Chapter 2

Topological Dynamics

In essence, symbolic dynamical systems are dynamical systems on a topological (in fact metric) space, and therefore share many of the topological properties that general dynamical systems can have. In this chapter, we discuss several of these general topological properties, such as minimality, entropy, versions of equicontinuity and mathematical chaos, as well as topological mixing and shadowing properties.

2.1. Basic Notions from Dynamical Systems

A **dynamical system** is a mathematical description of how a physical system evolves in time. It consists of

- a **phase space** X, usually a metric space, or at least topological space, describing the state of the system. For example, $\mathbb{R}^{2n}$ can be used to describe the positions and velocities of n point-particles moving along a line, or $\mathbb{R}^{6n}$ for the positions and velocities of n point-particles moving in $\mathbb{R}^3$.
- a **time space**, which could be $\mathbb{R}$ (for continuous time) or $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ (or $\mathbb{Z}$ if the dynamical system is time-invertible) if the observations are only made at **discrete** time steps. More complicated (multi-dimensional or group-valued) time can be considered too, but in this text, time is always discrete: $\mathbb{N}_0$ or $\mathbb{Z}$.
- an **evolution rule**, which for discrete time take the form of a transformation $T : X \to X$ satisfying
  (1) $T^0(x) = x$ for all $x \in X$.
  (2) $T^{m+n}(x) = T^m(T^n(x))$ for all $m, n \in \mathbb{N}_0$ (or $\mathbb{Z}$) and all $x \in X$.
  This is realized if we let T^n be the n-fold composition:
  $$T^n(x) = \underbrace{T \circ T \circ \cdots \circ T}_{n \text{ times}}$$
  and T^{-n} is the n-fold composition of its inverse transformation if it exists. If T is continuous, then (X, T) is called a **continuous dynamical system**.

Definition 2.1. Let (X, T) be a dynamical space on a topological space. The **orbit** of $x \in X$ is the set

$$\operatorname{orb}(x) = \begin{cases} \{T^n(x) : n \in \mathbb{Z}\} & \text{if } T \text{ is invertible;} \\ \{T^n(x) : n \geq 0\} & \text{if } T \text{ is non-invertible.} \end{cases}$$

The set $\operatorname{orb}^+(x) = \{T^n(x) : n \geq 0\}$ is the **forward orbit** of x. This notation is useful if T is invertible; if T is non-invertible, then $\operatorname{orb}^+(x) = \operatorname{orb}(x)$.

Exercise 2.2. Let $\sigma : \Sigma \to \Sigma$ be invertible. Is there a difference between $x \in \overline{\operatorname{orb}(x) \setminus \{x\}}$ and $x \in \overline{\operatorname{orb}^+(x) \setminus \{x\}}$?

We distinguish several types of orbit. Namely, a point x is

- **periodic** if $T^n(x) = x$ for some $n \geq 1$. The smallest such n is called the **period** of x. If the period is 1, then x is called a **fixed point**.
- **preperiodic** if $T^{m+n}(x) = T^m(x)$ for some $m, n \in \mathbb{N}$. The minimal such m, n are called the **preperiod** and **period** of x.
- **asymptotically periodic** if there is a periodic point $y \notin \text{orb}(x)$ such that $d(T^n(x), T^n(y)) \to 0$ as $n \to \infty$. The periodic point y is **attracting** if it is periodic and has a neighborhood[1] U such that $\bigcap T^n(U) = \{y\}$. If y has a neighborhood U such that $\bigcap T^n(U) = \{y\}$, then y is **repelling**.

For example, for the quadratic family with $a = 3.83187405528332\dots$ as in Exercise 1.36, the point $x = \frac{1}{2}$ has period 3, and since $Q'_a(\frac{1}{2}) = 0$, it is easy to show that $\frac{1}{2}$ is attracting. The two fixed points are 0 and $1 - \frac{1}{a}$; they are **repelling**. For the circle rotation R_α, every point is periodic if and only if $\alpha \in \mathbb{Q}$; $x = m/n$ in lowest terms, then it is periodic with period n, and can be called **neutral**. If $\alpha \notin \mathbb{Q}$, then every orbit is dense in $\mathbb{S}^1$.

Definition 2.3. Let (X, T) be a dynamical space on a topological space. The ω**-limit set** of x is the set of accumulation points of its forward orbit. In formula

$$\omega(x) = \bigcap_{n \in \mathbb{N}} \overline{\bigcup_{m \geq n} T^m(x)} = \{y \in X : \exists\, n_i \to \infty, \lim_{i \to \infty} T^{n_i}(x) = y\}.$$

We call x **recurrent** if $x \in \omega(x)$.

Analogously for invertible shifts, the α**-limit set** of x is the set of accumulation points of its backward orbit of x:

$$\alpha(x) = \bigcap_{n \in \mathbb{N}} \overline{\bigcup_{m \leq -n} T^m(x)} = \{y \in X : \exists\, n_i \to \infty, \lim_{i \to \infty} T^{-n_i}(x) = y\}.$$

Definition 2.4. Given a dynamical system (X, T), a point $x \in X$ is called **nonwandering** if for every neighborhood $U \ni x$ there is an $n \geq 1$ such that $T^{-n}(U) \cap U \neq \varnothing$. The **nonwandering set**, $\Omega(T)$, is the set of all nonwandering points.

Recurrent points are always nonwandering, but $\Omega(T)$ can contain nonrecurrent points. In the one-sided full shift, for instance, $x = 0111111\cdots$ is not recurrent but nonwandering. If (X, T) has a dense orbit, then $\Omega(T) = X$.

[1] If the space X is one-dimensional, then we can speak of one-sided attracting if there is a one-sided neighborhood U of y such that $\bigcap T^n(U \cup \{y\}) = \{y\}$.

Definition 2.5. Two dynamical systems (X, f) and (Y, g) are **(topologically) conjugate** if there is a homeomorphism $\psi : X \to Y$ such that $\psi \circ f = g \circ \psi$.

If $\psi \circ f = g \circ \psi$ and $\psi : X \to Y$ is a continuous, onto, but not necessarily one-to-one map, then ψ is called a **semi-conjugacy** or **factor map**, (Y, g) is called a **factor** of (X, f), and (X, f) is called an **extension** of (Y, g). This extension is **almost one-to-one** if there is a dense set Y' such that $\#\psi^{-1}(y) = 1$ for all $y \in Y'$.

A conjugacy $\psi : X \to Y$ is called **pointed** if it sends specified points $x \in X$ and $y \in Y$ to each other.

Lemma 2.6. *Let (X, f) and (Y, g) be dynamical systems that are conjugate via $g \circ \psi = \psi \circ f$. Then*

(1) If p is a (pre)periodic point for f, then $\psi(p)$ is a (pre)periodic point of g, and the (pre)periods are the same.

(2) If f, g are continuous, then the conjugacy preserves ω-limit sets: $\psi(\omega(x)) = \omega(\psi(x))$.

(3) If the periodic point p is attracting/repelling, then $\psi(p)$ is also attracting/repelling.

Proof. First note that

$$\begin{aligned} \psi \circ f^n &= \psi \circ f \circ \psi^{-1} \circ \psi \circ f \circ \psi^{-1} \circ \psi \circ f \circ \psi^{-1} \circ \cdots \circ f \\ &= g \circ \psi \circ \psi^{-1} \circ g \circ \psi \circ \psi^{-1} \circ g \circ \psi \circ \psi^{-1} \circ \cdots \circ g \circ \psi = g^n \circ \psi. \end{aligned}$$

1. Take p such that $f^n(p) = p$ and $q = \psi(p)$. Then $g^n(q) = g^n \circ \psi(p) = \psi \circ f^n(p) = \psi(p) = q$, so q if n-periodic for g. Next, suppose that $f^{m+n}(p) = f^m(p)$, and set $\psi(p) = q$. Then $g^{m+n}(q) = g^{m+n} \circ \psi(p) = \psi \circ f^{m+n}(p) = \psi \circ f^m(p) = g^m \circ \psi(p) = g^m(q)$.
2. Now assume that $x \in \omega_f(a)$, so there is a sequence $n_k \to \infty$ such that $f^{n_k}(a) \to x$. Set $y = \psi(x)$ and $b = \psi(a)$. Then, by continuity of f, $g^{n_k}(b) = g^{n_k} \circ \psi(a) = \psi \circ f^{n_k}(a) \to \psi(x) = y$, so $y \in \omega_g(b)$.
3. If $p = f(p)$ is asymptotically attracting, then for every $a \in X$ sufficiently close to p, we have $p = \omega_f(a)$. By part 1., $q := \psi(p)$ is a fixed point of g, and by part 2., $q = \omega_g(y)$ for $y = \psi(x)$. □

Exercise 2.7. Is the following true? If X is a factor of Y and Y a factor of X, then X and Y are conjugate.

Example 2.8. The quadratic **Chebyshev polynomial** $Ч_2(y) = 2y^2 - 1$ on $[-1, 1]$ is conjugate to the tent map $T(x) = \min\{2x, 2(\pi - x)\}$ on $[0, \pi]$. Indeed,

$$Ч_2 \circ \psi = \psi \circ T \quad \text{for} \quad \psi(x) = \cos x, \tag{2.1}$$

It is very unusual to find smooth conjugacies between maps, and even here, ψ is not diffeomorphic at the endpoints $0, 1$. But applying (2.1) n times and then differentiating, we find

$$(Ч_2^n)' \circ \psi(x) \cdot \psi'(x) = \psi'(T^n(x)) \cdot (T^n)'(x).$$

If x is a p-periodic point of T, and hence $y = \psi(x)$ an p-periodic point of $Ч_2$, we see that $|(Ч^p)'(y)| = 2^p$. The only periodic point where this fails is $y = \psi(0) = 1$, because $\psi'(0) = 0$.

Note that the same conjugacy works for the degree n Chebyshev polynomial $Ч_n$ and the slope n tent map with n branches. The characterization of Chebyshev polynomials $Ч_n(x) = \cos(n \arccos x)$ is the cause of this.

Example 2.9. We show that two circle rotations R_α and R_β are not conjugate if $0 \leq \alpha < \beta < 1$. Let $<$ denote the positive orientation on $\mathbb{S}^1$. Choose $n \in \mathbb{N}$ such that $n\alpha \leq k < n\beta$ and $(n-1)\beta \leq k$ for some integer k. Then, setting $y = \psi(0)$,

$$R_\alpha^n(0) \leq 0 \leq R_\alpha(0) \quad \text{and} \quad y \leq R_\beta^n(y) \leq R_\beta(y). \tag{2.2}$$

The homeomorphism $\psi : \mathbb{S}^1 \to \mathbb{S}^1$ must either preserve or reverse the orientation of the circle, but neither way is compatible with (2.2). Therefore there cannot be any conjugacy.

A more structural way to see this is using lifts and rotation numbers, see Theorem 4.54. Indeed, the rotation number $\rho(f)$ is preserved on conjugacy, and $\rho(R_\alpha) = \alpha \neq \beta = \rho(R_\beta)$.

Definition 2.10. Two dynamical systems (X, f) and (Y, g) are called **orbit equivalent** if there is a homeomorphism $\psi : X \to Y$ such that $\psi(\text{orb}_f(x)) = \text{orb}_g(\psi(x))$ for all $x \in X$, i.e., ψ sends orbits to orbits (set-wise, not necessarily point-wise).

Clearly, a conjugacy is an orbit equivalence. If f and g are themselves homeomorphisms, and $\psi \circ f = g^{-1} \circ \psi$, then ψ is called a **flip-conjugacy** and this is also an orbit equivalence. More generally, if ψ is a conjugacy or flip-conjugacy, then $\psi \circ f^k$ is an orbit equivalence for each $k \in \mathbb{Z}$.

