0, \ldots, n-1\}$, where $k = [n/N]$. For each $J = \{j_1 < j_2 < \ldots < j_m\} \subseteq I$, let $L(J) \subseteq L(X)$ denote the set of words w constructed as follows.

1. For each $j \in J$, set $w_j \ldots w_{j+k-1} = a$.

2. For each consecutive pair $i < j$ in J, choose $w_{i+k+N}w_{i+N+1}\dots w_{j-N} \in L_{j-i-(k+2N)}(Y)$. We do this also for the words $w_1 \dots w_{j_1-N}$ and $w_{j_m+k+N}\dots w_n$ at the beginning and end of w.

3. Now w consists of paths in G separated by "gaps" of length N before and after the copies of a defined in step 1. For each such "gap", complete the word to a path (we can do so by definition of N).

Note that for different $J \subseteq I$, the sets $L(J)$ are disjoint. This is because, given any $w \in L(J)$, we can identify J by the positions of copies of a in w. Indeed, in the part of w words defined in step (2), there are no a's, since $Y \cap [a] = \emptyset$ by choice of a. So an a can appear only at point $j \in J$, as defined in step (1), or possibly in positions which intersect the N locations to the right and left of the words defined in (1). But since $N > 10k$, these potential copies of a are distinct and each is within $3k$ of a unique $i \in I$. Thus, the locations of a in w determins J.

We estimate $|L(J)|$. Let $J = \{j_1 < j_2, \dots < j_m\}$. Then for each j_ℓ we make an independent choice in step (2) of a word of length $j_{\ell+1} - j_\ell - (k+2N)$ from Y. The number of words of length L in Y is $\geq 2^{h_{top}(Y)L}$ (because $h_{top}(Y) = \inf_L \frac{1}{n}\log|L_L(Y)|$); so the total number $\#(J)$ of patterns we have chosen so far is

$$\#(J) \geq \prod_{\ell=1}^{m-1} 2^{h(Y)(j_{\ell+1}-j_\ell-(k+2N)} = 2^{h(Y)\sum(j_{\ell+1}-j_\ell)-h(Y)m(k+2N)} = 2^{h(Y)n-C\cdot h(Y)\cdot m}$$

where $C = C(k, N, h(Y))$ depends only on $k, N, h(Y)$, hence on G and Y. Recall also that $m = |J|$. Thus we need to bound from below the sum

$$\sum_{J\subseteq I} \#(J) \geq \sum_{J \subseteq I} 2^{h(Y)n-C\cdot|J|}$$

(indeed, the sum bounds $|L_n(X)|$ from below). We decompose the sum on the right according to the density of J in I. In fact, it is enough to consider a single fixed density. Let $\varepsilon > 0$ be a parameter; by the previous lemma, there are $2^{h(\varepsilon)n+o(n)}$ such sets J. We use this to lower-bound the sum above:

$$\begin{aligned} \geq \sum_{J\sim\varepsilon I} 2^{h(Y)n-C\cdot\cdot|J|} &\geq 2^{h(\varepsilon)n+o(n)+h(Y)n-C\varepsilon n} \\ &\geq 2^{h(Y)n+(h(\varepsilon)-C\varepsilon)n+o(n)} \end{aligned}$$

It remains only to note that the function $h(t)$ has infinite right-derivative at zero, so for ε small enough relative to C, we have $\delta = h(\varepsilon) - C(\varepsilon) > 0$. Then for all large enough n, we have shown that $|L_n(X)| \geq 2^{(h(Y)+\delta/2)n}$, which proves the claim. □

Let $X = X_G$ be a shift of finite type and let a be a word in $L(X)$ (a path in G) . Let $X_G^{(a)} \subseteq X$ denote the set of paths in G which do not contain a as a sub-path. Then X_G is clealry shift-invariant, and also closed, since if $x_n \in X_G^{(a)}$ and $x_n \to x$ then $x \in X$ but every finite sub-sequence of x appears in x_n for all large n, hence cannot be equal to a.

Assuming that G is strongly irreducible, we know from the previous theorem that $h_{top}(X_G^{(a)}) < h_{top}(X_G)$. Certainly $X_G^{(a)}$ can also be empty, but the next theorem shows,that if a is long enough then not only is it not empty, but its entropy approaches that of X_G. First, a lemma. Observe that a word b belongs to $X_G^{(a)}$ if and only if for every M it can be extended forward and backwards by M symbols so that the extended word does not contain a copy of a.

Lemma 9.12. *Let X_G, a be as above. Then there exists an M with the following property: If b is any word that does not contain a copy of a, and if b can be extended M symbols forward and backward to a word b' also not containing a, then $b \in L(X_G^{(a)})$.*

Proof. Fix a. If M is large enough, then any word of length M contains at least two identical sub-words of length $|a|$. Now for this m suppose that we can extend b as stated. So we get a word of the form $c_1 b c_2$ where c_1, c_2 have length M. We can find in each a sub-word of length $|a|$ that repeats twice, so $c_1 = d_1 d d_2 d d_3$ and $c_2 = e_1 e e_2 e e_3$. But clearly $\ldots d d_2 d d_2 d d_2 \ldots$ and $\ldots e e_2 e e_2 e e_2 \ldots$ are infintie words not containing a, and so is $\ldots d d_2 d d_2 d d_3 b e_1 e e_2 e e_2 e \ldots$. The last sequence is therefore in $X_G^{(a)}$ and so $b \in L(X_G^{(a)})$. $\square$

Corollary 9.13. *The number of paths of length n in G which do not contain a fixed sub-path a is at most $C \cdot |L_n(X_G^{(a)})|$, where C depends only on G and a.*

Proof. Let M be associated to a as in the previous lemma. If $w = w_1 w_2 \ldots w_n$ is a word not containing a, then $b = w_{M+1} \ldots w_{n-M}$ can be extended by M letters in each direction so as not to contain a (just take w!) so by the lemma, $b \in L_{n-2M}(X_G^{(a)})$. The map $w \to b$ is at most $|V|^{2M}$-to-1, so the claim follows with $C = |V|^{2M}$. $\square$

Theorem 9.14. *Let $X = X_G$ be a shift of finite type with G strongly irreducible. Then there exists a sequence $\delta_k \to 0$ such that if $a \in V^\ell$ then $h_{top}(X_G^{(a)}) > h_{top}(X_G) - \delta_k$.*

Proof. The analysis is similar to that in the previous proof. Fix $a \in L_k(X_G)$ and set $Y = X_G^{(a)}$, so a doesn't appear in Y. Let n be large. We count the words $w \in L_n(G)$. For each such word w let $J \subseteq \{1, \ldots, n-k\}$ denote the set of indices j such that $w_j w_{j+1} \ldots w_{j+k-1} = a$.

Let $J_0 \subseteq J$ be a minimal subset with the property that $J \subseteq J_0 + [0, k-1]$ (such a set exists because $J \subseteq J + [0, k-1]$). We claim that $|J_0| < 2n/k$. Indeed, if $j \in J_0$, and if $j', j'' \in J_0 \cap [j, j+k-1]$, then we must have $j' = j''$, since otherwise we could delete the smaller of the two from J_0 and $J_0 + [0, k-1]$ would remain unchanged. It follows that for each $j \in J_0$ the interval $[j, j+k-1]$ contains at most one other elements of J_0 to in total, J_0 has density at most $2/k$ in $J_0 + [0, k-1]$, hence in $[1, n]$.

For each $J_0 \subseteq \{1, \ldots, n-k\}$ of size $\leq 2n/k$, let $L(J)$ denote the set of $w \in L(G)$ from which it could arise by the procedure above. To estimate $|L(J)|$, let $I_1, \ldots, I_m$ denote the maximal intervals in $[1, n] \setminus (J_0 + [0, k-1])$. For each w that gives rise to J_0, the words $w|_{I_i}$ do not contain a. By the previous lemma, the number of possible values for $w|_{I_i}$ is $\leq C \cdot L_{|I_i|}(X_G^{(a)})$, so the number of possibilities for $w|_{\bigcup I_i}$, and hence for $\#\{w$ which give rise to $J_0\}$, is

$$\leq \prod_{i=1}^{m} C L_{|I_i|}(X_G^{(a)}) \leq C^m L_{\sum I_i}(X_G^{(a)}) \leq C^{2n/k} L_n(X_G^{(a)})$$

Thus, using Lemma **??**,

$$\begin{aligned} |L_n(X_G)| &\leq \sum_{J_0 \subseteq \{1,\ldots,n-k\}, |J_0| \leq 2n/k} C^{2n/k} L_n(X_G^{(a)}) \\ &\leq C^{2n/k} \cdot 2^{h(2/k)n+o(n)} L_n(X_G^{(a)}) \\ &= 2^{h(X_G^{(a)})n+O(n/k)+h(2/k)n+o(n)} \end{aligned}$$

On the other hand,

$$|L_N(X_G)| = 2^{h(X_G)n+o(n)}$$

combining the last two inequalities and taking logs gives the statement. □

9.4 The Krieger embedding theorem

We say that a dynamical system (Y, T) can be embedded in a system (X, S) if there is a aubsystem $Y' \subseteq X$ isomorphic to Y.

