CHAPTER 8

Entropy and Chaos

In this chapter we look at two related notions that are important parameters for chaotic dynamical systems. The first is the fractal dimension of a set. By permitting noninteger values, this notion extends the topological concept of dimension to sets such as Cantor sets. While all Cantor sets are homeomorphic, they may look thicker or thinner depending on the parameters in their construction. Fractal dimension is a measure of the thickness of these sets. When the Cantor set in question arises as an invariant set of a hyperbolic dynamical system its dimension is related in deep ways to other dynamically important quantities, notably the contraction and expansion rates in the system. This is an active research topic, and we illustrate it with the Smale horseshoe.

The other notion is entropy. It measures the global orbit complexity on an exponential scale and is intimately related to the growth rate of periodic points and contraction and expansion rates. As an invariant of topological conjugacy, it also provides a means for telling apart dynamical systems that are not conjugate.

The values of dimension and entropy of an invariant set of a dynamical system are related, and so are the constructions involved in defining them. The common root is the notion of *capacity* of a set, with which we begin the chapter.

8.1 DIMENSION OF A COMPACT SPACE

8.1.1 Capacity

For a compact metric space there is a notion of the "size" or capacity inspired by the notion of volume. Suppose X is a compact space with metric d. Then a set $E \subset X$ is said to be *r-dense*if $X \subset \bigcup_{x\in E} B_d(x, r)$, where $B_d(x, r)$ is the r-ball with respect to d around x (see Section 2.6.1). Define the *r-capacity* of (X, d) to be the minimal cardinality $S_d(r)$ of an r-dense set.

For example, if $X = [0, 1]$ with the usual metric, then $S_d(r)$ is approximately $1/2r$ because it takes over $1/2r$ balls (that is, intervals) to cover a unit length, and the $\lfloor 2 + 1/2r\rfloor$-balls centered at $ir(2 - r)$, $0 \le i \le \lfloor 1 + 1/2r\rfloor$ suffice. As another example, if $X = [0, 1]^2$ is the unit square, then $S_d(r)$ is roughly r^{-2}

because it takes at least $1/\pi r^2$ r-balls to cover a unit area, and, on the other hand, the $(1+1/r)^2$-balls centered at points (ir, jr) provide a cover. Likewise, for the unit cube $(1+1/r)^3$, r-balls suffice.

In the case of the ternary Cantor set with the usual metric we have $S_d(3^{-i}) = 2^i$ if we cheat a little and use closed balls for simplicity; otherwise, we could use $S_d((3-1/i)^{-i}) = 2^i$ with honest open balls.

8.1.2 Box Dimension

One interesting aspect of capacity is the relation between its dependence on r [that is, with which power of r the capacity $S_d(r)$ increases] and dimension.

If $X = [0, 1]$, then

$$\lim_{r\to 0} -\frac{\log S_d(r)}{\log r} \geq \lim_{r\to 0} -\frac{\log(1/2r)}{\log r} = \lim_{r\to 0} \frac{\log 2 + \log r}{\log r} = 1$$

and

$$\lim_{r\to 0} -\frac{\log S_d(r)}{\log r} \leq \lim_{r\to 0} -\frac{\log\lfloor 2 + 1/2r\rfloor}{\log r} \leq \lim -\frac{\log(1/r)}{\log r} = 1,$$

so $\lim_{r\to 0} -\log S_d(r)/\log r = 1 = \dim X$. If $X = [0, 1]^2$, then $\lim_{r\to 0} -\log S_d(r)/\log r = 2 = \dim X$; and if $X = [0, 1]^3$, then $\lim_{r\to 0} -\log S_d(r)/\log r = 3 = \dim X$. This suggests that $\lim_{r\to 0} -\log S_d(r)/\log r$ defines a notion of dimension.

Definition 8.1.1 If X is a totally bounded metric space (Definition A.1.20), then

$$\operatorname{bdim}(X) := \lim_{r\to 0} -\frac{\log S_d(r)}{\log r}$$

is called the *box dimension* of X.

8.1.3 Examples

Let us test this notion on less straightforward spaces.

1. The Ternary Cantor Set. If C is the ternary Cantor set, then

$$\operatorname{bdim}(C) = \lim_{r\to 0} -\frac{\log S_d(r)}{\log r} = \lim_{n\to\infty} -\frac{\log 2^i}{\log 3^{-i}} = \frac{\log 2}{\log 3}.$$

If C_α is constructed by deleting a middle interval of relative length $1-(2/\alpha)$ at each stage, then $\operatorname{bdim}(C_\alpha) = \log 2/\log\alpha$. This increases to 1 as $\alpha \to 2$ (deleting ever smaller intervals), and it decreases to 0 as $\alpha \to \infty$ (deleting ever larger intervals). Thus we get a small box dimension if in the Cantor construction the size of the remaining intervals decreases rapidly with each iteration.

This illustrates, by the way, that the box dimension of a set may change under a homeomorphism, because these Cantor sets are pairwise homeomorphic.