Orbit equivalence implies the existence of two functions $\boldsymbol{m}, \boldsymbol{n} : X \to \mathbb{Z}$, called **orbit cocycles**, such that

$$\psi \circ f(x) = g^{\boldsymbol{n}(x)} \circ \psi(x) \qquad \text{and} \qquad \psi \circ f^{\boldsymbol{m}(x)} = g \circ \psi(x).$$

Thus the orbit cocycles of a conjugacy are constant 1 and of a flip conjugacy are constant -1. Another special case of orbit equivalence is a **speed-up**: (Y, g) is a speed-up of (X, f) if it is orbit equivalent and the orbit cocycle $\boldsymbol{m} : X \to \mathbb{Z}$ is non-negative.

Definition 2.11. Two dynamical systems (X, f) and (Y, g) are **strongly orbit equivalent** if their orbit cocycles are continuous on X, except for at most one point each.

2.2. Transitive and Minimal Systems

The following definition expresses that all parts of a dynamical system connect to each other:

Definition 2.12. A dynamical system (X, T) is **(topologically) transitive** if for every two non-empty open[2] sets $U, V \subset X$, there is an $n \geq 0$ such that $U \cap T^{-n}(V) \neq \varnothing$.[3] It is called **totally transitive** if T^N is transitive for each $N \in \mathbb{N}$.

Clearly totally transitive implies transitive. The other implication fails; for example, σ is transitive on the two-point subshift $\{(10)^\infty, (01)^\infty\}$ but σ^2 is not.

Proposition 2.13. Let X be a compact regular Hausdorff space[4] without isolated points and which is **second countable**, i.e., it possesses a countable basis of its topology. A continuous map $T : X \to X$ is a topologically transitive map if and only if there is a dense orbit.

Remark 2.14. The notion of dense orbit may need further explanation if the subshift is two-sided. Consider the sequence

$$x = \cdots 000000000000000000.101000101000000000101000101 \cdots \tag{2.3}$$

This sequence emerges from the Cantor substitution

$$\chi_{\text{Cantor}} : \begin{cases} 0 \to 000 \\ 1 \to 101 \end{cases}$$

from the seed 0.1. This sequence has a dense forward orbit $\text{orb}^+(x)$ within its forward orbit closure $\overline{\text{orb}^+(x)}$ as well as a dense backward orbit $\text{orb}^-(x)$ within its forward orbit closure $\overline{\text{orb}^-(x)}$. However, $\text{orb}^-(x)$ is not dense in its two-sided orbit closure.

Proof. Let $\{U_j\}_{j \in \mathbb{N}}$ be a countable basis of the topology. Let $U, V \subset X$ be arbitrary open sets. Take $j, k \in \mathbb{N}$ such that $U_j \subset U$, $U_k \subset V$. Since $\text{orb}(x)$ is dense, and X has no isolated points, x visits each U_j infinitely

[2] Some authors use the abbreviation **opene** for open and non-empty.

[3] Many texts write $T^n(U) \cap V \neq \varnothing$, which may be more intuitive but the fact that $T^n(U)$ need not be open (or not measurable even if U is measurable) might in some case lead to inadvertent problems.

[4] Regular Hausdorff means that singletons $\{x\}$ are closes and for all closed sets A and $x \notin A$ there are neighborhoods $U \ni x$ and $V \supset A$ such that $\overline{U} \cap \overline{V} = \varnothing$.

often. Hence there is $m, n \in \mathbb{N}$ such that $T^m(x) \in U_j$ and $T^{m+n}(x) \in U_k$. This shows that $U \cap T^{-n}(V) \neq \varnothing$.

Conversely, by topological transitivity applied to U_1 and U_2, we can find n_1 such that $U_1 \cap T^{-n_1}(U_2) \neq \varnothing$. By continuity of T, $U_1 \cap T^{-n_1}(U_2)$ is an open set. Choose V_2 open such that $\overline{V}_2 \subset U_1 \cap T^{-n_2}(U_2)$. Here we use the regular Hausdorff property of X.

Next, using topological transitivity applied to V_2 and U_3, choose $n_2 > n_1$ such that $V_2 \cap T^{-n_2}(U_3) \neq \varnothing$. Then choose an open set V_3 such that $\overline{V_3} \subset V_2 \cap T^{-n_2}(U_3)$.

Continuing this way we find a nested sequence of open sets V_k, with $\overline{V}_k \subset V_{k-1}$, and a sequence of integers (n_k) such that $V_k \subset T^{-n_k}(U_{k+1})$.

Let $V_\infty = \cap_k V_k$. Since $\overline{V}_k \subset V_{k-1}$, and closed sets in X are automatically compact, this intersection is non-empty, and every $x \in V_\infty$ has a dense orbit. This concludes the proof. □

A strong form of transitivity is minimality:

Definition 2.15. A dynamical system (X, T) is **minimal** if every orbit is dense in X.

Remark 2.16. It is a straightforward application of Zorn's Lemma that every dynamical system on a compact space[5] contains at least one minimal subsystem. For compact metric spaces, this fact can also be shown without the use of Zorn's Lemma, see [**296**, Chapter 1, Theorem 2.2.1].

Proposition 2.17. Let X be a compact topological space. We have the following equivalent characterizations for a continuous dynamical system (X, T) to be **minimal**:

(i) There is no closed T-invariant proper subset of X;
(ii) Every orbit is dense in X;
(iii) There is a dense orbit and T is **uniformly recurrent**[6], i.e., for every open set $U \subset X$ there is an $N \in \mathbb{N}$ such that for every $x \in U$ there is $1 \le n \le N$ such that $T^n(x) \in U$.

Proof. We prove the three implications by the contrapositive.
(i) ⇒ (ii): Suppose that $x \in X$ has an orbit that is not dense. Then $\overline{\text{orb}(x)}$ is a T-invariant closed proper subset, so (i) fails.

[5] Compactness is important, otherwise one could take a single non-recurrent orbit (without its closure) as the phase-space. An interesting example with only recurrent orbits but no minimal subset is due to Auslander [**37**, page 27].

[6] The word "almost periodic" is frequently used as well, e.g. in [**278**, **369**, **386**, **450**], but it is not the same with all authors, and sometimes refers to a different notion. For instance, in [**467**] it is used as "periodically recurrent" in Definition 2.19.

(ii) ⇒ (iii): By (ii) every orbit is dense, so there is at least one dense orbit.

Now to prove uniform recurrence, let U be any open set and U_0 an open subset such that $\overline{U_0} \subset U$.

Suppose that for every $N \in \mathbb{N}$ there is $x_N \in U_0$ such that $T^n(x_N) \notin U_0$ for all $1 \le n \le N$. Let $x \in \overline{U_0} \subset U$ be an accumulation point of $(x_N)_{N\in\mathbb{N}}$. Suppose by contradiction that there is $n \ge 1$ such that $T^n(x) \in U_0$. Take an open set $V \ni x$ such that $T^n(V) \subset U_0$. Next take $N \ge n$ so large that $x_N \in V$. But this means that $T^n(x_N) \in U_0$, which is against the definition of x_N. Hence no such n exists, and therefore $\mathrm{orb}(x)$ is not dense, and (ii) fails.
Now take $y \in U$ arbitrary (so not necessarily in U_0), and take $x \in U_0$ with a dense orbit. Find a sequence k_i such that $T^{k_i}(x) \to y$. For each i there is $1 \le n_i \le N$ such that $T^{k_i+n_i}(x) \in U_0$. Passing to a subsequence, we may as well assume that $n_i \equiv n$. Then $T^n(y) = T^n(\lim_i T^{k_i}(x)) = \lim_i T^{k_i+n}(x) \in U_0 \subset U$. This proves the uniform recurrence of U.

(iii) ⇒ (i): Let x be a point with a dense orbit. Suppose that Y is a closed T-invariant proper subset of X and let $U \subset X$ be nonempty open such that $\overline{U} \cap Y = \varnothing$. Let $n \ge 0$ be minimal such that $u := T^n(x) \in U$. Let $N = N(U) \ge 1$ be as in the definition of uniform recurrence, and let $y \in Y$ be arbitrary. Since $\mathrm{orb}(y) \subset Y$, there is an open set $V \ni y$ such that $V \cap T^{-i}(U) = \varnothing$ for $0 \le i \le N$.

Take $n'' > n$ minimal such that $T^{n''}(u) \in V$, and let $n' < n''$ be maximal such that $T^{n'}(u) =: u' \in U$. Then $T^i(u') \notin U$ for all $1 \le i \le n'' - n' + N$. Since N was arbitrary, this contradicts the uniform recurrence and hence such Y cannot exist. □

Definition 2.18. Uniform recurrence means that the set

$$\mathcal{N}(x, U) := \{n \in \mathbb{Z} \text{ or } \mathbb{N} : x \in T^{-n}(U)\}$$

is **syndetic** for every $x \in X$, i.e., it has **bounded gaps** (from the Greek $\sigma\upsilon\nu\delta\varepsilon\tau\iota\kappa o\varsigma$ = bound together). A set that is not syndetic has a complement that is **thick**: for every $N \in \mathbb{N}$ it contains blocks $\{n, n+1, \dots, n+N\}$.

Definition 2.19. A dynamical system is called **periodically recurrent** if for every nonempty open set U, there is N such that $U \subset T^{-kN}(U)$ for all $k \in \mathbb{N}$ (or $k \in \mathbb{Z}$ if T is invertible).

Since periodic recurrence is obviously stronger than uniform recurrence, we have the following corollary.

Corollary 2.20. Every periodically recurrent dynamical system is minimal.

Definition 2.21. Given a dynamical system (X,T), a point $x \in X$ is **uniformly recurrent** (resp. **periodically recurrent**) if for every neighborhood $U \ni x$, the set $\mathcal{N}(x,U)$ is syndetic (resp. $\mathcal{N}(x,U) \supset \{bk : k \in \mathbb{N} \text{ or } \mathbb{Z}\}$ for some $b \in \mathbb{N}$).

Corollary 2.22. Let (X,T) be a continuous dynamical system and $x \in X$ have a dense orbit. Then (X,T) is minimal (resp. periodically recurrent) if and only if x is uniformly recurrent (resp. periodically recurrent).

Proof. If (X,T) is minimal, then x is uniformly recurrent by Proposition 2.17, part (iii).

Conversely, assume that x is uniformly recurrent. First observe that every $u \in \text{orb}(x)$ is uniformly recurrent too. Indeed, suppose $u = T^n(x)$, and let V be an open neighborhood of x. Then for every open neighborhood U of u, also $U' = T^{-n}(U) \cap V$ is an open neighborhood of x, and $\mathcal{N}(u,U) \supset \mathcal{N}(x,U') + n$. Now minimality of (X,T) follows precisely as in the step (iii) $\Rightarrow$ (i) in the proof of Proposition 2.17.

The proof for x periodically recurrent is analogous. □

Definition 2.23. A dynamical system (X,T) on a metric space (X,d) is **uniformly rigid** if for every $\varepsilon > 0$ there is an iterate $n \geq 1$ such that $d(T^n(x),x) < \varepsilon$ for all $x \in X$.

Lemma 2.24. *A continuous dynamical system (X,T) on a Cantor set (or compact zero-dimensional set) is uniformly rigid if and only if it is periodically recurrent.*

For this result, it is important that the space X is zero-dimensional. For example, irrational rotations R_α on the circle are uniformly rigid but only uniformly, so not periodically, recurrent. The uniform rigidity follows immediate because a circle rotation is an isometry and every point is recurrent. But periodic recurrence fails because for every $n \in \mathbb{N}$ and $x \in \mathbb{S}^1$, the set $\{R_\alpha^{kn}(x) : k \in \mathbb{N}\}$ is dense in $\mathbb{S}^1$. The below proof, however, shows that a periodically recurrent dynamical system on a compact space is uniformly rigid.