Let $P_n(Y, T)$ denote the set of points of period n in Y, i.e.

$$P_n(Y, T) = \{y \in Y \, : \, T^n y = y\}$$

Theorem 9.15 (Krieger embedding theorem)**.** *Let G be a strongly irreducible graph and $X = X_G$ the associated shift of finite type. Let (Y, T) be a dynamical system satisfying*

1. *Y is zero-dimensional (i.e. there is a basis for the ctopology consisting of open and closed sets).*
2. *(Y, T) is expansive.*
3. *$|P_n(Y, S)| \leq |P_n(X, T)|$ for all n.*
4. *$h_{top}(Y) < h_{top}(X)$*

Then Y can be embedded in X.

The hypotheses of the theorem are almost necessesary. If Y embeds in X it is isomorphic to a subsystem $Y' \subseteq X$. The isomorphism is an particular a homeomorphism, and Y' is totally disconnected since X is, so (1) is necessary. Similarly, Y' is expansive since it is a symbolic systme, so Y is, showing that (2) is necessary. Clearly $P_n(Y') \subseteq P_n(X)|$ so $|P_n(Y)| = |P_n(Y')| \leq |P_n(X)|$, so (3) is necessary. Finally, $Y' \subseteq X$ implies $h_{top}(Y) \leq h_{top}(X)$, and since entropy is an isomorphism invariant, $h_{top}(Y) = h_{top}(Y')$, so (4) is nearly necesary. One cannot hope for the theorem to hold with a weak inequality in (4), because if $h_{top}(Y') = h_{top}(X)$ then, by Theorem **??**, $Y' = X$. On the other hand there are many sytems which satisfy the first three properties and have entropy $h_{top}(X)$, which are not isomorphic to X.

We will not prove the full theorem. Instead we prove a weaker version:

Theorem 9.16 (Krieger embedding theorem for minimal systems)**.** *Let G be a strongly irreducible graph and $X = X_G$ the associated shift of finite type. Let (Y, T) be a dynamical system satisfying*

1. *Y is zero-dimensional (i.e. there is a basis for the ctopology consisting of open and closed sets).*

2. *(Y,T) is expansive.*

3. *(Y,T) is an infinite minimal system.*

4. *$h_{top}(Y) < h_{top}(X)$*

Then Y can be embedded in X.

A minimal system containing a periodic point is equal to the orbit of that point and hence is finite. Thus the assumption (3) in the last theorem implies that $P_n(Y,T) = \emptyset$ for all n, and so is a special case of (3) in the previous theorem.

We begin the proof with a reduction to the case that Y is a symbolic system. Indeed let $\varepsilon > 0$ be a constant for the expansivity of Y, and let $\mathcal{U} = \{U_1, \ldots, U_k\}$ be a finite partition of Y into closed and open sets of diameter $< \varepsilon$., which exists since the topology is totaly disconnected and the space is compact. Let $\tau : Y \to \{1,\ldots,k\}^{\mathbb{Z}}$ be the itinerary map: $\tau(y)_i = j$ if and only if $T^i y \in U_j$. As we saw in **??**, the fact that $\mathcal{U}$ is a partition implies that τ is well defined, the fact that the U_i are open implies that it is continuous, and expansiveness implies that it is injective: indeed, if $x, y \in Y$ and $x \neq y$ then by expansiveness there is some i with $d(T^i x, T^i y) \geq \varepsilon$, hence $T^i x, T^i y$ cannot be in the same set U_j (since these sets have diameter $< \varepsilon$; so $\tau(x)_i \neq \tau(y)_i$, and hence $\tau(x) \neq \tau(y)$. Finally, τ also intertwines the action. So τ is an embedding of Y into $\{1,\ldots,k\}^{\mathbb{Z}}$, and so we can assume from the start that $Y \subseteq \{1,\ldots,k\}^{\mathbb{Z}}$ and T is the shift.

Thus, Y is an infinite minimal symbolic system. For a set U and $y \in Y$ consider the set $I(y) = \{i : T^i y \in U\}$. When U is open and closed, the map $y \to I(y)$ is continuous in the following sense: For every N, if $y_n \to y$ then $I(y) \cap [-N,N] = I(y_n) \cap [-N,N]$ for all large enough n. To see this, note that we must verify for each $i \in [-N,N]$ that for all large enough n, either y, y_n are both in U or neither are. But U is open, so if $y \in U$ then $y_n \to y$ implies $y_n \in U$ for all large n; and if $y \notin U$ then we have $y_n \notin U$ for large n by applying the same logic to the open set $Y \setminus U$.

Also note, that if U is open, then $I(y)$ as defined above is unbounded above and below. This is because of minimality; in fact, the gaps in $I(y)$ are bounded, uniformly in y.

A first approximation for the construction of τ is obtained as follows.

(A) Let ℓ_0 be large enough that $L_\ell(Y) < L_\ell(X)$ for all $\ell \geq \ell_0$; this exists since $L_\ell(Y) = 2^{h(Y)\ell + o(\ell)}$ and $L_\ell(X) = 2^{h(X)\ell + o(\ell)}$, and $h(Y) < h(X)$.

(B) For each $\ell \geq \ell_0$ choose an ijective function $\tau_\ell : L_\ell(Y) \to L_\ell(X)$, which exists by choice of ℓ_0.

(C) Choose a closed and open set $U \subseteq Y$ such that the gaps in $I(y)$ are at least ℓ_0 for all $y \in Y$. To see that this can be done, fix any $y \in Y$. Since y is not periodic, the points $y, Ty, \ldots, T^{\ell_0} y$ are all distinct. Thus by continuity of T, any small enough neighborhood U of y has the property that $U, Tu, \ldots, T^{\ell_0} U$ are pairwise disjoint. Since Y is totally disconnected we can choose U to be clopen. Now for any z, if $i \in I(z)$ then $T^i z \in U$. Then $i + k \in I(z)$ if and only if $T^{i+k} z \in U$, if and only if $T^k(T^i z) \in U$, so by choice of U if $k > 0$ then $k > \ell_0$.

We now define $\tau : Y \to X$ as follows. Given y, first compute $I(y) = \{\ldots i_{-1} < i_0 < i_1 < \ldots\}$. Now replace each word $y|_{[i_k, i_{k+1}-1]}$ with the word $\tau_{i_{k+!}-i_k}(y|_{[i_k,i_{k+1}-1]})$. The result of all these replacements is $\tau(y)$.

The map τ defined in this way is continuous. Indeed, for y and k, we will show that if y' is close enough to y then $\tau(y')_i = \tau(y)_i$ for all $-k \leq i \leq k$. Indeed fix y and let $i_- < -k$ and $i_+ > k$ be elements of $I(y)$. Then any y' close enough to y will have $I(y') \cap [i_-, i_+] = I(y) \cap [i_-, i_+]$ and $y'|_{[-i_-,i_+]} = y|_{[-i_-,i_+]}$. Thus, in the parsing of y and y' into blocks according to $I(y)$,$I(y')$ respectively, thes blocks in coordinates $[i_-, i_+]$ agree in both location and content. Thus the same substitution is performed on both of them, which implies directly that $\tau(y)|_{[i_-,i_+]} = \tau(y')|_{[i_-,i_+]}$.

Also, τ intertwines the action: Indeed, it is clear that $I(Ty) = I(y) - 1$. Thus the parsing of Ty into blocks is the shift of the parsing of y. Also the corresponding content of the blocks is the same because Ty is y shifted by one. So the substitution of Ty (i.e., $\tau(y)$) is the shift of the substitution of y.

What is not ensured in this construction, is that τ is injective. What is true is that, if we know $\tau(y)$ and also know $I(y)$, then we can recover y, because $I(y)$ tells us the parsing into blocks, and on eahc block we applied τ_ℓ for some ℓ, and these maps were chosen to be injective. However, it is not clear that $I(y)$ is in fact encoded in $\tau(y)$.

Also, the image of τ need not be contained in X. Each $\tau(y)$ is a concetenation of blocks from $L(X)$ but the concatenation need not itself form a legal path in G.