Proof. $\Rightarrow$: Take $\varepsilon > 0$ arbitrary with corresponding iterate $n \geq 1$, and let $k \in \mathbb{N}$ be the smallest integer such that $2^{-k} < \varepsilon$. Thus the distance between every two distinct k-cylinders Z in X is at least ε. By uniform rigidity $T^n(Z) = Z$, and therefore $T^{kn}(Z) = Z$ for all $k \geq 0$, proving periodic recurrence.
$\Leftarrow$: Let $\varepsilon > 0$ be arbitrary. For each $x \in X$, we can find a neighborhood U_x of $\text{diam}(U_x) < \varepsilon$ and iterate n_x such that $T^{n_x}(U_x) \subset U_x$. By compactness, there is a finite collection $x_1, \dots, x_N$ such that $X = \bigcup_{i=1}^N U_{x_i}$. Take $n = \text{lcm}\{n_{x_1}, \dots, n_{x_N}\}$. Then $d(T^n(x),x) < \varepsilon$ for each $x \in X$, as required. □

The following weakening of minimality is of importance for e.g. Toeplitz shifts and $\mathcal{B}$-free shifts, see Sections 4.5 and 4.6.

Definition 2.25. A dynamical system (X,T) is called **essentially minimal** if it contains a unique minimal set Y, i.e., a unique non-empty closed set Y such that $T(Y) = Y$.

Clearly, essentially minimal maps can have at most one periodic orbit, but as the subshift $X := \{\sigma^k(\cdots 000001000000 \cdots)\}_{k\in\mathbb{Z}} \cup \{0^\infty\}$ shows, $X \setminus Y \neq \varnothing$ is possible. However, the two-sided orbit closure of (2.3) does not give an essentially minimal shift.

Proposition 2.26. Given a dynamical system (X,T) and a point $y \in X$, the following are equivalent:

(i) (X,T) is essentially minimal and y is contained in its minimal set.
(ii) For every $x \in X$, $\omega(x) \ni y$.

If in addition, T is invertible, then two further equivalent statements are:

(iii) For every $x \in X$, $\alpha(x) \ni y$.
(iv) For every open set $U \ni y$, $\cup_{n\in\mathbb{Z}} T^n(U) = X$.

Proof. **(i) ⇒ (ii)**: $\omega(x)$ is a closed non-empty T-invariant set, so by Zorn's Lemma, it contains a minimal set. But Y is the unique minimal set, so $y \in \omega(x)$.
(ii) ⇒ (i): Assume by contradiction that $y \in Y$ and Y' are minimal sets, and take $x \in Y$, $x' \in Y'$. By assumption $y \in \omega(x) \cap \omega(x')$, so $y \in Y \cap Y'$. Thus $Y \cap Y'$ is a non-empty, closed and T-invariant subset of both Y and Y'. Since Y and Y' are minimal, $Y = Y \cap Y' = Y'$.
(i) ⇔ (iii): Use the above with T^{-1} instead of T.
(i) ⇒ (iv): Let U be an arbitrary neighborhood of y. Since $\cup_{n\in\mathbb{Z}} T^n(U)$ is an open (two-sided!) T-invariant set, its complement Y' is closed and T-invariant. If $Y' \neq \varnothing$, then it contains a minimal set that is disjoint from y, contradicting essential minimality. Hence $\cup_{n\in\mathbb{Z}} T^n(U) = X$.
(iv) ⇒ (iii): Let $x \in X$ be arbitrary; we can assume without loss of generality that $x \neq T^k(y)$ for all $k \geq 0$, because if y is periodic then $\alpha(x) = \mathrm{orb}(y) \ni y$, and otherwise we replace x by $T^{-(k+1)}(x)$ to get it outside the forward orbit of y. Let $(U_r)_{r\in\mathbb{N}}$ be a nested sequence of neighborhoods of y such that $\cap_r U_r = \{y\}$. Since $\cup_{n\in\mathbb{Z}} T^n(U_r) = X$ and X is compact, there is a finite N_r such that $\cup_{n=-N_r}^{N_r} T^n(U) = X$. Applying T^{N_r} to both sides, we obtain $\cup_{n=0}^{2N_r} T^n(U) = X$. Thus there is $n_r \leq 2N_r$ such that $T^{-n_r}(x) \in U_r$. As we can do this for every r, we have found a sequence (n_r) (and $n_r \to \infty$ because $x \neq T^k(y)$ for any $k \geq 0$) such that $T^{-n_r}(x) \to y$. Thus $y \in \alpha(x)$, as required. □

2.3. Equicontinuous and Distal Systems

The opposite to expansive (recall Definition 1.38) is equicontinuous.

Definition 2.27. A dynamical system (X, T) on a metric space (X, d) is called **equicontinuous** if for all $\varepsilon > 0$ there exists $\delta > 0$ such that if $d(x, y) < \delta$, then $d(T^n(x), T^n(y)) < \varepsilon$ for all $n \geq 0$ (or $n \in \mathbb{Z}$ if T is invertible). This is sometimes also called **Lyapunov stability**.

Naturally, if T is not injective, then distality fails immediately. Every **isometry**, i.e., a dynamical system such that $d(T(x), T(y)) = d(x, y)$ for all $x, y \in X$ is equicontinuous.

Exercise 2.28. Let (X, T) be an equicontinuous dynamical system. Show that it is topologically transitive if and only if it is minimal.

Lemma 2.29. *Let (X, T) be an equicontinuous surjection on a compact metric space (X, d). Then the nonwandering set $\Omega(T) = X$.*

Proof. Suppose by contradiction that $x \in X$ is wandering, i.e., there is an $\varepsilon > 0$ such that $T^k(B_\varepsilon(x)) \cap B_\varepsilon(x) = \varnothing$ for all $k \geq 1$. In particular, x is not periodic. By equicontinuity, there is $\delta > 0$ such that $d(a, b) < \delta$ implies $d(T^n(a), T^n(b)) < \varepsilon/2$ for all $n \geq 0$. Construct a backward orbit $(x_{-n})_{n\geq 0}$, i.e., $T^n(x_{-n}) = x$ and $T^k(x_{-n}) \notin B_\varepsilon(x)$ for all $k \in \mathbb{N} \setminus \{n\}$. By compactness of X, $(x_{-n})_{n\geq 0}$ has an accumulation point $y \in X$. Let $m < n$ be so that $d(y, x_{-m}) < \delta$ and $d(y, x_{-n}) < \delta$. Then $T^n(x_{-n}) = x \in B_\varepsilon(x)$ and $T^n(x_{-m}) = T^{n-m}(x) \notin B_\varepsilon(x)$, so $d(T^n(x_{-m}), T^n(y)) \geq \varepsilon/2$ or $d(T^n(x_{-n}), T^n(y)) \geq \varepsilon/2$. This contradicts equicontinuity of T and hence there cannot be a wandering point. □

Lemma 2.30. *If an equicontinuous dynamical system (X, T) on a compact metric space (X, d) is topologically transitive, then it is uniformly rigid.*

See [**315**, Proposition 1.1] for more general results in this direction.

Proof. Suppose $z \in X$ that has a dense orbit. Take $\varepsilon > 0$ arbitrary and choose $\delta \in (0, \varepsilon/3)$ such that $d(x, y) < \delta$ implies $d(T^n(x), T^n(y)) < \varepsilon/3$ for all $n \geq 0$. Choose $N \in \mathbb{N}$ so large that $\bigcup_{n=0}^{N-1} B_\delta(T^n(z)) = X$ and $d(T^N(z), z) < \delta$. Now let x arbitrary and take $0 \leq n < N$ such that $d(T^n(z), x) < \delta$. Then

$$\begin{aligned} d(T^N(x), x) &\leq d(T^N(x), T^{N+n}(z)) + d(T^{n+N}(z), T^n(z)) + d(T^n(z), x) \\ &\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \delta < \varepsilon \end{aligned}$$

as required. □

If $T : X \to X$ is equicontinuous on a metric space (X, d), then the metric

$$d_\infty(x, y) := \sup_{n \geq 0} d(T^n(x), T^n(y))$$

is well-defined (i.e., not infinite) and non-expanding because

$$d_\infty(T(x), T(y)) = \sup_{n \geq 1} d(T^n(x), T^n(y)) \leq \sup_{n \geq 0} d(T^n(x), T^n(y)) = d_\infty(x, y).$$

However, equicontinuity also gives for every $\varepsilon > 0$ some $\delta > 0$ (and $\delta \to 0$ as $\varepsilon \to 0$) such that $d(x, y) < \delta$ implies $d_\infty(x, y) < \varepsilon$, and therefore $x_n \to x$ in the metric d if and only if $x_n \to x$ in the metric d_∞. Hence both metrics generate the same topology.

If T is itself a strict contraction, then also $d_\infty(T(x), T(y)) < d_\infty(x, y)$, but if X is compact and T surjective, then the dynamical system (X, T) is an isometry in the metric d_∞.

Proposition 2.31. If a dynamical system (X, T) is equicontinuous and surjective on a compact metric space (X, d), then T preserves d_∞.

Proof. We already have seen that $d_\infty(T(x), T(y)) \leq d_\infty(x, y)$ for all $x, y \in X$. Assume by contradiction that we have strict inequality for some choice $a \neq b$, say $d_\infty(a, b) = d_\infty(T(a), T(b)) + 9\varepsilon$ for some $\varepsilon > 0$.

Consider the product system $T_2 : X^2 \to X^2$ with metric

$$d_2((x, x'), (y, y')) := \max\{d_\infty(x, y), d_\infty(x', y')\}.$$

Clearly T_2 is non-expanding on (X^2, d_2). Let $B \subset X^2$ be the ε-ball w.r.t. d_2 around (a, b). So, if $(x, y) \in B$, then $d_\infty(x, a) < \varepsilon$ and $d_\infty(y, b) < \varepsilon$. Assume by contradiction that there is $n \geq 1$ such that $B \cap T_2^n(B) \neq \varnothing$. This would mean that $d_\infty(T^n(x), a) < 3\varepsilon$ and $d_\infty(T^n(y), b) < 3\varepsilon$. But then

$$\begin{aligned} d_\infty(a, b) &\leq d_\infty(a, T^n(x)) + d_\infty(T^n(x), T^n(y)) + d_\infty(T^n(y), b) \\ &\leq 3\varepsilon + d_\infty(T(x), T(y)) + 3\varepsilon \\ &\leq 6\varepsilon + d_\infty(T(x), T(a)) + d_\infty(T(a), T(b)) + d_\infty(T(b), T(y)) \\ &\leq 3\varepsilon + \varepsilon + d_\infty(a, b) - 9\varepsilon + \varepsilon = d_\infty(a, b) - \varepsilon. \end{aligned}$$

This contradiction shows that $T_2^n(B) \cap B = \varnothing$ for all $n \geq 1$. But then (a, b) is a wandering point for T_2, contradicting Lemma 2.29. □

Related notions to equicontinuity are distality and its opposite: proximality.

Definition 2.32. A dynamical system (X, T) on a metric space (X, d) is **distal** if $\liminf_n d(T^n(x), T^n(y)) > 0$ for every $x \neq y$. Conversely, a pair $(x, y) \in X^2$ is called **proximal** if $\liminf_n d(T^n(x), T^n(y)) = 0$. That is, a distal dynamical system has no proximal pairs (except (x, x)). A dynamical system (X, T) is called **proximal** if every pair $(x, y) \in X^2$ is proximal.

Auslander & Ellis (see e.g. [**13**]) proved that for every $x \in X$, there exists a $y \in X$ such that $\overline{\mathrm{orb}(y)}$ is a minimal subset of X and (x, y) is a proximal pair. Note that proximality is not an equivalence relation: it is not transitive. For example, $(101)(000)^2(101)^3(000)^4 \cdots$ and $(000)(101)^2(000)^3(101)^4 \cdots$ are both proximal to 0^∞ under the shift, but not to each other. A stronger version of proximality that does give an equivalence relation the following:

Definition 2.33. Let (X, T) be a dynamical system on a metric space (X, d). Then a pair of points (x, y) is **syndetically proximal** if for every $\varepsilon > 0$, the set $\{n \in \mathbb{N} \text{ or } \mathbb{Z} : d(T^n(x), T^n(y)) < \varepsilon\}$ is syndetic.

The following result for subshifts goes back to [**152**, **542**], see also [**421**, Theorem 19] for the proof.

Theorem 2.34. *Given a subshift (X, σ), the following are equivalent.*

(1) Proximality is an equivalence relation.