The last problem is easier to fix. First,

(D) Let $N = N_G$ be as in the definition of strong irreducibility,

Now in the construction above, choose ℓ_0 large enough that $L_\ell(Y) < L_{\ell-N}(X)$; this can be done for the same reson as before. Now when defining τ, each block $y|_{[i_k,i_{k+1}-1]}$ is replaced by is image under $\tau_{i_{k+1}-i_k}$, but the latter is shorter by N than the original block, so $\tau(y)$ now consists of blocks from X separated by undetermined gaps of length N. We fill in each gap to form a legal path in G; we do this deterministically, i.e. for each $u, v \in V$ we choose a single path $u \xrightarrow{N} v$, and use this to fill in each gap with the endpoints u, v. It is not hard to check that τ is still equivariant and continuous and its range is now in X.

To fix non-injectivity, we need to somehow "mark" $I(y)$ in the output sequence $\tau(y)$. To do this we introduce the following definition.

Definition 9.17. A Marker for X_G is a word $a \in L(X_G)$ such that if a appears in a word b at indices i, j, then $|i - j| \geq |a|$. That is, it cannot overlap itself non-trivially.

Proposition 9.18. *Let G be strongy irreducible. Then there exist arbitrarily long markers for X_G.*

Proof. Let $u_0 u_1 \ldots u_{k-1} u_0$ be a simple cycle in G, and of minimal length among all cycles in G. As in the proof of Lemma **??**, there must be some $0 \leq i < k$ such that u_i has two outgoing edges in G, ending at u_{i+1} and another vertex v. Withoug loss of generality we can assume $i = k - 1$ has this property (otherwise, permute the vertices cyclically). Also, we must have $v \neq u_j$ for all j, since otherwise we could form the shorter cycle $u_j u_{j+1} \ldots u_{k-1} u_j$, contradicting minimality.

Write $a = u_0 u_1 \ldots u_{k-1}$. For each n we claim that $a^n v$ is a marker. Indeed, suppose that $a^n v$ appears in b at indices $i < j$. If $j < nk + 1$, then the last symbol in $a^n v$ appears as part of the word $a^n v$ starting at j, but not the last symbol, and this is impossible, since v does not appear in a. Thus $a^n v$ is a marker and its length is $nk + 1$, which can be made arbitrarily large by choce of n. ∎

We now return to the proof of Krieger's theorem for minimal systems. We introduce the following parameter:

(E) Let $\delta = \frac{1}{2}(h_{top}(X_G) - h_{top}(Y))$.

(F) Let M be large enough that for every $m \geq M - 2N$ and $a \in L_m(X_G)$ we have $h_{top}(X_G^{(a)}) > h_{top}(Y) + \delta$. Such M exists by Theorem **??**.

(G) Choose a marker a for X_G of length $|a| \geq M$. Write $m = |a|$. Also, let $\widetilde{a}$ denote was remains of a after deleting the first N symbols and the last N symbols.

Now we modify step (A), replacing it with:

(A') Let ℓ_0 be large enough that $L_\ell(Y) < L_{\ell-m-2N}(X_G^{(\widetilde{a})})$ for all $\ell \geq \ell_0$; this exists since $L_\ell(Y) = 2^{h(Y)\ell + o(\ell)}$ and $L_\ell(X_G^{(\widetilde{a})}) = 2^{h(X_G^{(\widetilde{a})})\ell + o(\ell)}$, and $h(X_G^{(\widetilde{a})}) > H(Y) + \delta$.

We proceed now to choose τ_ℓ as in (B) and U as in (C), using the ℓ_0 defined in (A').

Now define $\tau : Y \to X_G$ as follows. Given $y \in Y$, first find $I(y) = \{\ldots < i_{-1} < i_0 < i_1 < i_2 < \ldots\}$, and in $\tau(y)$ put a copy of the marker a starting at every $i_k \in I(y)$. Then parse y into blocks $w_i = y|_{[i_k, i_{k+1}-1]}$, and put the block $\tau_{|w_i|}(w_i)$ at $i_k + m + N$. Note that by construction this block ends at index $i_{k+1} - N$.

The sequence $\tau(y)$ is now defined on the intervals $[i_k, i_k + m - 1]$ and $[i_k + m + N, i_{k+1} - N]$ and on these intervals we have valid paths in G. The final step is to fill inthe gaps, each of length $\geq N$, to form legitimate paths. We do this in a manner depending only on the vertices at the ends of the gaps.

As defined, it is clear that the image of τ is in X_G and that it is continuous and intertwines the action – the argument for these is the same as in the preliminary version of the construction.

It remains to show that τ is injective. For this we will show that $I(y)$ can be recovered from $\tau(y)$; once this is done, we can recover $\tau_{|w_i|}(w_i)$ and, since τ_ℓ are injective, recover w_i, which make up y.

We claim that we obtain $I(y)$ from $\tau(y)$ as the set of locations i at which a copy of a appears in $\tau(y)$. By construction, this set certainly contains $I(y)$, so we need to show that there are no other copies of a in $\tau(y)$ besides these. Indeed, such a copy cannot intersect the copies coming from $I(y)$, because a is a marker. Hence any other copy of a is contained in the gaps between the a's coming from $I(y)$. The sequence in such a gap consists of N symbols, then a word from $L(X_G^{(\widetilde{a})})$, and another N symbols. Thus if a occurs in a gap, then, except for at most N symbols at its start and end, it lies in a word from $X_G^{(\widetilde{a})}$, and, in particular, $\widetilde{a}$ (which is a without its first and last N symbols) lies entirely in word that comes from $L(X_G^{(\widetilde{a})})$; which is impossible.

10 Symbolic coding of toral automorphisms

10.1 Symbolic representation of dynamical systems

Let (X,T) be an invertible dynamical system and $\mathcal{A} = \{A_1,\ldots,A_r\}$ a cover of X. To avoid trivialities we assume that $A_i \neq \emptyset$.

A point $\omega \in \{1,\ldots,r\}^{\mathbb{Z}}$ is called the *itinerary* of a point $x \in X$ if

$$T^n x \in A_{\omega_n}$$

for all $n \in \mathbb{Z}$, that is, (ω_n) records where (in which set(s) A_i) $T^n x$ is for each n. Similarly one defines theforward itinerary $(\omega_n) \in \{1,\ldots,r\}^{\mathbb{N}}$ which is defined also when T is not invertible. We focus on the invertible case, the non-invertible one is similar.

The idea of taking itineraries is that we observe the orbit at “finite resolution”: at each time we take a “measurament” of the point T^n which gives us only a finite amount of information about it, namely, the identity of a set A_i to which it belongs. One can hope – and sometimes it is true – that this finite information taken along the entire history of the orbit may allow us to reconstruct the point x completely, or with high accuracy; and that the association $x \leftrightarrow \omega$ between points and itineraries will give us a “good” representation of the dynamics.

For there to be any hope of this we need (a) to better understand the association of itineraries to points, and (b) for this to be useful we need the set of itineraries to be simple enough to analyze (otherwise we have not gained any insight). We start with (a).

Remarks and basic properties

- Every $x \in X$ has some itinerary (because $T^n x$ belongs to some A_i, since $\mathcal{A}$ covers X).
- A sequence $\omega_{-N},\ldots,\omega_N \in \{1,\ldots,r\}^{2N+1}$ can be extended to an itinerary of x if and only if

$$x \in \bigcap_{n=-N}^{N} T^{-n} A_{\omega_n}$$

  (and similalry for forward itineraries taking the intersection from 0 to N).
- If ω is an itinerary of x then $S\omega$ (the shift of ω) is an itinerary of Tx.

Let

$$\Omega_{\mathcal{A}} = \{\text{all itineraries of points in } X\} \subseteq \{1,\ldots,r\}^{\mathbb{Z}}$$

More properties

- If $A_1,\ldots,A_r$ are closed then $\Omega_{\mathcal{A}}$ is closed. Indeed, let $\omega^1,\omega^2,\ldots$ be itineraries and $\omega^i \to \omega$. Let $x^1,x^2,\ldots \in X$ such that ω^i is an itinerary of ω^i and pass to a subsequence so $x^i \to x$. Then $T^n x^i \in A_{\omega^i_n}$ for all n,i; since $\omega^i_n = \omega_n$ for all large enough i, and A_i is closed, it follows that $T^n x \in A_{\omega_n}$, hence ω is an itinerary of x, and $\Omega_{\mathcal{A}}$ is closed to taking limits.
- Each $x \in X$ has a unique itinerary, if and only if $\mathcal{A}$ is a partition. This is clear, since if $x \in A_i$ and $x \in A_j$ for some i,j then x has itineraries starting with both i and j. Note that if $\mathcal{A}$ is a partition then each $A_i = X \setminus \bigcup_{j\neq i} A_j$ is open (since it is the complement of a finite union of closed sets). Thus X admits sets non-trivial sets which are both open and closed. In other words, it is disconnected. This is a non-trivial topological assumption (for example no such partition exists of a ball in a normed space, of tori, etc.).