(2) Every proximal pair is syndetically proximal.

(3) The orbit closure $\overline{\{\sigma^n \times \sigma^n(x, y) : n \in \mathbb{N} \text{ or } \mathbb{Z}\}}$ of every $(x, y) \in X \times X$ contains exactly one minimal set in the product shift.

Distality doesn't imply equicontinuity, see Exercise 2.37. Neither does equicontinuity imply distality; think of $T(x) = x/2$ on $X = [0, 1]$ or on $X = \mathbb{R}$. However:

Corollary 2.35. Every equicontinuous surjection (X, T) on a compact metric space (X, d) is distal.

Proof. Assume by contradiction that T is not distal. Then there are $x \neq y$ and a sequence $(n_k)_{k \in \mathbb{N}}$ such that $d(T^{n_k}(x), T^{n_k}(y)) \to 0$. Since X is compact, and passing to a subsequence, we can assume $\lim_k T^{n_k}(x) = \lim_k T^{n_k}(y) = z$ in the metric d. But then also $\lim_k T^{n_k}(x) = \lim_k T^{n_k}(y) = z$ in the metric d_∞, and this contradicts that T is an isometry in d_∞. □

In particular, equicontinuous surjections on compact metric space are invertible, because distal dynamical systems are.

Corollary 2.36. An equicontinuous surjection (X, T, d) on a compact metric space has an equicontinuous inverse.

Proof. Take $K_\varepsilon = \{(x, y) \in X^2 : d(x, y) \geq \varepsilon\}$ for $\varepsilon > 0$. We claim that

$$\delta(\varepsilon) := \inf\{d(T^n x, T^n y) : (x, y) \in K_\varepsilon, n \in \mathbb{N}\} > 0.$$

Indeed, assume by contradiction that there are sequences $(x_k, y_k) \subset K_\varepsilon$ and $(n_k) \subset \mathbb{N}$ such that $d(T^{n_k} x_k, T^{n_k} y_k) \leq 1/k$ for all $k \in \mathbb{N}$ and $(x_k, y_k) \to (x_\infty, y_\infty) \in K_\varepsilon$. By Corollary 2.35, T is distal, so $\eta := \inf\{d(T^n(x_\infty), T^n(y_\infty)) :$

$n \geq 0\} > 0$. By equicontinuity, there is $\gamma(\eta) > 0$ is such that $d(x,y) < \gamma(\eta)$ implies that $d(T^n(x), T^n(y)) < \eta/3$ for all $n \geq 0$. Take $k > 3/\eta$ so large that $(x_k, y_k) \in B_{\gamma(\eta)}(x_\infty, y_\infty)$. Then by the triangle inequality

$$\begin{aligned} d(T^{n_k}(x_\infty), T^{n_k}(y_\infty)) &\leq d(T^{n_k}(x_\infty), T^{n_k}(x_k)) + d(T^{n_k}(x_k), T^{n_k}(y_k)) \\ &\quad + d(T^{n_k}(y_k), T^{n_k}(y_\infty)) \\ &< \eta/3 + \eta/3 + \eta/3 = \eta, \end{aligned}$$

contradicting the choice of η. Hence two points $u, v \in X$ with $d(u,v) < \delta(\varepsilon)$ have $d(T^{-n}(u), T^{-n}(v)) < \varepsilon$ for all $n \in \mathbb{N}$. This is equicontinuity of T^{-1}. □

Exercise 2.37. a) Show that the map $T(x,y) = (x, x+y)$ on the two-torus $\mathbb{T}^2$ is distal but not equicontinuous.
b) Let $\alpha \in [0,1]$ be irrational. Show that the map $T(x,y) = (x+\alpha, x+y)$ on the two-torus $\mathbb{T}^2$ is distal but not equicontinuous. (Here showing minimality is the hard part, see Proposition 6.26).

Proposition 2.38. Every subshift (X, σ) with a non-periodic minimal set is proximal (so not equicontinuous by Corollary 2.35).

The non-periodicity is essential, otherwise $X = \{(01)^\infty, (10)^\infty\}$ is an equicontinuous counterexample. Non-periodicity implies in particular that X is uncountable.

Proof. First assume that the shift is one-sided. If it is distal, then it has to be invertible, and therefore a homeomorphism. But a one-sided shift is locally expanding, and locally expanding homeomorphisms only exist on finite spaces, see Proposition 1.41. Hence, there are no distal one-sided shifts other than finite unions of periodic orbits.

Now if (X, σ) is a two-sided shift, then its one-sided restriction (X^+, σ) is a subshift too. Here we need to check that $\sigma : X^+ \to X^+$ is surjective, but this follows because if x^+ is the one-sided restriction of $x \in X$, then $y^+ := \sigma^{-1}(x)^+ \in X^+$ and $\sigma(y) = x$. Furthermore, since X has a non-periodic minimal set, X^+ has a non-periodic minimal set too. Thus the above argument shows that (X^+, σ) cannot be distal. □

Definition 2.39. Given a dynamical system (X, T), we say that (Y, S) is the **maximal equicontinuous factor (MEF)** if it is equicontinuous and semi-conjugate to (X, T), and every other equicontinuous factor of (X, T) is also a factor of (Y, S).

Every dynamical system has a MEF, and it can be shown that the MEF is unique up to conjugacy. This goes back to a result of Ellis & Gottschalk

[**233**]. The proof we give is for invertible dynamical systems[7] and relies on the notion of regional proximality:

Definition 2.40. Let (X,T) be a dynamical system on a metric space (X,d). Two points $x,y \in X$ are **regionally proximal** if there are sequences $x_i \to x$ and $y_i \to y$ and $(n_i) \subset \mathbb{N}$ such that $d(T^{n_i}(x_i), T^{n_i}(y_i)) \to 0$. In this case we write $x \sim_{\rm rp} y$. It is not obvious that $\sim_{\rm rp}$ is a transitive relation[8], and therefore we take the transitive hull $x \sim_{\rm trp} y$ if there is a sequence $x = z_0 \sim_{\rm rp} z_1 \sim_{\rm rp} \cdots \sim_{\rm rp} z_N = y$.

Proposition 2.41. Every continuous invertible dynamical system (X,T) on a compact metric space (X,d) has a maximal equicontinuous factor.

Proof. First we note that if (X,T) is equicontinuous and $x \sim_{rp} y$, then $x = y$. Indeed, otherwise for any $\varepsilon > 0$ and $\delta = \delta(\varepsilon)$ as in the definition of equicontinuity, there is $x_i \in B_\delta(x)$, $y_i \in B_\delta(y)$ and n_i such that $d(T^{n_i}(x_i), T^{n_i}(y_i)) < \varepsilon$. But then also

$$\begin{aligned} d(T^{n_i}(x), T^{n_i}(y)) &\leq d(T^{n_i}(x), T^{n_i}(x_i)) + d(T^{n_i}(x_i), T^{n_i}(y_i)) \\ &\quad + d(T^{n_i}(y_i), T^{n_i}(y)) < 3\varepsilon. \end{aligned}$$

Therefore (x,y) is not a distal pair, but equicontinuous maps are distal, see Corollary 2.35.

The (transitive hull) relation $\sim_{trp}$ is an equivalence relation that is T-invariant and also T^{-1}-invariant. The equivalence classes are closed, and if $x_k \to x$, $y_k \to y$ are such that $x_k \sim_{\rm trp} y_k$, then also $x \sim_{\rm trp} y$. Therefore the quotient space $X_{\rm eq} = X/\sim_{\rm trp}$ is a well-defined Hausdorff space (and in fact metric space with quotient metric $d_{\rm eq}$), and the maps T and T^{-1} are well-defined on it.

Now suppose by contradiction that T and hence T^{-1} is not equicontinuous on the quotient space $X_{\rm eq}$. Then there is $\varepsilon > 0$ such that for all $i \in \mathbb{N}$, there are $x_i', y_i' \in X_{eq}$, $d_{\rm eq}(x_i', y_i') < 1/i$ and $n_i \in \mathbb{N}$ such that $d_{\rm eq}(x_i, y_i) > \varepsilon$ for $x_i = T^{-n_i}(x_i')$ and $y_i = T^{-n_i}(y_i')$. By passing to a subsequence, we can assume that $x_i \to x$ and $y_i \to y$ and $d_{\rm eq}(x,y) \geq \varepsilon$. But $x \sim_{\rm trp} y$ by construction, contradicting that $X_{\rm eq}$ has only trivial regionally proximal pairs. □

[7]see [**369**, Theorem 2.44] for a proof of the non-invertible case, which is not constructive if it comes to the factor map

[8]See e.g. [**312, 421, 481**] for further information

2.3.1. Mean Equicontinuity. Instead of assuming that nearby points always remain close under iteration, mean equicontinuity stipulates that iterates of nearby points remain close on average. This notion was first used by Fomin [**247**] under the name of mean Lyapunov stability[9].

Definition 2.42. A dynamical system (X, T, d) on a metric space is called **mean equicontinuous** if for every $\varepsilon > 0$ there is a $\delta > 0$ such that $d(x, y) < \delta$ implies

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} d(T^i x, T^i y) < \varepsilon.$$

Mean equicontinuity is more versatile than its strict version. Clearly, circle rotations $R_\alpha : \mathbb{S}^1 \to \mathbb{S}^1$ with angle $\alpha \notin \mathbb{Q}$ are isometries and therefore equicontinuous . Their symbolic versions, i.e., Sturmian shifts, see Section 4.3, are expansive and therefore not equicontinuous. Indeed, equip $\mathbb{S}^1$ with an orientation and a partition $\{[0, \alpha)\,, [\alpha, 1)\}$, with symbols 1 and 0 respectively, as is done in Example 1.33. If $x < y \in \mathbb{S}^1$ are very close together, then there are still iterates $n \in \mathbb{N}$ such that $R^n_\alpha(x) < 0 < R^n_\alpha(y)$, so the symbolic distance $d_\sigma(\sigma^n \circ \boldsymbol{i}(x), \sigma^n \circ \boldsymbol{i}(y)) = 1$. However, since this happens less frequently as the distance $|x - y|$ becomes smaller, mean equicontinuity of a Sturmian shift is still achieved.

Another variation of equicontinuity, which is a priori weaker than mean equicontinuity, is **Weyl mean equicontinuity**: for every $\varepsilon > 0$ there is a $\delta > 0$ such that $d(x, y) < \delta$ implies

$$\limsup_{n-m\to\infty} \frac{1}{n-m} \sum_{i=m}^{n-1} d(T^i x, T^i y) < \varepsilon.$$

However, it was proved in [**207**] for minimal dynamical systems (and [**260**, **449**] in more generality) that (X, T) is mean equicontinuous if and only if for every $\varepsilon > 0$ there is a $\delta > 0$ and $N \in \mathbb{N}$ such that $d(x, y) < \delta$ implies

$$\frac{1}{n-m} \sum_{i=m}^{n-1} d(T^i x, T^i y) < \varepsilon \quad \text{for all } m \text{ and } n \geq m + N.$$

Some of the stronger results on mean equicontinuity rely on invariant measures, and therefore don't quite fit in this section on topological dynamics. We present some of this nonetheless, and refer to Chapter 6 for the relevant details. Given a T-invariant Borel probability measure μ, we call (X, T) **μ-mean equicontinuous** if for every $\eta > 0$, there is a set $Y \subset X$ of measure $\mu(Y) > 1 - \eta$ such that T is mean equicontinuous on Y.

[9]This was defined as for every $\varepsilon > 0$ there is $\delta > 0$ such that $d(x, y) < \delta$ implies $d(T^n(x), T^n(y)) < \varepsilon$ for all $n \in \mathbb{N}$ except for a set of density zero. This is equivalent to Definition 2.42 by Lemma 8.52.

As shown in [**204, 435**], if (X, T) is an almost one-to-one extension of a minimal equicontinuous dynamical system (Y, S), then (Y, S) is the maximal equicontinuous factor of (X, T).

It follows from Theorem 6.22 (or more precisely the remarks that follow it) that transitive mean equicontinuous dynamical systems are uniquely ergodic. Thus the following characterization of mean equicontinuity, due to [**207**] for minimal dynamical systems and to [**260**] in general, makes sense:

Theorem 2.43. *A continuous dynamical system (X, T) is mean equicontinuous if and only if its semi-conjugacy to its maximal equicontinuous factor is at the same time a measure-theoretic isomorphism between the unique invariant probability measures of (X, T) and its maximal equicontinuous factor*[10]*.*

Let us show that the symbolic version of an equicontinuous homeomorphism with a reasonable partition is mean equicontinuous.

Theorem 2.44. *Let (X, T) be an equicontinuous homeomorphism on a compact metric space (X, d) with T-invariant measure*[11] *μ. Let $\mathcal{P} = \{P_0, \dots, P_{r-1}\}$ be a finite partition such that*

(1) $\mathcal{P}$ is generating (c.f. Theorem 6.47), i.e., for every $x \neq x' \in X$ there is $n \in \mathbb{Z}$ such that $T^n(x)$ and $T^n(x')$ lie in different partition elements;

(2) $\lim_{\varepsilon \to 0} \mu(U_\varepsilon) = 0$ where U_ε is the ε-neighborhood of $\partial\mathcal{P} = \{x \in X : x \in \overline{P}_i \cap \overline{P}_j$ for some $0 \leq i < j < r\}$.

Let (Y, σ) be the symbolic system associated to $(X, T, \mathcal{P})$, i.e., the smallest subshift such that the itinerary $\boldsymbol{i}(x) \in Y$ for every $x \in X$. Then (Y, σ) is mean equicontinuous.

Proof. Choose $N \in \mathbb{N}$ arbitrary and $0 < \varepsilon' < 2^{-N}/(2N+1)$. Choose $\varepsilon > 0$ so small that $\mu(U_\varepsilon) < \varepsilon'$. By equicontinuity of (X, T) there is $\delta > 0$ such that $d(T^n(x), T^n(x')) < \varepsilon$ for all $n \in \mathbb{Z}$ whenever $d(x, x') < \delta$. Next take $M \in \mathbb{N}$ so large that the diameter $\operatorname{diam}(\boldsymbol{i}^{-1}([e_{-M} \cdots e_M])) < \delta$ for every two-sided $(2M+1)$-cylinder $[e_{-M} \cdots e_M]$.

Now take $y, y' \in Y$ such that $d_\sigma(y, y') \leq 2^{-M}$, where d_σ is the symbolic metric, i.e., y, y' are in the same two-sided $(2M+1)$-cylinder. The sequences y, y' may not be well-defined itineraries of points in X, but this is remedied by assuming that points $x \in X$ such that $T^n(x) \in \partial P$ get multiple itineraries, according to which $\overline{P}_i$ contains $T^n(x)$. In this sense there are x, x' such that at least one of their multiple itineraries equal y and y' respectively. In particular, $d(x, x') < \delta$ and therefore $d(T^n(x), T^n(x')) < \varepsilon$ for all $n \in \mathbb{Z}$. The

[10] In this case, (X, T) is called a topo-isomorphic extension of its MEF.

[11] If (X, T) is minimal, then μ is unique.

points $T^n(x)$ and $T^n(x')$ can only lie in different partition elements if they both lie in U_ε. Unless $T^n(x), T^n(x') \in \cup_{j=-N}^N T^j(U_\varepsilon)$, their itineraries satisfy $d_\sigma(\boldsymbol{i}(T^n(x)), \boldsymbol{i}(T^n(x'))) \leq 2^{-N}$. But the measure $\mu(\cup_{j=-N}^N T^j(U_\varepsilon)) \leq (2N+1)\varepsilon'$ and by Oxtoby's Ergodic Theorem 6.20, x and x' visit $\cup_{j=-N}^N T^j(U_\varepsilon)$ with frequency $\leq (2N+1)\varepsilon'$. Therefore

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} d_\sigma(\sigma^j(\boldsymbol{i}(x)), \sigma^j(\boldsymbol{i}(x'))) \\ &\leq \limsup_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} d_\sigma(\boldsymbol{i}(T^j(x)), (T^j(\boldsymbol{i}(x'))) \\ &\leq (2N+1)\varepsilon' + (1-(2N+1)\varepsilon')2^{-N} \leq 2^{-N+1}.\end{aligned}$$

This proves mean equicontinuity. □

2.4. Topological Entropy

The notion of **topological entropy** was introduced, by Adler, Konheim & McAndrew [**9**] in 1969. Nowadays, the definition due to the American mathematician Rufus Bowen [**98**] and, independently, his Russian colleague Efim Dinaburg [**198**] is most often[12] used.

Entropy is a measure of disorder of the dynamical system, and one popular definition of chaos is that the topological entropy is positive.

Let (X, T) be continuous dynamical dynamical system on a compact metric space (X, d). If my eyesight is not so good, I cannot distinguish two points $x, y \in X$ if $d(x, y) \leq \varepsilon$. I may still be able to distinguish their orbits, if $d(T^k x, T^k y) > \varepsilon$ for some $k \geq 0$. Hence, if I'm willing to wait up to $n-1$ iterations, I can distinguish x and y if

$$d_n(x, y) := \max\{d(T^k x, T^k y) : 0 \leq k < n\} > \varepsilon.$$

If this holds, then x and y are said to be **(n, ε)-separated**. Among all the subsets of X of which all elements are mutually (n, ε)-separated, choose one, say $E_n(\varepsilon)$, of maximal cardinality. Then $s_n(\varepsilon) := \#E_n(\varepsilon)$ is the maximal number of n-orbits I can distinguish with my ε-poor eyesight.

Remark 2.45. Compactness of X together with continuity of T ensures that $s_n(eps) < \infty$. However, also for discontinuous maps, such as β-transformations, it can be proven that $s_n(eps) < \infty$ for all $\varepsilon > 0$ and $n \in \mathbb{N}$. Consequently, this approach to topological entropy usually also works for discontinuous functions.

[12] Note, however, that the Adler, Konheim & McAndrew definition requires only a topology, whereas the Bowen-Dinaburg definition is metric.

The **topological entropy** is defined as the limit (as $\varepsilon \to 0$) of the exponential growth-rate of $s_n(\varepsilon)$:

$$h_{top}(T) = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log s_n(\varepsilon). \tag{2.4}$$

Note that $s_n(\varepsilon_1) \geq s_n(\varepsilon_2)$ if $\varepsilon_1 \leq \varepsilon_2$, so $\limsup_n \frac{1}{n} \log s_n(\varepsilon)$ is a decreasing function in ε, and the limit as $\varepsilon \to 0$ indeed exists.

Instead of (n,ε)-separated sets, we can also work with (n,ε)**-spanning** sets, that is, sets that contain, for every $x \in X$, a point y such that $d_n(x,y) \leq \varepsilon$. Let $r_n(\varepsilon)$ denote the minimal cardinality among all (n,ε)-spanning sets. Due to its maximality, $E_n(\varepsilon)$ is always (n,ε)-spanning, and no proper subset of $E_n(\varepsilon)$ is (n,ε)-spanning. Each $y \in E_n(\varepsilon)$ must have a point of an $(n,\varepsilon/2)$-spanning set within an $\varepsilon/2$-ball (in d_n-metric) around it, and by the triangle inequality, this $\varepsilon/2$-ball is disjoint from the $\varepsilon/2$-balls centered around all other points in $E_n(\varepsilon)$. Therefore,

$$r_n(\varepsilon) \leq s_n(\varepsilon) \leq r_n(\varepsilon/2). \tag{2.5}$$

Thus we can equally well define

$$h_{top}(T) = \lim_{\varepsilon\to 0} \limsup_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon). \tag{2.6}$$

Example 2.46. Let (X,σ) be the full shift on N symbols. Let $\varepsilon > 0$ be arbitrary, and take m minimal such that $2^{-m} < \varepsilon$. If we select a point from each $n+m$-cylinder, this gives an (n,ε)-spanning set, whereas selecting one point from each n-cylinder gives an (n,ε)-separated set. Therefore

$$\begin{aligned} \log N = \limsup_{n\to\infty} \frac{1}{n} \log N^n &\leq \limsup_{n\to\infty} \frac{1}{n} \log s_n(\varepsilon) \leq h_{top}(\sigma) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log r_n(\varepsilon) \leq \limsup_{n\to\infty} \frac{1}{n} \log N^{n+m} \\ &= \log N. \end{aligned}$$

Exercise 2.47. Show that for subshifts the definition of (1.2) coincides with (n,ε)-definition in this section.

Example 2.48. Consider the β-transformation $T_\beta : [0,1) \to [0,1)$, $x \mapsto \beta x \bmod 1$ for some $\beta > 1$. Take $\varepsilon < \frac{1}{2\beta^2}$, and $G_n = \{\frac{k}{\beta^n} \ : \ 0 \leq k < \beta^n\}$. Then G_n is (n,ε)-separating, so $s_n(\varepsilon) \geq \beta^n$. On the other hand, $G'_n = \{\frac{2k\varepsilon}{\beta^n} \ : \ 0 \leq k < \frac{\beta^n}{2\varepsilon}\}$ is (n,ε)-spanning, so $r_n(\varepsilon) \leq \frac{\beta^n}{2\varepsilon}$. Therefore

$$\log \beta = \limsup_{n\to\infty} \frac{1}{n} \log \beta^n \leq h_{top}(T_\beta) \leq \limsup_{n\to\infty} \frac{1}{n} \log \frac{\beta^n}{2\varepsilon} = \log \beta.$$

Circle rotations, or in general isometries, have zero topological entropy. Indeed, if $E(\varepsilon)$ is an ε-separated set (or ε-spanning set), it will also be (n,ε)-separated (or (n,ε)-spanning) for every $n \geq 1$. Hence $s_n(\varepsilon)$ and $r_n(\varepsilon)$ are independent of n, and their exponential growth-rates are equal to zero. In more generality:

Proposition 2.49. Every equicontinuous transformation (X, T) on acompact metric space (X, d) has zero entropy.

Proof. Let $\varepsilon > 0$ be arbitrary and choose $\delta > 0$ as in the definition of equicontinuity. Then $\text{diam}(T^n(B_\delta(x)) \leq 2\varepsilon$ for all $x \in X$ and $n \geq 0$ (or $n \in Z$ if T is invertible). Take $M = \lceil \text{diam}(X)/\delta \rceil$. Hence, a single cover of X by M δ-balls constitutes a cover of (n, ε)-balls for all n. Therefore $h_{top}(T) \leq \lim_{\varepsilon \to 0} \lim_{n \to \infty} \frac{1}{n} \log M = 0$. □

Corollary 2.50. Given a continuous map $T : X \to X$, $h_{top}(T^k) = kh_{top}(T)$ for all $k \geq 0$, and if T is invertible, then $h_{top}(T^k) = |k|h_{top}(T)$ for all $k \in \mathbb{Z}$.

Proof. For any $k \in \mathbb{N}$, a (kn, ε)-separated set for T is also an (n, ε)-separated set for T^k. Therefore

$$h_{top}(T^k) = \lim_{n \to \infty} \frac{1}{n} \log s_n(\varepsilon, T^k) = k \lim_{n \to \infty} \frac{1}{kn} \log s_n(\varepsilon, T) = kh_{top}(T).$$

Clearly the identity T^0 has zero entropy. If T is invertible, and $E_n(\varepsilon)$ is an (n, ε)-separated set, then $T^{n-1}(E_n(\varepsilon))$ is an $(n\varepsilon)$-separated set for T^{-1}. Therefore $h_{top}(T^{-1}) = h_{top}(T)$. Combined with the first part, it follows that $h_{top}(T^k) = |k|h_{top}(T)$ for all $k \in \mathbb{Z}$. □

Corollary 2.51. If (Y, S) is a continuous factor of (X, T) (where (X, d) is a compact metric space), then $h_{top}(S) \leq h_{top}(T)$. In particular, conjugate dynamical systems on compact metric spaces have the same topological entropy.