- Suppose that $\mathcal{A}$ is a partition, so each x has a unique itinerary. Then we have a well defined function $i = i_{\mathcal{A}} : X \to \Omega_{\mathcal{A}}$ taking a point to its itinerary. Then this map is continuous. Indeed, since cylinder sets generate the topology in $\Omega_{\mathcal{A}}$, we need to show that for any $a_{-N}, \ldots, a_N \in \{1, \ldots, r\}^{2N+1}$, and writing $C = \{\omega : \omega_n = a_n \text{ for } |n| \leq N\}$, the set $i^{-1}(C)$ is open. But this set is just $\bigcap_{n=-N}^{N} T^{-n} A_{\omega_n}$, and since T is a continuous bijection and the A_i is open, this is a finite intersection of open sets, hence open, as desired.

  In conclusion, when $\mathcal{A}$ is a partition into closed (and open) sets, we obtain a factor map from X to $\Omega_{\mathcal{A}}$. The factor map will be an isomorphism when it is injective; a condition for this is given below.

The cover $\mathcal{A}$ is said to be *generating* if no ω is the itinerary of more than one point. A more useful condition is the following (in fact the two are equivalent, we leave the equivalence as an exercise):

$$\forall \varepsilon > 0 \quad \exists N \quad \forall \omega_{-N} \ldots \omega_N \in \{1, \ldots, r\}^{2N+1} \qquad \operatorname{diam} \bigcap_{n=-N}^{N} T^{-n} A_{\omega_n} < \varepsilon \tag{5}$$

- If (5) holds, then every $\omega \in \Omega_{\mathcal{A}}$ is the itinerary of a unique $x \in X$. This defines a map $\pi : \Omega_{\mathcal{A}} \to X$. The maps is, furthermore, continuous. Indeed let $\varepsilon > 0$. Let N correspond to ε in (5). Let $\delta = 2^{-N+1}$. Then $d(\omega, \omega') < \delta$ implies $\omega_{-N}, \ldots, \omega_N = \omega'_{-N}, \ldots, \omega'_N$. Since $\pi(\omega)$ and $\pi(\omega')$ are both in $\bigcap_{n=-N}^{N} T^{-n} A_{\omega_n}$, and since this is a set of diameter $< \varepsilon$ but our choice of N, we have $d(\pi(\omega), \pi(\omega')) < \varepsilon$, which gives continuity.

  In conclusion, assuming (5), we obtain a factor map $\pi : \Omega_{\mathcal{A}} \to X$. The map will generally not be an isomorphism – for that to happen we would need also that $\mathcal{A}$ is a partition into close and open sets, as above.

Example

Let $X = \mathbb{R}/\mathbb{Z}$ and $Tx = 2x \bmod 1$ (you can replace 2 by any larger integer).

Let A_i denote the cover of $\mathbb{R}/\mathbb{Z}$ by half-open intervals $A_0 = [0, 1/2]$ and $A_1 = [1/2, 1]$.

Since T is non-invertible, we will work with forward itineraries. Let $\Omega_{\mathcal{A}}$ denote the set of these.

Claim. For each $\omega_0 \ldots \omega_N \in \{0, 1\}^{\mathbb{N}}$ the set $\bigcap_{n=0}^{N} T^{-n} A_{\omega_n}$ is a closed interval of length $2^{-(N+1)}$.

Proof. By induction. It is true for $N = 0$. Assume it it true for $N - 1$. Then

$$\begin{aligned} \bigcap_{n=0}^{N} T^{-n} A_{\omega_n} &= A_{\omega_0} \cap \bigcap_{n=1}^{N} T^{-n} A_{\omega_n} \\ &= A_{\omega_0} \cap T^{-1} \left(\bigcap_{n=0}^{N-1} T^{-n} A_{\omega_{n+1}} \right) \end{aligned}$$

By the induction hypothesis, $\bigcap_{n=0}^{N-1} T^{-n} A_{\omega_{n+1}}$ is an interval of length 2^{-N}, and its inverse image under T is the union of two intervals of half that length, one contained in A_0 and one in A_1. Thus, the intersection of the preimage with A_{ω_0} is a single interval of length $2^{-(N+1)}$, as claimed. □

This proves (5). It also shows that $\Omega_{\mathcal{A}} = \{0,1\}^{\mathbb{N}}$, since in particular for every $\omega_0 \dots \omega_N$ there exist points with itineraries starting with the given sequence. We obtain a factor map $\pi : \{0,1\}^{\mathbb{N}} \to \mathbb{R}/\mathbb{Z}$. It is easy to see that this is just the binary coding map!

10.2 Hyperbolic toral automorphisms

We now describe (by example) a class of systems which allow a concrete representation via the factor map $\pi : \Omega_{\mathcal{A}} \to X$ described above. In this class, the symbolic system $\Omega_{\mathcal{A}}$ (or, rather a certain subset of it) is a shift of finite type, i.e. the set of paths through a directed graph; and thus easy to analyze; and π is injective on a large part of the space. The difficulty is in constructing the cover $\mathcal{A}$ which gives this representation. The example we give is the simplest possible beyond the example at the end of the previous section.

Recall that if A is an integer 2×2 matrix then it induces a map of $\mathbb{T}^2 = (\mathbb{R}/\mathbb{Z})^2$ which intertwines the projection $x \mapsto x \bmod 1$. If $\det A = \pm 1$ then by Cramérs rule, A^{-1} is also an integer matrix. Then A^{-1} induces a map on $\mathbb{T}^2$ as well and since $AA^{-1} = A^{-1}A = \mathrm{id}$ in $\mathbb{R}^2$, the same relation holds for the induced maps, i.e. they are invertible.

A 2×2 matrix is said to be *hyperbolic* if none of its eigenvalues has absolute value 1. In the 2×2 case we are considering this means that it has two distinct real eigenvalues λ_+, λ_-, and since their product is $\pm 1 = \det A$, and since the are not of modulus one, we must havge $\lambda_+ > 1 > \lambda_- > 0$ and $|\lambda_-| = 1/|\lambda_+|$.

We consider for simplicity the example

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

whose eigenvalues are $\lambda_\pm = \frac{1}{2}(3 \pm \sqrt{5})$ with corresponding eigenvectors $v_\pm = (\frac{1}{2}(1 \pm \sqrt{5}), 1)$. Since A is symmetric these are orthogonal.

We say that a set $R \subseteq \mathbb{R}^2$ is an A-rectangle if in the basis v_+, v_- it is a product of closed intervals, i.e. has the form $\{sv_+ + tv_- : (s,t) \in [a,b] \times [c,d]\}$. We shall speak of the v_- and v_+ sides of R with the obviousmeaning. Note that such a set is indeed a rectangle because $v_\pm$ are orthogonal; for general hyperbolic matrices the eigenvectors will not be orthogonal and A-rectangles we would get parallelograms instead.

We say that a subset $R \subseteq \mathbb{T}^2$ is an A-rectangle if it is the image of a rectangle in $\mathbb{R}^2$ under the mod-1 projection.

Properties of rectangles. If R is an A-rectangle with sides of length $a_\pm$ in directions $v_\pm$ respectively, then TR is an A-rectangle with sides of length $\lambda_+ a_+$ and $\lambda_- a_-$, respectively; and $T^{-1}R$ is an A-rectangle with sides of length $\lambda_+^{-1} a_+ = \lambda_- a_+$ and $\lambda_-^{-1} a_- = \lambda_+ a_-$. the proof is from the definition of the eigendirections and eigenvalues.

Also, a finite intersections of A-rectangles is an A-rectangle.

Fact. There exists a cover $\mathcal{B}' = \{B'_0, B'_1\}$ of $\mathbb{T}^2$ into A-rectangles which intersect only at their boundaries, and such that TB'_i is the union of finitely many A-rectangles, each contained in, and sharing the v_--sides of, one of the B'_i. The construction of $\mathcal{B}'$ can be found in chapter 5.12 of Brin and Stuck.

Now, one can verify that $TB'_1 \cap B'_1$ consists of two such rectangles B_1, B_3, and $TB'_1 \cap B'_2$ consists of one rectangle B_5; and $TB'_2 \cap B'_1$ consists of one rectangle B_2 and $TB'_2 \cap B'_1$ consists of one rectangle, B_4. (the indexing is to be consistent with the figure in Brin and Stuck).

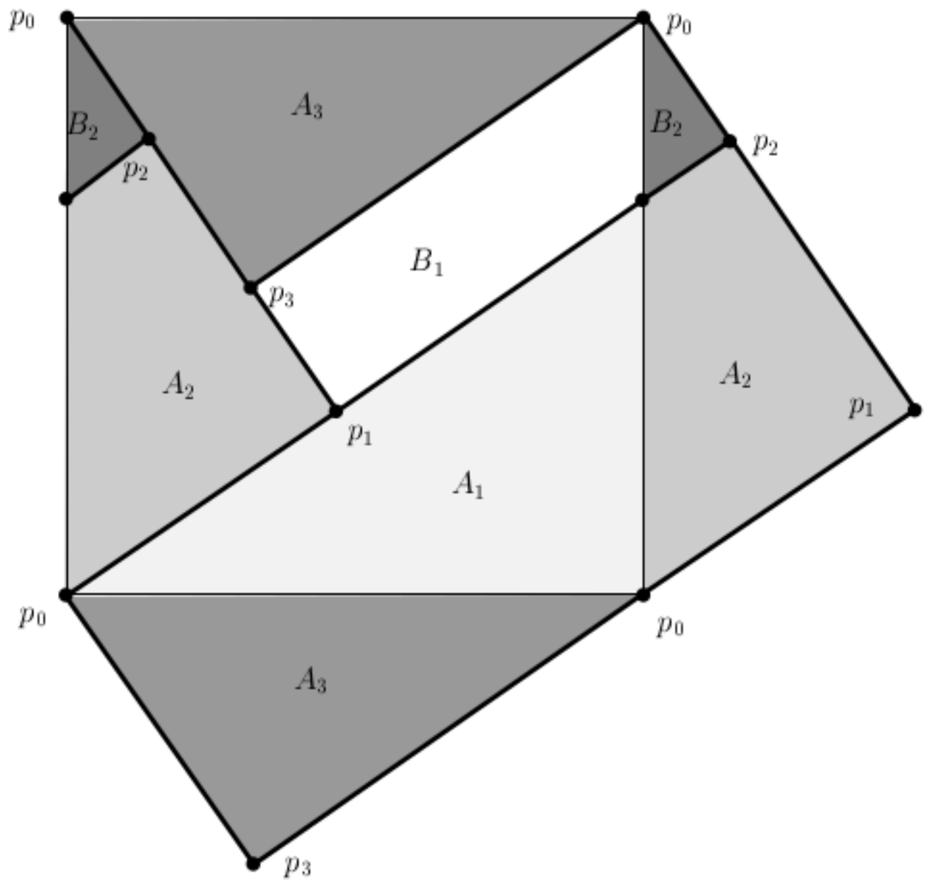

Figure 1: The cover $\{B_1', B_2'\}$ (notation in the figure is different ours). Taken from Brin & Stuck.

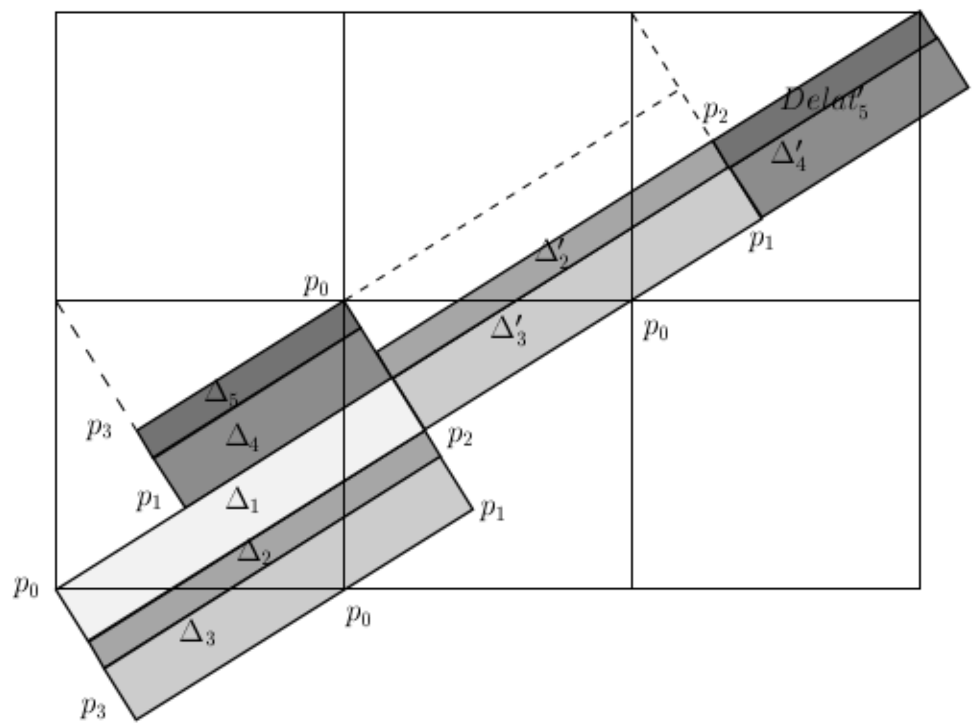

Figure 2: The cover $\{B_1, \ldots, B_5\}$ (notation in the figure is different from ours). Taken from Brin & Stuck.

Let $\mathcal{B} = \{B_1, \ldots, B_5\}$ denote the cover thus obtained. Further, let $G = (V, E)$ denote the graph $V = \{1, 2, 3, 4, 5\}$ (the index set of the B_i), and $(i, j) \in E$ if $TB_i \cap B_j$ has non-empty interior. That is, intersect not only at the boundaries.

The proof of the following should be compared with the example in the previous section.

Claim. For any $\omega_0 \ldots \omega_N \in \{1, \ldots, 5\}^{\mathbb{N}}$, $\bigcap_{n=0}^{N} T^n B_{\omega_n}$ is an A-rectangle whose side length in the V_+ direction is $\leq C$ and in the v_- direction is $\leq C\lambda_-^{N+1}$ for some constant $C > 0$. Furthermore, if (ω_n) form a path in G then the intersection is not empty.

Proof. By induction. For $N = 0$ it is trivial. Assume it for $N - 1$ and let $\omega_0, \ldots, \omega_N$ be given. Then

$$\begin{aligned}\bigcap_{n=0}^{N} T^n B_{\omega_n} &= B_{\omega_0} \cap \bigcap_{n=1}^{N} T^n B_{\omega_n} \\ &= B_{\omega_0} \cap T^1 \left(\bigcap_{n=0}^{N-1} T^n B_{\omega_{n+1}} \right)\end{aligned}$$

The set B in parentheses is, by the induction hypothesis, an A-rectangle of proportions $a \times b$ with $a \leq C$ and $b \leq C \cdot \lambda_-^N$ and is contained in B_j for some j. So $T^1 B$ has proportions $\lambda_- a \times \lambda_+ b$ and is contained in TB_j. The intersection of TB_j with B_{ω_0} is contained in the intersection of B'_u and TB'_v where B'_u, B'_v are the rectangles containing B_{ω_0} and B_j, and this is one of the B_w's; thus $\bigcap_{n=0}^{N} T^n B_{\omega_n} = B_{\omega_0} \cap TB$ has the proportioned claimed.

For the last claim we do the same induction, noting that since B is a rectangle running the entire v_+-side of B_{ω_1}, the interior of TB runs the entire length of the v_+-direction of the interiors of its intersection with B_i for all i such that $i\omega_1 \in E$, and in particular intersects each of them non-trivially. This proves the claim. ☐

Corollary 10.1. *The cover $\mathcal{B}$ satisfies (5).*

Proof. Use $\bigcap_{n=-N}^{N} T^{-n} B_{\omega_n} = T^{-N}(\bigcap_{n=0}^{2N+1} T^n B_{\omega_n})$. By the previous claim, the set in parentheses is an A-rectangle of dimension $O(1) \times O(\lambda_-^{2N+1})$, so T^{-N} of it has dimensions $O(\lambda_-^N) \times O(\lambda_-^N)$. Since $\lambda_- < 1$ this means that the diameter of $\bigcap_{n=-N}^{N} T^{-n} B_{\omega_n}$ decays exponentially independently of ω, proving the claim. ☐

It follows that we have a factor map $\pi : (\Omega_{\mathcal{B}}, S) \to (\mathbb{T}^2, T)$. Let Y be the space of paths through G. The claim above shows that $Y \subseteq \Omega_{\mathcal{B}}$.