Proof. Let $\pi : X \to Y$ be a continuous factor map. Since X is compact, π is uniformly continuous, so for $\varepsilon > 0$, we can find $\delta > 0$ such that $d(x, y) < \delta$ implies $d(\pi(x), \pi(y)) < \varepsilon$. Therefore, if $E_n(\delta)$ is an (n, δ)-spanning set for T, then $\pi(E_n(\delta))$ is an (n, ε)-spanning set for S (but possibly not a minimal (n, ε)-spanning set, even if $E_n(\delta)$ is minimal). It follows that $r_n(\delta, T) \geq r_n(\varepsilon, S)$, and hence $h_{top}(T) \geq h_{top}(S)$. □

Proposition 2.52. The entropy of a dynamical system (X, T) restricted to the nonwandering set $\Omega(T)$ satisfied $h_{top}(T) = h_{top}(T|_{\Omega(T)})$.

Since T-invariant measures have to be supported on the nonwandering set, this proposition follows from the Variational Principle (Theorem 6.62).

Example 2.53. The nonwandering set $\Omega(\sigma)$ of the subshift

$$X = \overline{\{0^{n_1}1^{n_2}0^{n_3}1^{n_4}\cdots : 0 \leq n_1 \leq \max\{n_1, 1\} \leq n_2 \leq n_3 \leq n_4 \leq \dots\}}$$

consists of periodic orbits $0^k1^k0^k1^k\cdots$ or $1^k0^k1^k0^k\cdots$, i.e., with period $2k$. Therefore the number of $2n$-periodic points (not necessarily prime period $2n$) equals twice the number of divisors of n, and hence is $\leq 2n$. In view of Proposition 2.52, we have $h_{top}(\sigma) = 0$.

2.4.1. Amorphic Complexity. If the cardinalities of (n, ε)-separated and of (n, ε)-spanning sets increase subexponentially, then one could compute polynomial growth-rates instead. This is called **power entropy**:

$$h_{pow}(T) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{\log s_n(\varepsilon)}{\log n}, \tag{2.7}$$

see [**296**]. However, in practice this isn't a very powerful tool to distinguish between dynamical systems, because for instance, all dynamical systems with linear word-complexity have $h_{pow}(T) = 1$. A recent approach [**261**], which turns out to distinguish between many zero-entropy systems (even of linear complexity and between some semi-conjugate dynamical systems), is amorphic complexity[13]. It is based on the average time v that orbits are δ apart. Given a dynamical system (X, T) on a metric space (X, d), two points $x, y \in X$ are (δ, v)-separated for some $\delta > 0$ if

$$\limsup_{n \to \infty} \frac{1}{n} \#\{0 \leq j < n : d(T^j(x), T^j(y)) \geq \delta\} \geq v.$$

A set $S \subset X$ is (δ, v)-separated if every $x \neq y \in S$ are (δ, v)-separated. Let $\mathrm{Sep}(\delta, v)$ denote the maximal cardinality of the (δ, v)-separated sets. We say that (X, T) has **finite separation numbers** if $\mathrm{Sep}(\delta, v) < \infty$ for all $\delta, v > 0$. If $\mathrm{Sep}(\delta, v) = \infty$ for some $\delta, v > 0$, then (X, T) has **infinite separation numbers**, and in this case the amorphic complexity defined below is infinite, hence not so useful. This occurs, for instance, in the following cases, see [**261**, Theorem 1.1]:

Theorem 2.54. *Let (X, T) be a continuous dynamical system on a compact metric space (X, d). If $h_{top}(T) > 0$ or T is weakly mixing w.r.t. some non-atomic invariant probability measure (see Definition 6.82), then T has infinite separation numbers.*

Hence we are only interested in dynamical systems with separation numbers that are finite, but potentially unbounded in v.

Definition 2.55. Assume that (X, T) has finite separation numbers. The **upper/lower amorphic complexity** is the polynomial growth-rate of the

[13] This notion was first used in the context of aperiodic tilings that approximate "amorphous" material. The name was coined for this reason.

separation numbers as function of v tending to zero:

$$\begin{cases} \overline{\mathrm{ac}}(T) = \sup_{\delta>0} \limsup_{v\to 0} \frac{\log \mathrm{Sep}(\delta,v)}{-\log v}, \\ \underline{\mathrm{ac}}(T) = \sup_{\delta>0} \liminf_{v\to 0} \frac{\log \mathrm{Sep}(\delta,v)}{-\log v}. \end{cases} \tag{2.8}$$

If these quantities are the same, then $\mathrm{ac}(T) = \sup_{\delta>0} \lim_{v\to 0} \frac{\log \mathrm{Sep}(\delta,v)}{-\log v}$ is the **amorphic complexity** of T.

Remark 2.56. Amorphic complexity can also be defined by spanning sets [**261**, Section 3.2]. A set $S \subset X$ is (δ, v)-spanning if for every $y \in X$ there is an $x \in S$ such that

$$\limsup_{n\to\infty} \frac{1}{n} \#\{0 \le j < n : d(T^j(x), T^j(y)) \ge \delta\} < v.$$

Letting $\mathrm{Span}(\delta, v)$ denote the minimal cardinality of the (δ, v)-spanning sets, (2.8) holds with $\mathrm{Sep}(\delta, v)$ replaced by $\mathrm{Span}(\delta, v)$.

If T is an isometry, then the frequency of two points $x, y \in X$ being $\ge \delta$ apart is 0 or 1, depending on whether $d(x, y) < \delta$ or $\ge \delta$. Therefore $\mathrm{Sep}(\delta, v)$ is independent of v, so $\mathrm{ac}(T) = 0$. More generally:

Proposition 2.57. If (X, T) is equicontinuous, then the amorphic complexity $\mathrm{ac}(T) = 0$.

Proof. Let $\varepsilon > 0$ be arbitrary. By equicontinuity and the compactness of X, we can take $\delta > 0$ such that $T^n(B_\delta(x)) \subset B_{\varepsilon/2}(T^n(x))$ for all $x \in X$ and $n \in \mathbb{N}$ or $\mathbb{Z}$. Thus two points in $B_\delta(x)$ are never (ε, v)-separated for any $v \in (0, 1]$. Let $N(\delta)$ be the number of such δ-balls that can be packed in X, so that no such ball contains the center of another. Then $\frac{\log \mathrm{Sep}(\varepsilon,v)}{-\log v} \le \frac{\log N(\delta)}{-\log v} \to 0$ as $v \to 0$. Therefore $\mathrm{ac}(T) = 0$. □

Further properties concern iterates and factors, see [**261**, Proposition 1.3].

Lemma 2.58. *Let (X, T) and (Y, S) be two dynamical systems on compact metric spaces.*

- *If (Y, S) is a topological factor of (X, T), then* $\mathrm{ac}(S) \le \mathrm{ac}(T)$*. In particular, amorphic complexity is preserved under conjugacy.*
- $\mathrm{ac}(T^n) = \mathrm{ac}(T)$ *for every $n \in \mathbb{N}$.*
- $\mathrm{ac}(S \times T) = \mathrm{ac}(S) + \mathrm{ac}(T)$.

In later sections, we compute the amorphic complexity of some particular dynamical systems, such as Sturmian shifts, see Section 4.3.1, and Toeplitz shifts, see Section 4.5.

2.5. Mathematical Chaos

Mathematical **chaos** doesn't have a single definition, but the basic idea it tries to capture is that forward orbits are unpredictable. The computation of orbits in any (physical) dynamical systems inherently brings errors: measurement errors, round-off errors, error in the mathematical model. Unpredictability means that initial errors blow up over time (sometimes exponentially fast, as is the case with subshifts). Therefore distal dynamical systems on compact spaces (in particular isometries) are not chaotic in any common definition. On the other hand, expansivity is in general too strong a property to require for chaos. For instance, a **tent map**

$$T_s : [0,1] \to [0,1], \quad T_s : x \mapsto \min\{sx, s(1-x)\}$$

is chaotic if the slope $s \in (\sqrt{2}, 2]$, but not expansive. Indeed, $x = \frac{1+\varepsilon}{2}$ and $y = \frac{1-\varepsilon}{2}$ are ε apart, but $T_s^n(x) = T_s^n(y)$ for all $n \geq 1$. A weaker, more appropriate, definition in this context is the following:

Definition 2.59. A dynamical system (X, T) on a metric space (X, d) has **sensitive dependence on initial conditions** if there is $\delta > 0$ such that for all $\varepsilon > 0$ and $x \in X$, there is $y \in B_\varepsilon(x)$ and $n \geq 0$ such that $d(T^n(x), T^n(y)) > \delta$.

This leads to one of the most common definitions of chaos [**192**]:

Definition 2.60. A dynamical system (X, T) on a metric space (X, d) is **chaotic in the sense of Devaney** if

(1) X has a dense set of periodic orbits;

(2) T has a dense orbit;

(3) (X, T) has sensitive dependence on initial conditions.

As was soon realized by Banks et al. [**44**], unless X is a single periodic orbit, 3. follows automatically from 1. and 2. See also Silverman's study [**494**] on chaos and topological transitivity.

Proposition 2.61. Let (X, T) be a continuous dynamical system on an infinite metric space (X, d). If T has a dense set of periodic orbits as well as a dense orbit, then T has sensitive dependence on initial conditions.

Proof. Since X is infinite and has a dense orbit, no periodic point is isolated, and are at least two periodic orbits, say $\mathrm{orb}(p)$ and $\mathrm{orb}(q)$. Let $\delta := \min\{d(x,y) : x, y \in \mathrm{orb}(p) \cup \mathrm{orb}(q), x \neq y\}/6 > 0$. Take $x \in X$ and $\varepsilon > 0$ arbitrary. Then $B_\varepsilon(x)$ contains a periodic point $r \notin \mathrm{orb}(p) \cup \mathrm{orb}(q)$. If there is $n \geq 0$ such that $d(T^n(x), T^n(r)) > \delta$, then sensitive dependence is established at x. Therefore assume that $d(T^n(x), T^n(r)) \leq \delta$ for all $n \geq 0$.

Since there is a dense orbit, we can find $y \in B_\varepsilon(x)$ such that $p, q \in \overline{\mathrm{orb}(y)} = X$. Take $j, k \in \mathbb{N}$ such that

$$d(T^{j+i}(y), T^i(p)) < \delta \quad \text{and} \quad d(T^{k+i'}(y), T^{i'}(q)) < \delta$$

for all $0 \le i, i' \le \mathrm{per}(r)$. We can choose $0 \le i, i' \le \mathrm{per}(r)$ such that

$$d(T^i(p), r) \le d(T^i(p), T^{j+i}(y)) + d(T^{j+i}(y), T^{j+i}(x)) + d(T^{j+i}(x), r) \le 3\delta$$

and

$$d(T^{i'}(q), r) \le d(T^{i'}(q), T^{k+i'}(y)) + d(T^{k+i'}(y), T^{k+i'}(y)) + d(T^{k+i'}(x), r) \le 3\delta.$$

But then $d(T^i(p), T^{i'}(q)) \le 6\delta$, contradicting the choice of δ. This proves the result. □

The requirement of a dense set of periodic orbits in Devaney chaos is restrictive, because it precludes minimal systems to be chaotic. The following notion doesn't have this drawback.

Definition 2.62. A dynamical system (X, T) on a metric space (X, d) is **chaotic in the sense of Auslander-Yorke** if

(1) T has a dense orbit;

(2) (X, T) has sensitive dependence on initial conditions.

The following results is known as the Auslander-Yorke dichotomy [**38**]:

Theorem 2.63. *Every minimal dynamical system (X, T) is either equicontinuous or has sensitive dependence on initial conditions.*

In fact, there is a version of the Auslander-Yorke dichotomy, see [**14**, **410**, **533**, **543**], saying that a transitive dynamical system either has sensitive dependence on initial conditions (see Definition 2.59), or is uniformly rigid. This implies in particular that for minimal dynamical systems, equicontinuity is equivalent to uniform rigidity.