Claim 10.2. $\pi(Y) = \mathbb{T}^2$, so π is a factor map $Y \to \mathbb{T}^2$ and $|\pi^{-1}(x)| = 1$ for all $x \in \mathbb{T}^2$ outside of a countable union of line segments (in particular for a.e. x and for x in a dense G_δ set).

Proof. The fact that π is a factor map follows from the previous discussion.

For the last statement note that x belongs to two of the rectangles $B_1, \ldots, B_5$ only when it belongs to their boundaries. Therefore π is non-injective only on the union of the A-images of these line segments, which form a zero measure et and a set of first category. ☐

We can be a little more precise about when π is non-injective. If $\pi^{-1}(x)$ contains more than one point then $T^n x \in \partial B_i$ for some $n \in \mathbb{Z}$ and $i \in \{1, \ldots, 5\}$. If the boundary component is in the v_+ direction then, since the boundaries in this direction are mapped by T into themselves, we see that for all future times, $T^{n+k}x$ is contained in such a boundary component also; so every itinerary of x is, fom some point on, not euqal to 2. In the same way if the boundary component is in the v_- direction then at all previous times he same holds, so any itinerary omits the symbols $2, 3$ from some point on. Thus, for example, any point that is forward and backward transitive in $\mathbb{T}^2$ has a unique pre-image.

10.3 Remarks

The existence of a shift of finite type symbolic extension for hyperbolic toral automorphisms extends to higher dimensions. There, a matrix is hyperbolic if all its eigenvalues have modulus $\neq 1$. However, the cover obtained is not as nice as the one we described. In dimension ≥ 3, the sets in the cover are not rectangles or even parallelograms. Rather, they have a "fractal" boundary, and in particular their boundaries are not piecewise smooth.

More generally, there is a broad class of dynamical systems, the so-called hyperbolic systems, which admit similar covers. These are, essentially, systems on manifolds such that the differential at typical points consist of hyperbolic matrices.

There are still quesitons in this area. It is not clear for non-hyperbolic system when such symbolic covers exist, and this is still an active area of research.

11 Multiple recurrence and Van der Warden's theorem

Van der Warden's theorem is the following combinatorial statement of Ramsey type:

Theorem 11.1 (Van der Warden)**.** *Let $\mathbb{N} = A_1 \cup \ldots \cup A_r$ be a finite partition of the natural numbers. Then one of the sets A_i contains arbitrarily long arithmetic progressions, i.e., for every d there are a, b such that $a, a+b, a+2b, \ldots, a+db \in A_i$.*

There is a dynamical proof of this theorem, due to Furstenberg and Weiss, which uses the following generalization of the recurrence phenomenon for single maps.

Theorem 11.2. *Let X be a compact metric space and $T_1, \ldots, T_d : X \to X$ continuous commuting maps, i.e. $T_iT_j = T_jT_i$. Then there exists an $x \in X$ which is jointly recurrent under the maps, i.e. not only is it recurrent for all of them, but it is recurrent along a common sequence of times: there exists a sequence $n_k \to \infty$ such that $T_1^{n_k}x \to x$, $T_2^{n_k} \to x$, $\ldots$, $T_d^{n_k}x \to x$ as $k \to \infty$.*

Proof that the multiple recurrence theorem implies Vann der Waerden. Let $\mathbb{N} = A_1 \cup \ldots \cup A_r$ be a partition of the integers.

Let $x \in \{1, \ldots, r\}^{\mathbb{N}}$ denote the point

$$x_n = i \qquad \text{if } n \in A_i$$

Let $X \subseteq \{1, \ldots, r\}^{\mathbb{N}}$ denote the orbit closure of x.

Fix $d \geq 1$ and set $T_i = S^i$ (where S is the shift map). Then $T_1, \ldots, T_d$ are commuting, continuous maps of X.

By the multiple recurrence theorem there exists $y \in X$ and $n \in \mathbb{N}$ such that $d(T_i^n y, y) < 1/2$ for $i = 1, \ldots, d$. Here $d(\cdot, \cdot)$ denotes the metric on the shift space and it is defined in such a way that when the distance between two points is less than $1/2$, they must agree in their first cordinate.

Since $T_i^n = S^{in}$, the condition $d(T_i^n y, y) < 1/2$ implies that $y_1 = y_{in}$. Thus all the symbols $y_n, y_{2n}, \ldots, y_{dn}$ agree and are equal to y_1.

Since $y \in \overline{O_S(x)}$ there is a sequence (n_k) such that $S^{n_k}x \to y$. In particular for large enough k, the points $S^{n_k}x$ and y agree on their first dn symbols. It follows, that $x_{n_k+n} = y_n$, $x_{n_k+2n} = y_{2n}$, ... $x_{n_k+dn} = y_{dn}$. In particular,

$$x_{n_k+n} = \ldots = x_{n_k+dn} = y_1$$

which means that the set A_{y_1} contains the arithmetic progression $n_k + n, n_k + 2n, \ldots, n_k + dn$.

We have thus shown that for every d, one of the sets A_i contains an arithmetic progression of length d. Since there are finitely many sets A_i, one of them must contain arithmetic progressions of arbitrarily large length, and thus of all lengths. □

The proof of the multiple recurrence theorem presented below is taken from Chapter 3 (and end of Chapter 2) of Furstenberg's book "Recurrence in ergodic theory and combinatorial number theory".

11.1 Commuting maps : a quick introduction

Commuting continuous onto maps $T_1, \ldots, T_d$ of X can be viewd as the action of an abelian group or semigroup G on X, and all the notions we have developed apply. In particular the orbit of a point is the set of images under all elements of the semigroup, the action is minimal if there are no closed non-empty proper subsets which are invariant under the action (equivalently under all the maps $T_1, \ldots, T_d$), and minimality is the same as every orbit being dense. n a minimal action, if $U \neq \emptyset$ is an open set then $\bigcup g^{-1}U$ covers X (and so there is a finite sub-cover).

11.2 Recurrent sets and homogeneous sets

Definition 11.3. Let (X, T) be a dynamical system. A subset $A \subseteq X$ is called recurrent for T, if for every $x \in A$ and every $\varepsilon > 0$ there exists $y \in A$ and $n \in \mathbb{N}$ such that $d(T^n y, x) < \varepsilon$.

Note that we did not require A to be T-invariant!
For example, when T is onto, X is recurrent.

Lemma 11.4. *If $A \subseteq X$ is recurrent then for every $\varepsilon > 0$ there exists $z \in A$ and $n \in \mathbb{N}$ with $d(T^n z, z) < \varepsilon$.*

Proof. Fix $\varepsilon > 0$. Choose any $z_0 \in A$ and $\varepsilon_0 = \varepsilon/2$ and define a sequence recursively as follows:

Given x_n, ε_n, we use the fact that A is recurrent to find $x_{n+1} \in A$ and $k_{n+1} \in \mathbb{N}$ such that $T^{k_{n+1}} x_{n+1} \in B_{\varepsilon_n}(X_n)$, and use continuity of T to find $\varepsilon_{n+1} < \varepsilon/2$ so that $T^{k_{n+1}}(B_{\varepsilon_{n+1}}(x_{n+1})) \subseteq B_{\varepsilon_n}(x_n)$.

By compactness, there will be two points in the sequence, say x_i and x_j with $< j$, such that $d(x_i, x_j) < \varepsilon/2$. Now one has $T^{k_j} x_j \in B_{\varepsilon_{j-1}}(x_{j-1})$, hence $T^{k_{j-1}}(T^{k_j} x_j) \in B_{\varepsilon_{j-2}}(x_{j-2})$, and so on, until we find that, writing $k = k_j + k_{j-1} + \ldots + k_{i+1}$,

$$T^k x_j = T^{k_{i+1}} T^{k_{i+1}} \ldots T^{k_j} x_j \in B_{\varepsilon_i}(x_i)$$

Therefore,

$$d(T^k x_j, x_j) \leq d(T^k x_j, x_i) + d(x_i, x_j) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

and x_j is the desired point. □

Definition 11.5. Let (X, T) be a dynamical system. A subset $A \subseteq X$ is called **homogeneous** if there exists a group G of homeomorphism of X commuting with T (i.e. $gT = Tg$ for $g \in G$) and such that A is a minimal subset for G (i.e. A is closed and G-invariant and contains no smaller G-subsystem).

Note again that A was not required to be T-invariant!

For example, if G is a compact group and $\varphi : X \to G$ is a continuous function, we can form the skew product $X \times G$ with the map $\widetilde{T}(x, g) = (Tx, \varphi(g)x)$. The group G acts on $X \times G$ via the maps $R_g(x, h) = (x, hg^{-1})$, and these maps commute with T and act minimally on each set $\{x\} \times G$. So these sets are homogeneous in the sense above.

Lemma 11.6. *If $A \subseteq X$ is recurrent and homogeneous then it contains a recurrent point for T.*

Proof. Define $F : A \to [0, \infty)$ by

$$F(x) = \inf\{d(T^n x, x\} \ : \ n \in \mathbb{N}\}$$

Then F is the infimum of continuous functions, so it is upper semi-continuous (if $y_n \to y$ then $\limsup F(y_n) \le F(y)$).

Let $x_0 \in A$ be a point of continuity of F (every upper semi-continuous function is continuous on a dense G_δ set of points).

We claim that $F(x_0) = 0$, which implies that x_0 is recurrent.

For suppose not. Let $\varepsilon > 0$ and let V be an open neighborhood of x_0 such that $f|_V > \varepsilon$.

Let G be the group from the definition of homogeneity and choose a finite set $\Gamma \subseteq G$ such that $\bigcup_{g\in\Gamma} gV \supseteq A$ (we can do this because G acts minimally on A).

Let $\delta > 0$ be such that if $d(x, y) < \delta$ then $d(gx, gy) < \varepsilon$ for all $g \in \Gamma$.

By the previous lemma we can find $z \in A$ and $n \in \mathbb{N}$ with $d(T^n z, z) < \delta$.

Also there exists $g \in \Gamma$ with $gz \in V$.

Therefore,

$$\begin{aligned} d(T^n gz, gz) &= d(g(T^n z), gz) \qquad \text{because } g, T \text{ commute} \\ &< \varepsilon \qquad \text{because } d(T^n z, z) < \delta \text{ and by choice of } \delta \end{aligned}$$

Hence $F(gz) < \varepsilon$ contradicting $gz \in V$ and the definition of ε, V. □

We next show that we require les than recurrene of A to draw the same conclusion.

Proposition 11.7. *Let $A \subseteq X$ be homogeneous and suppose that*

For every $\varepsilon > 0$ there exist $x, y \in A$ and $n \in \mathbb{N}$ with $d(T^n x, y) < \varepsilon$

Then A contains a recurrent point for T.

Proof. We want to show that the stated condition implies that A is recurrent; then the conclusion follows from the previous lemma.

Fix any $z \in A$ and $\varepsilon > 0$, and let G be as in the definition of homogeneity.

We can choose a finite set $\Gamma \subseteq G$ such that for any $x, y \in A$ there exists $g \in \Gamma$ with $d(gx, y) < \varepsilon/2$. Indeed, let $B_1, \ldots, B_N$ be a cover of A by balls of diameter $< \varepsilon/2$. By minimality, for each i there are elements $g_{i,1}, \ldots, g_{i,n(i)}$ such that $\bigcup_{j=1}^{n(i)} g_{i,j}^{-1} B_i \supseteq A$. Now given $x, y \in A$, we have $y \in B_i$ for some i, and then $x \in g_{i,j}^{-1} B_i$, so $g_{i,j} x \in B_i$, and hence $d(g_{i,j} x, y) < \operatorname{diam} B_i < \varepsilon/2$. So $\Gamma = \{g_{i,j}\}$ satisfies our requirement.

Let $\delta > 0$ be such that if $d(x, y) < \delta$ then $d(gx, gy) < \varepsilon/2$ for all $g \in \Gamma$.

Now use the hypothesis of the proposition, to find $x, y \in A$ and $n \in \mathbb{N}$ with $d(T^n y, x) < \delta$.

Let $g \in \Gamma$ be such that $d(gx, z) < \varepsilon/2$. Then

$$\begin{aligned} d(T^n gy, gx) &= d(gT^n y, gx) \qquad \text{because } gT = Tg \\ &< \varepsilon/2 \qquad \text{because } d(T^n y, x) < \delta \text{ and definition of } \delta \end{aligned}$$

Therefore

$$d(T^n gy, z) \le d(T^n gy, gx) + d(gx, z) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

so gy is the desired point verifyling recurrence of A. □

11.3 Proof of the multiple recurrence theorem

We want to show that if$T_1, \ldots, T_N$ are commuting continuous maps of X then there is a jointly recurent point, that is, a point $x \in X$ and a seqeuence $n_k \to \infty$ such that $T_i^{n_k} x \to x$ for $i = 1, \ldots, N$.

By passing to a subsystem we may assume that the action of $T_1, \ldots, T_N$ is minimal.

We also may assume they are invertible, for if not, one can pass to an extension in which the maps are invertible (as in the natural extension construction for a single map).

We prove the claim by induction on N.

For $N = 1$, it is just the Birkhoff recurrence theorem.

Suppose we have proved it for $N - 1$. Let

$$\Delta = \{(x, \ldots, x) \in X^N \ : \ x \in X\}$$

denote the diagonal and

$$\widetilde{T} = T_1 \times T_2 \times \ldots \times T_N$$

This is a map of X^N. We wish to find a point $(x, \ldots, x) \in \Delta$ that is recurrent for $\widetilde{T}$, for then x is jointly recurrent for the T_i.

Let

$$\widetilde{T}_i = T_i \times \ldots \times T_i$$

Then $\widetilde{T}_1, \ldots, \widetilde{T}_N$ comute with $\widetilde{T}$ and act iminimally on Δ. This shows that Δ is homogeneous.

Thus, by the proposition from the previous section, it is enough if we show that Δ satisfies:

$$\forall \varepsilon > 0 \ \exists \widetilde{x}, \widetilde{y} \in \Delta \ \exists n \in \mathbb{N} \ d(\widetilde{T}^n y, x) < \varepsilon$$

Define the maps

$$R_i = T_i T_N^{-1}$$

These are continuous commuting maps of X, so by the induction hypothesis, there is a point $x \in X$ and $n_k \to \infty$ such that $R_i^{n_k} x \to x$.

Define

$$\widetilde{x} = (x, x, \ldots, x) \in \Delta$$

Also, let

$$\widetilde{y}_k = (T_N^{-n_k} x, \ldots, T_N^{-n_k} x) \in \Delta$$

and observe that

$$\begin{aligned} \widetilde{T}^{n_k} \widetilde{y}_k &= (T_1^{n_k} T_N^{-n_k} x, \ldots, T_{N-1}^{n_k} T_N^{-n_k} x, x) \\ &= (R_1^{n_k} x, \ldots, R_{N-1}^{n_k} x, x) \\ &\to \widetilde{x} \end{aligned}$$

so for large enough k the choice $\widetilde{y} = \widetilde{y}_k$ and $n = n_k$ satisfy $d(\widetilde{T}^n \widetilde{y}, \widetilde{x}) < \varepsilon$, which is what we wanted.

12 Furstenberg's $\times 2, \times 3$ theorem

Furstenberg's $\times 2, \times 3$ theorem is a result on the "rigidity" of certain commuting transformations, and although the proof we give is simple it is an extremely important result which has lead to many generalizations, questions and applications.

For an integer $a \geq 2$ let $T_a : \mathbb{T} \to \mathbb{T}$ denote the times-a map $x \mapsto ax \bmod 1$. As we know, the system $(\mathbb{T}, T_a)$ s a factor of $\{0, 1, \ldots, a-1\}^{\mathbb{N}}$ by a map that is 1-1 outside a countable set of rational points. By the Krieger embedding theorem, this system is universal for expansive zero-dimensional subshifts which obey mild conditions on their entropy and periodic points.

The maps T_a, T_b commute, and in fact, $T_a T_b = T_{ab}$. However not every pair a, b gives a genuinely 2-dimensional action. For example the maps T_2, T_4 together just give an action of $\mathbb{N}$ (the powers of T_2), and similarly T_4, T_8 generate a semigroup which is isomorphic to $\mathbb{N} \times \mathbb{Z}/2\mathbb{Z}$, which is not very different from $\mathbb{Z}$. In particular any T_2-invariant set is invariant under T_4, T_8, hence we again find that there are many subsystems of the joint T_4, T_8 action.

Definition 12.1. Let $a, b \geq 2$ be integers. We write $a \sim b$ if a, b are powers of some integer, or equivalently if $\log a / \log b \in \mathbb{Q}$. In this case we say a, b are multiplicatively dependent. When they are not we call them **multiplicatively independent** and write $a \not\sim b$.

By our discussion, when $a \sim b$ the joint action of T_a, T_b admits many subsystems. Surprisingly when $a \not\sim b$ they do not!