Remark 2.64. There is also an analogue for mean equicontinuity, see [**383**] and also [**266**], saying that every minimal dynamical system is either mean equicontinuous or **mean sensitive**, which means that there is a $\delta > 0$ such that for every $x \in X$ and neighborhood $U \ni x$, there is $y \in U$ such that $\limsup_n \frac{1}{n} \sum_{i=0}^{n-1} d(T^i x, T^i y) > \delta$. A measure-theoretic version of the dichotomy is due to [**265**], which states that given an ergodic T-invariant Borel measure μ, (X, T) is either μ-mean equicontinuous or **μ-mean sensitive**, i.e., mean sensitive with "neighborhood U" replaced by "Borel set $U \ni x$ with $\mu(U) > 0$".

The paper of Li & Yorke [**375**] from 1973 might be called a popular (partial) rediscovery of Sharkovskiy's theorem [**482**] from 1964[14], but it started the following characterization as well.

Definition 2.65. Let (X, T) be a dynamical system on a metric space (X, d). A pair of points $x, y \in X$ is called a **Li-Yorke pair** if

$$\liminf_{n\to\infty} d(T^n(x), T^n(y)) = 0 \quad \text{and} \quad \limsup_{n\to\infty} d(T^n(x), T^n(y)) > 0.$$

A set $S \subset X$ is called **scrambled** if (x, y) is a Li-Yorke pair for every two distinct $x, y \in S$. The dynamical system is **chaotic in the sense of Li and Yorke** if there is an uncountable scrambled set.

Huang & Ye [**315**, Theorem 4.1] proved that if a continuous dynamical system is transitive and **properly** contains a periodic orbit, then it is chaotic in the sense of Li-Yorke. In particular, Devaney chaos implies Li-Yorke chaos.

Example 2.66. Let us construct an uncountable scrambled set in the full shift space $X = \{0, 1\}^{\mathbb{N}_0}$. First define an equivalence relation $\sim$ by setting $x \sim y$ if there is $n_0 \in \mathbb{N}$ such that either $x_n = y_n$ for all $n \geq n_0$ or $x_n \neq y_n$ for all $n \geq n_0$. That is, x and y have either the same or opposite tails. Each equivalence class is countable, because for each fixed n_0 there are finitely many equivalent points with the same n_0. Since X is uncountable, there are uncountably many equivalence classes.

Next, using the axiom of choice, construct a set $Y \subset X$ that contains exactly one point in each equivalence class.

Now define an injection $\pi : X \to X$ by $\pi(x)_j = x_n$ for each $2^n - 1 \leq j < 2^{n+1} - 1$. Then $S = \pi(Y)$ is uncountable and scrambled. Indeed, for every $x \neq y \in Y$, there are infinitely many n such that $x_n = y_n$ and then $d(\sigma^{2^n-1} \circ \pi(x), \sigma^{2^n-1} \circ \pi(y)) \leq 2^{-n}$. Also there are infinitely many n such that $x_n \neq y_n$ and then $d(\sigma^{2^n-1} \circ \pi(x), \sigma^{2^n-1} \circ \pi(y)) \geq 1 - 2^{-n}$.

Similarly, all non-trivial subshifts of finite type (SFTs) are Li-Yorke chaotic, but Sturmian subshifts (or more generally distal maps) are not Li-Yorke chaotic ($\liminf_n d(T^n(x), T^n(y)) > 0$ for distinct $x \neq y \in X$).

[14] Sharkovskiy's Theorem states that if a continuous map of the real line has a periodic point of period n, it also has a periodic point of period m for every $m \prec n$ in the Sharkovskiy order $1 \prec 2 \prec 4 \prec 8 \prec \cdots \prec 4 \cdot 7 \prec 4 \cdot 5 \prec 4 \cdot 3 \cdots \prec 2 \cdot 7 \prec 2 \cdot 5 \prec 2 \cdot 3 \cdots \prec 7 \prec 5 \prec 3$. Sharkovskiy related during the 2018 IWCTA: International Workshop and Conference on Topology & Applications (Kochi, India) in honor of his 1000-th moon that the printer of his original publication didn't have the sign $\prec$ at his disposal, and therefore he suggested to use the letter Y turned side-ways. The publisher followed this suggestion, but turned the Y in the different direction as Sharkovskiy had intended, and therefore the Sharkovskiy order was first printed as $3 \mathrel{\prec\!\!\!-} 5 \mathrel{\prec\!\!\!-} 7 \mathrel{\prec\!\!\!-} \dots 6 \mathrel{\prec\!\!\!-} 10 \mathrel{\prec\!\!\!-} 14 \mathrel{\prec\!\!\!-} \dots 12 \mathrel{\prec\!\!\!-} 20 \mathrel{\prec\!\!\!-} 28 \mathrel{\prec\!\!\!-} \dots\dots \mathrel{\prec\!\!\!-} 4 \mathrel{\prec\!\!\!-} 2 \mathrel{\prec\!\!\!-} 1$ in [**482**]. Štefan [**507**] in his 1977 proof used $3 \vdash 5 \vdash 7 \vdash \dots$ and the English translation of Sharkovskiy's proof [**483**] by Tolosa used $3 \triangleright 5 \triangleright 7 \triangleright \dots$

Exercise 2.67. Let $X = \mathcal{A}^{\mathbb{N}_0}$ be the full shift space for some alphabet $\mathcal{A}$ containing a. Define $\pi : X \to X$ by

$$\pi(x)_k = \begin{cases} x_{k-n^2} & n^2 \leq k \leq n^2 + n; \\ a & n^2 + n < k \leq n^2 + 2n. \end{cases}$$

Show that $\pi(X)$ is a scrambled set.

An important, long conjectured, result ties Li-Yorke chaos to topological entropy.

Theorem 2.68. *Every dynamical system of positive entropy is Li-Yorke chaotic.*

This is the main result of [**79**], see also [**467**, Chapter 5] and [**206**]. The converse is, however, not true. There exist examples of continuous (so-called 2^∞) interval maps, which have periodic points of period 2^n for each $n \in \mathbb{N}$ and no periodic points with other periods, which have (therefore) zero topological entropy, but which still are Li-Yorke chaotic, see [**500**, **546**]. Example 2.53 gives a subshift which has zero entropy but is Li-Yorke chaotic.

Theorem 2.69. *Let $X = \{1, \dots, d\}^{\mathbb{N}}$. For every probability vector $\boldsymbol{p} = (p_1, \dots, p_d)$, every scrambled set has zero $\boldsymbol{p}$-Bernoulli measure.*

Proof. Let $(X, \mathcal{B}, \mu_{\boldsymbol{p}}, \sigma)$ be the $\boldsymbol{p}$-Bernoulli shift and assume by contradiction that $S \subset X$ is a scrambled set with $\mu_{\boldsymbol{p}}(S) > 0$. Take two distinct Lebesgue density points a and b of S', and for any n, let $Z_n(a)$ and $Z_n(b)$ be the corresponding n-cylinders of a and b respectively. Because a and b are density points, the Lebesgue fractions of $\mu_{\boldsymbol{p}}(\sigma^n(S \cap Z_n(a)))$ and $\mu_{\boldsymbol{p}}(\sigma^n(S \cap Z_n(b)))$ tend to 1 as $n \to \infty$. That means that there are distinct $x, y \in S$ and some $n \in \mathbb{N}$ such that $\sigma^n(x) = \sigma^n(y)$. But then (x, y) is not a Li-Yorke pair. This contradiction shows that $\mu_{\boldsymbol{p}}(S) = 0$. $\square$

2.6. Transitivity and Topological Mixing

Transitivity prevents that the phase space consist of multiple pieces that don't communicate with each other. Topological mixing prevents that they communicate with each other only at a periodic sequence of iterates. There are several related concepts in addition to (totally) transitive from Definition 2.12:

Definition 2.70. A dynamical systems (X, T) on a topological space is called **topologically mixing** if for every two open set U, V there is $N \geq 0$ such that $U \cap T^{-n}(V) \neq \varnothing$ for all $n \geq N$.

Topologically mixing dynamical systems on metric space are sensitive to initial conditions (provided X consists of at least two points), and therefore equicontinuous dynamical systems cannot be topologically mixing. In particular, since topological mixing is inherited by factors, the maximal equicontinuous factor of a topologically mixing dynamical system is trivial.

Definition 2.71. A dynamical system (X,T) on a topological space is called **topologically exact** (also called **locally eventually onto** or **leo** for short) if for every open set U there is $N \geq 0$ such that $T^N(U) = X$.

Invertible dynamical systems (other than the identity on a singleton) are never topologically exact, and neither are nontrivial dynamical systems with zero entropy.

Lemma 2.72. *If a dynamical system (X,T) on a non-trivial metric space (X,d) is topologically exact, then $h_{top}(T) > 0$.*

Proof. Take $x_0 \neq x_1 \in X$, and choose $0 < \varepsilon < d(x_0,x_1)/3$. Let U_0 and U_1 be the ε-neighborhoods of x_0 and x_1 respectively. By topological exactness, there is $N \in \mathbb{N}$ such that $T^N(U_0) = X = T^N(U_1)$. Hence, for an arbitrary $n \in \mathbb{N}$ and every $w = w_0w_1\cdots w_{n-1} \in \{0,1\}^n$, there is $x_w \in X$ such that $T^{kN}(x_w) \in U_{w_k}$ for all $0 \leq k < n$. If $w \neq w' \in \{0,1\}^n$, then the nN-distance $d_{nN}(x_w, x_{w'}) > \varepsilon$. Hence, every (nN,ε)-spanning set must contain at least 2^n elements and $h_{top}(T) \geq \frac{1}{N}\log 2 > 0$. □

Theorem 2.73. *If $T : [0,1] \to [0,1]$ is a continuous transitive interval map, then $h_{top}(T) \geq \frac{1}{2}\log 2$. If in addition T is topologically mixing, then $h_{top}(T) > \frac{1}{2}\log 2$.*

This result is due to Blokh, see [**86**, **88**]. A compact exposition of this and related results can be found in [**467**, Proposition 4.70].

Definition 2.74. A dynamical system (X,T) on a topological space is called **weakly topologically mixing** if for every four non-empty open sets U_1, U_2, V_1, V_2, there is n such that $U_1 \cap T^{-n}(V_1) \neq \varnothing$ and $U_2 \cap T^{-n}(V_2) \neq \varnothing$, or equivalently, the product system $T \times T$ on $X \times X$ is transitive.

When presenting these notions, we consistently write the adjective "topological" because there are also measure-theoretic versions of exact, mixing and weak mixing. These are discussed in Section 6.7. Some specific differences exist, for instance, ther eis no topological analog of Theorem 6.85.

From the definition it is clear that topological weak mixing implies that the product system $(X^2, T \times T)$ is transitive. In fact, Furstenberg [**263**] showed that this holds for every N-fold Cartesian product $(X^N, T \times \cdots \times T)$. An important result on topological weak mixing is the following multiple recurrence (a dynamical version of Van der Waerden's Theorem) due to Furstenberg & Weiss [**264**]: if (X,T) is minimal then for every open set $U \subset X$ and $m \in \mathbb{N}$, there is $n \in \mathbb{N}$ such that $U \times T^n(U) \times T^{2n}(U) \times \cdots \times T^{mn}(U) \neq \varnothing$. Glasner [**272**] extended this to multiple transitivity: if (X,T) is minimal and topologically weak mixing, then for x in a residual subset of X, the m-tuple $(x,\dots,x)$ has a dense orbit under $T \times T^2 \times \cdots \times T^m$. Further results can be found in e.g. [**135**, **411**].

The following hierarchy (which also holds for the measure-theoretic analog) will not come as a surprise:

Theorem 2.75. *The following implications hold:*

top. exact	$\Rightarrow$	*topologically mixing*	$\Rightarrow$	*topologically weak mixing*	$\Rightarrow$	*totally transitive*	$\Rightarrow$	*topologically transitive*

The reverse implications are in general false.

Counter-examples to the reverse implications can be found among subshifts:

full shift	Petersen's shift	Chacon substitution shift	Fibonacci substitution shift	Thue-Morse shift

where the Fibonacci, Chacon and Thue-Morse substitution shifts are defined in Examples 1.3, 1.27 and 1.6, respectively. Petersen's shift [**440**] is an example of a zero entropy subshift that is topologically mixing. Lemma 2.72 shows that it cannot be topologically exact.

Remark 2.76. Although none of the reverse implications in Theorem 2.75 holds in all generality, for many subshifts, some of these notions are equivalent. For instance, sofic shifts and density shifts that are totally transitive are topologically mixing, cf. [**234**] and Theorem 3.61. For coded and synchronized shifts, total transitivity is equivalent to topologically weak mixing, see [**234**, Theorem 1.1].

In terms of the set of **visit times** for sets $U, V \subset X$,

$$\mathcal{N}(U,V) = \{n \in \mathbb{N}_0 \text{ or } \mathbb{Z} : U \cap T^{-n}(V) \neq \varnothing\}. \tag{2.9}$$

the notions in this section can be expressed as follows. For all $U, U', V, V' \subset X$ open and non-empty:

- topologically exact: $\bigcap_{x \in X} \mathcal{N}(U, \{x\})$ is cofinite.
- topologically mixing: $\mathcal{N}(U,V)$ is cofinite.
- topologically weak mixing: $\mathcal{N}(U,V) \cap \mathcal{N}(U',V')$ is infinite.
- topologically transitive: $\mathcal{N}(U,V)$ is non-empty.
- totally transitive: $\forall k\ \mathcal{N}(U,V) \cap k\mathbb{N}$ is non-empty.

2.7. Shadowing and Specification

Definition 2.77. Let (X, T) be a dynamical system on a metric space (X, d). A sequence $(x_n)_{n\in\mathbb{N}_0 \text{or } \mathbb{Z}}$ is called a **δ-pseudo-orbit** if $d(T(x_n), x_{n+1}) < \delta$ for all $n \in \mathbb{N}_0$ or $\mathbb{Z}$.

Given that every floating-point calculation has round-off errors, orbits that a computer calculates numerically are always pseudo-orbits for some small δ. Whether such pseudo-orbits represent an approximations of actual orbits is captured in the following definition.

Definition 2.78. A dynamical system (X, T) on a metric space (X, d) has the **shadowing property** if for every $\varepsilon > 0$ there is $\delta > 0$ such that for every δ-pseudo-orbit (x_n), there is $y \in X$ so that $\text{orb}(y)$ **ε-shadows** (x_n), i.e., $d(x_n, T^n(y)) < \varepsilon$ for all n.

By now, many variations of shadowing have been studied, for example average shadowing (the average error need to be smaller than ε), periodic shadowing (periodic pseudo-orbits are ε-shadowed by actual periodic orbits), limit shadowing (the ε in the shadowing tends to zero as the iterates $|n| \to \infty$). We refer to the monograph by Pilyugin [**447**], although many variations of shadowing are from a later date, cf. [**53**, **274**, **410**].

The seminal result for shadowing is the Anosov Shadowing Lemma [**26**] for hyperbolic sets. Work by Bowen [**95**] showed that hyperbolic dynamical systems, and this includes SFTs, have the shadowing property.

Definition 2.79. Let $f : M \to M$ be a C^1 diffeomorphism of a C^1 Riemannian manifold M. An f-invariant set Λ is called **hyperbolic** if there is a uniformly transversal splitting $T_qM = E^s_q \oplus E^u_q$ of the tangent spaces that is continuous in $q \in \Lambda$, invariant under f, i.e., $Df_q(E^s_q) = E^s_{f(q)}$ and $Df_q(E^u_q) = E^u_{f(q)}$, and tangent vectors in E^s_q resp. E^u_q decrease exponentially fast under forward resp. backward iteration.

If $f : M \to M$ is not invertible, then we need to select inverse branches in order for E^u_q to be well-defined. The manifold contains stable and unstable local manifolds $W^s_{loc}(q)$ and $W^u_{loc}(q)$ of q, tangent to E^s_q and E^u_q respectively, such that

$$d(f^n(q), f^n(x)) \to 0 \text{ exponentially, as } \begin{cases} n \to \infty & \text{if } x \in W^s_{loc}(q), \\ n \to -\infty & \text{if } x \in W^u_{loc}(q). \end{cases}$$

In the symbolic setting, i.e., a subshift (X, σ) takes the place of (M, f), we can define

$$\begin{cases} W^s_{loc}(q) = \{x \in X : x_n = q_n \text{ for all } n \geq 0\}, \\ W^u_{loc}(q) = \{x \in X : x_n = q_n \text{ for all } n \leq 0\}. \end{cases}$$

Theorem 2.80 (Anosov Shadowing Lemma)**.** *Let Λ be a hyperbolic set of a C^1 diffeomorphism $f : M \to M$, and let Λ_ε denote the ε-neighborhood of Λ. Then for every $\varepsilon > 0$ there is $\delta > 0$ such that every δ-pseudo orbit $(x_k) \subset \Lambda_\delta$ (finite, one-sided or two-sided infinite), there is $x \in \Lambda_\varepsilon$ such that $d(f^k(x), x_k) < \varepsilon$ for all k.*

The analogue of this theorem for periodic shadowing is called the Anosov Closing Lemma. See [**336**, Sections 6.4 and 18.1].

One may think that uniform expansion is enough to guarantee shadowing, but it is not as simple as that. For example (see [**167**], and [**112**, Theorem 6.3.5]), tent maps T_s with slope $s \in (1, 2)$ have the shadowing property if and only if the critical point c is recurrent **or** its kneading map is unbounded (in the terminology of Section 3.6.3).

An important variation of shadowing, also introduced by Bowen [**97**], is specification. In this case, no pseudo-orbits are involved, but particular pieces of orbits are to be ε-shadowed for particular intervals of time, allowing gaps in between that are inverse proportional to $\log \varepsilon$.

Definition 2.81. A dynamical system (X, T) on a metric space (X, d) has **specification** for K points if for every $\varepsilon > 0$ there is a gap size N with the following property: for all points $x_1, \dots, x_K \in X$ and iterates $m_1 \le n_1 < m_2 \le n_2 < \dots < m_K \le n_K$ with $m_{k+1} - n_k \ge N$, there is $x \in X$ such that

$$d(T^j(x), T^{j-m_k}(x_k)) < \varepsilon \quad \text{for all } k \in \{1, \dots, K\},\ m_k \le j < n_k. \tag{2.10}$$

Sometimes specification includes the requirement that x is periodic as well (**periodic specification**) and that specification holds for all $K \in \mathbb{N}$ (**strong specification**).

Remark 2.82. For subshifts (X, σ), this definition can be simplified. We give the version for strong specification, because it is the one in most frequent use in this context. There is a gap size N^* such that for all $K \in \mathbb{N}$ and every K-tuple $x^1, \dots, x^K \in X$ and iterates $m_1 \le n_1 < m_2 \le n_2 < \dots < m_K \le n_K$ with $m_{k+1} - n_k \ge N^*$, there is $x \in X$ such that

$$x_j = x^k_{j-m_k} \quad \text{for all } k \in \{1, \dots, K\},\ m_k \le j < n_k. \tag{2.11}$$

Since $d(x, x^k) \le \frac{1}{2}$ if and only if $x_0 = x^k_0$, condition (2.11) implies (2.10) with $N(\varepsilon) = N^*$ for $\varepsilon > \frac{1}{2}$. For $\varepsilon \in (0, \frac{1}{2}]$, condition (2.11) implies (2.10) with $N(\varepsilon) = N^* + n$ where n is minimal such that $2^{-n} < \varepsilon$.

The strength of specification is that a single orbit can shadow many other orbits consecutively, in particular orbits that have different dynamical behaviors.

Lemma 2.83. *A dynamical system with specification is topologically mixing and if the specification is periodic, then the set of periodic orbits is dense.*

Proof. Specification allows one to connect ε-neighborhoods of any two points x_1, x_2 by an orbit of length $N = N(\varepsilon)$. To show topological mixing, take $n \geq 1$ arbitrary and $m_1 = n_1 = 0$, $m_2 = n_2 = n_1 + N(\varepsilon)$ and $m_2' = n_2' = n_1 + N + n$ as in the definition of specification. Then there are $x, x' \in B_\varepsilon(x_1)$ such that $T^N(x) \in B_\varepsilon(x_2)$ and $T^{N+n}(x) \in B_\varepsilon(x_2)$ as required. Finally, for any $x_1 \in X$ and $\varepsilon > 0$, we can find a periodic point $x \in B_\varepsilon(x_1)$, so the set of periodic points is dense. □

The next result is due to Bowen [**97**] and in more generality to Sigmund [**492**, Proposition 3].

Proposition 2.84. Every continuous dynamical system with specification on a compact metric space has positive topological entropy.

Proof. Take distinct point $a, b \in X$ and let $\varepsilon = d(a,b)/3$. Let N be the gap size associated to ε. Now for every $K \in \mathbb{N}$ and chain $\{x_1, \dots, x_K\} \subset \{a, b\}^K$ and the integers Let $m_k = n_k = m_{k+1} - N$, there is a point x such that $d(T^{m_k}(x), x_k) < \varepsilon$ for $k = 1, \dots, K$. There are 2^K choices of $\{x_1, \dots, x_K\}$ and the corresponding points x are (n_K, ε)-separated. Hence, according to Definition (2.4), $h_{top}(T) \geq \frac{1}{1+N} \log 2 > 0$. □

The following was first shown by Bowen [**99**].

Lemma 2.85. *Every continuous factor of a dynamical system (X, T) with specification on a compact metric space (X, d) has specification.*

Proof. Let (Y, S) be a factor of (X, T) such that $\pi : X \to Y$ is the semi-conjugacy. Since X is compact, π is uniformly continuous. Choose $\varepsilon > 0$ arbitrary, and take $\delta > 0$ such that the π-image of every δ-neighborhood in X is contained in an ε-neighborhood in Y. Find $N = N(\delta)$ as in Definition 2.81 of specification for (X, T). Choose $K \in \mathbb{N}$ and $m_1 \leq n_1 < m_2 \leq n_2 < \cdots < m_K \leq n_K$ with gaps $m_{k+1} - n_k \geq N$ and points $y_1, \dots, y_K \in Y$ arbitrary. Choose $x_k \in \pi^{-1}(y_k)$ for each $1 \leq k \leq K$. Since (X, T) has specification, there is $x \in X$ that δ-shadows the pieces of orbits of the x_k's at the required time intervals. Thus $y := \pi(x)$ ε-shadows the pieces of orbits of the y_k's at the required time intervals. This completes the proof. □

Theorem 2.86. *Let (X, T) be an expansive continuous dynamical system on a compact metric space (X, d). If T has specification then it is intrinsically ergodic, i.e., T has a unique measure of maximal entropy.*

This was proven in [**99**], and it applies of course to subshifts. Strong specification makes it possible, and even easy, to approximate invariant measures in the weak* topology by equidistributions on periodic orbits. Indeed, if x is a typical[15] point for an ergodic T-invariant measure μ, then for arbitrarily

[15] in the sense that the Ergodic Theorem 6.13 holds for x.

large n, we can find an n-periodic point p_n that ε-shadows the orbit of x up to iterate $n - N$. The equidistribution $\mu_n := \frac{1}{n}\sum_{i=0}^{n-1}\delta_{T^i(p)}$ then tends to μ as $n \to \infty$. Similar ideas work for non-ergodic measures, see Definition 1.30. An extended version of this argument yields that the measure of maximal entropy is the weak* limit of $\frac{1}{\#\{p:Per(p)\leq n\}}\sum_{Per(p)\leq n}\delta_p$, see **[99]** and **[155]**. Further variations of specification were designed to extend this proof of intrinsic ergodicity to dynamical systems where specification fails, see Buzzi **[131]**, Climenhaga & Thompson **[155, 156]** and Kwietniak and coauthors **[371, 372]**. This applies for instance to (factors of) β-shifts and gap shifts.