Theorem 12.2 (Furstenberg's $\times - 2, \times - 3$ Theorem)**.** *Let $a \not\sim b$ be integers greater than one. Then there is no infinite, closed, T_a, T_b-invariant proper subset of $\mathbb{T}$.*

The name "$\times 2, \times 3$ theorem" derives from the fact that $2 \not\sim 3$ are prototypical examples.

The theroem classifies the infinite invariant subsystems, but in fact, it is also easy to classify the finite ones. Any irrational point clearly has an infinite orbit under T_a (since if $a^n x = x \bmod 1$ $a^n x = x + k$ for some integer k and then x is rational). So the only finite invariant sets must consist of rationals, and in fact, the orbit of every rational point is finite under the joint action of T_a, T_b (since both maps preserve the denominator).

Thus, we can reformulate the theorem as follows:

> The only subsystems of the joint T_a,T_b action are $\mathbb{T}$ itself, and certain finite sets of rational numbers.
>
> Equivalently: if $x \in [0, 1]$ is irrational, then $\{a^n b^k x \bmod 1\}_{k,n \in \mathbb{N}}$ is dense in $[0, 1]$.

We turn to the proof, which will require several steps.

Fix $a \not\sim b$ as in the theorem, set

$$S = \{a^k b^n\}_{n \in \mathbb{N}}$$

and order S as

$$S = \{s_1 < s_2 < s_3 < \ldots\}$$

The main analytic input for the proof is provided by the next lemma, which is the only place that the assumption $a \not\sim b$ will be used.

Lemma 12.3. $s_{n+1}/s_n \to 1$.

Proof. Let $\varepsilon > 0$. Observe that $\alpha = \log 2/\log 3$ is irrational, so there exist $p, q \in \mathbb{N}$ such that $p\log 2 - q\log 3 \in (0, \varepsilon)$; indeed, this is equivalent to $p\alpha \in (q, q+\varepsilon)$ and we have seen that $(p\alpha)_{p=1}^{\infty}$ is dense modulo 1.

Similarly, we can find p', q' such that $-p'\log 2 + q'\log 3 \in (0, \varepsilon)$.

Write $n_t = 2^{k(t)}3^{\ell(t)}$, and note that $k(t) + \ell(t) \to \infty$ as $t \to \infty$. Consider any t large enough that $k(t) + \ell(t) > p' + q$. Then either $k(t) > p'$ or $\ell(t) > q$. Suppose for instance the first of these holds. Let $k = k(t) - p'$ and $\ell = \ell(t) + q'$. We then have

$$\begin{aligned} 2^k 3^\ell &= 2^{k(t)}3^{\ell(t)}2^{-p'}3^{q'} \\ &= 2^{k(t)}3^{\ell(t)}\exp(-p'\log 2 + q\log 3) \end{aligned}$$

By choice of $p'q'$ this is greater than 2^k3^ℓ but less than $2^{k(t)}3^{\ell(t)} \cdot e^{\varepsilon}$. Thus $2^k3^\ell \geq n_{k+1}$ and we conclude that

$$n_t \leq n_{t+1} \leq e^{\varepsilon} \cdot n_t$$

If we happened to be in the case $\ell(t) > q$, we would define $k = k + p$ and $\ell = \ell - q$ and proceed in the same way.

In both cases, we found that $1 \leq n_{t+1}/n_t \leq e^{\varepsilon}$. Since ε was arbitrary, this proces the claim. □

Lemma 12.4. *Suppose that $X \subseteq \mathbb{T}$ is closed and T_a, T_b invariant. If X contains a non-isolated rational point, then $X = \mathbb{T}$.*

Proof. First suppose that $0 \in X$ is not isolated. Then we can find a monotone sequence in X converging to 0. We may assume that the sequence converges from the right; if it converges from the left, replace X by $-X$, noting that it is still T_a, T_b invariant, and now 0 can be approximated from the right.

Thus let $x_n \in [0, 1] \cong \mathbb{T}$ with $x_n \searrow 0$.

Fix $y \in \mathbb{T}$.

For each n, let $k(n)$ be the largest integer with $s_{k(n)}x_n < y$, so

$$s_{k(n)}x_n < y < s_{k(n)+1}x_n$$

Then

$$\begin{aligned} |y - s_{k(n)}x_n| &< s_{k(n)+1}x_n - s_{k(n)}x_n \\ &= \left(\frac{s_{k(n)+1}}{s_{k(n)}} - 1\right)s_{k(n)}x_n \\ &\leq \frac{s_{k(n)+1}}{s_{k(n)}} - 1 \\ &\to 0 \qquad \text{as } n \to \infty \end{aligned}$$

where we used $s_{k(n)}x_n < y \leq 1$.

We conclude that $y \in \overline{\{sx_n \,:\, s \in S \,,\, n \in \mathbb{N}\}} \subseteq X$, and since y was arbitrary, $X = \mathbb{T}$.

Now suppse $r \in X$ is a rational accumulation point. Then so is sr for all $s \in S$. since the orbit of r under T_a, T_b is finite, we can find $s \in S$ and n_0, k_0 such that $r' = sr$ is fixed by multiplication by $a' = a^{n_0}$ and $b' = b^{k_0}$. Note that $r' \in sX = X$.

Now let $X' = X - r'$. Since r' is fixed by a', b' the set X' is $T_{a'}, T_{b'}$ invariant. Also, $a' \not\sim b'$ since $\log a'/\log b' = (n_0/k_0)\log a/\log b$. Finally, 0 is an accumulation point in X'. So by the first part of the proof, $X' = \mathbb{T}$ and it follows that also $X = \mathbb{T}$. $\square$

Observe that if $X \subseteq \mathbb{T}$ is T_a, T_b-invariant and infinite it contains a not necessarily rational accumulatoin point x_0. Furthermore,

$$X - X = \{x_1 - x_2 \ : \ x_1, x_2 \in X\}$$

is also an infinite closed T_a, T_b-invariant that contains $0 = x_0 - x_0$ and 0 is an accumulation point, since if $x_n \to x_0$ monotonically in X then $0 \neq x_n - x_{n-1} \to 0$ in $X - X$. We then conclude that $X - X = \mathbb{T}$.

Proof of Furstenberg's theorem. Let X be an infinite closed invariant set.

Fix $\varepsilon > 0$; we shall show that X is ε-dense, and since ε is arbitrary and X is closed, this will mean $X = \mathbb{T}$.

Fix a prime number $p > 1/\varepsilon$.

Replace a by a^{p-1} and b by b^{p-1}. These are still multiplicatively independent, X is still jointly invariant, and now $T_a(1/p) = T_b(1/p) = 1/p$ by Fermat's little theorem.

Let X_1 denote what is left after removing all isolated points from X. This set is closed and stil T_a, T_b-invariant (since the image of a non-isolated point under T_a is non-isolated), so if X_1 is finite, it consitsts only of rational points. But then X would contain non-isolated rational points and by the previous lemma we would be done. So we can assume X_1 is infinite.

By the previous observation, $X_1 - X_1 = \mathbb{T}$ and in particular, $1/p \in X_1 - X_1$.

Let $X_2 = X_1 \cap (X_1 + 1/p)$ and note that $1/p \in X_1 - X_1$ implies $X_2 \neq \emptyset$.

Since $1/p$ is fixed by T_a, T_b we check that X_2 is jointly T_a, T_b-invariant.

If X_2 is finite, then it consists of rational points, and so X_1 contains a rational point (since $X_2 \subseteq X_1$). But all points in X_1 are non-isolated, so we conclude by the previous lemms that $X_1 = \mathbb{T}$ and hence $X = \mathbb{T}$.

So we may assume that X_2 is infinite.

We now repeat the argument. Remove all isolated points from X_2 and conclude that either we are left with a finite set of rationals, in which case $X = X_2 = \mathbb{T}$, or else, writing again X_2 for the set after the removel, $1/p \in X_2 - X_2$, and can define $X_3 = X_2 \cap (X_2 + 1/p)$.

We proceed in this way to define also $X_4 \supseteq X_5 \supseteq \ldots \supseteq X_p \neq \emptyset$ (we can do so unless at some point we find already that $X = \mathbb{T}$).

Now let $x_p \in X_p$. Then since $X_p = X_{p-1} \cap (X_{p-1} + 1/p)$ there exist $x_{p-1} \in X_{p-1}$ with $x_p = x_{p-1} + 1/p$.

We proceed recursively to find $x_i \in X_i$ with $x_{i+1} = x_i + 1/p$. Thus we have found that all the numbers

$$x_p - \frac{k}{p} \qquad k = 0, 1, \ldots, p-1$$

belong to X. This is a $1/p$-dense, and hence ε-dense, set, as required. $\square$