2. The Sierpinski Carpet. It is easy to handle other Cantor-like sets, such as the Sierpinski carpet S from Section 2.7.2. For the square Sierpinski carpet we can cheat as in the capacity calculation for the ternary Cantor set and use closed balls (sharing their center with one of the small remaining cubes at a certain stage) for covers. Then $S_d(3^{-i}/\sqrt{2}) = 8^i$ and

$$\text{bdim}(S) = \lim_{n\to\infty} -\frac{\log 8^i}{\log 3^{-i}/\sqrt{2}} = \frac{\log 8}{\log 3} = \frac{3\log 2}{\log 3},$$

which is three times that of the ternary Cantor set (but still less than 2, of course). For the triangular Sierpinski carpet we similarly get box dimension $\log 3/\log 2$.

3. The Koch Snowflake. The Koch snowflake K from Section 2.7.2 has $S_d(3^{-i}) = 4^i$ by covering it with (closed) balls centered at the edges of the ith polygon. Thus

$$\text{bdim}(K) = \lim_{n\to\infty} -\frac{\log 4^i}{\log 3^{-i}} = \frac{\log 4}{\log 3} = \frac{2\log 2}{\log 3},$$

which is less than that of the Sierpinski carpet, corresponding to the fact that the iterates look much "thinner". Notice that this dimension exceeds 1, however, so it is larger than the dimension of a curve. All of these examples have (box) dimension that is not an integer, that is, fractional or "fractal". This has motivated calling such sets *fractals*.

4. The Smale Horseshoe. Suppose that in the construction of the Smale horseshoe (Section 7.4.4) the expansion rate on the linear pieces is $\lambda > 2$ and the contraction rate is $\mu < 1/\lambda$ (without loss of generality). Given $n \in \mathbb{N}$, the invariant set $\Lambda = \bigcap_{n=-\infty}^{\infty} f^{-n}(\Delta)$ is contained in $\Lambda = \bigcap_{i=-n}^{n} f^{-i}(\Delta)$, which consists of 4^n rectangles with sides λ^{-n} and μ^n and can therefore be covered by about $4^n/(\lambda^n\mu^n)$ squares with sides μ^n. Thus $S_d(\mu^{-n}) \asymp 4^n/(\lambda^n\mu^n)$ and

$$\begin{aligned}\text{bdim}(\Lambda) &= \lim_{n\to\infty} -\frac{\log S_d(\mu^{-n})}{\log \mu^{-n}} = \lim_{n\to\infty} -\frac{n(\log 4 - \log\lambda - \log\mu)}{n\log\mu}\\ &= 1 + \frac{\log 4 - \log\lambda}{-\log\mu}.\end{aligned}$$

5. Sequence Spaces. Consider the two-sided sequence space Ω_N with the metric d_λ of (7.3.4). According to (7.3.5), there is a disjoint cover by N^{2n-1} balls of radius λ^{1-n}, namely, the cylinders $C_{\alpha_{1-n}\ldots\alpha_{n-1}} = \{\omega \in \Omega_N \mid \omega_i = \alpha_i \text{ for } |i| < n\}$. Therefore $S_{d_\lambda}(\lambda^{1-n}) = N^{2n-1}$ and hence the box dimension is

$$\text{bdim}(\Omega_N, d_\lambda) = \lim_{r\to 0} -\frac{\log S_d(r)}{\log r} = \lim_{n\to\infty} -\frac{\log N^{2n-1}}{\log\lambda^{1-n}} = \lim_{n\to\infty}\frac{2n-1}{n-1}\frac{\log N}{\log\lambda} = 2\frac{\log N}{\log\lambda}.$$

Analogously to the Cantor set example, the box dimension decreases as λ increases, corresponding to the rapid decrease of the radius of cylinders (as a function of the length of the specified string) for large λ.

8.1.4 Dependence on the Metric

A different issue related to capacity is the dependence of $S_d(r)$ on the metric for a given r. If one replaces a metric by a larger one (with finer resolution, as it were), then balls become smaller and hence $S_d(r)$ increases. The rate at which it does so is a new measure of the rate of refinement of the metrics. A simple example is scaling of the metric, that is, multiplying by a positive factor a. Clearly $S_{ad}(ar) = S_d(r)$ and

$$\lim_{r\to 0} -\frac{\log S_{ad}(r)}{\log r} = \lim_{r\to 0} -\frac{\log S_{ad}(ar)}{\log ar} = \lim_{r\to 0} -\frac{\log S_d(r)}{\log ar}$$
$$= \lim_{r\to 0} -\frac{\log S_d(r)}{\log a + \log r} = \lim_{r\to 0} -\frac{\log S_d(r)}{\log r}.$$

Thus, scaling does not affect the box dimension. However, one may study the asymptotic behavior of $S_{d_i}(r)$ for a sequence d_i of metrics as $i \to \infty$ for fixed r. We presently do this in our study of entropy.

EXERCISES

Exercise 8.1.1 Prove that the cardinality of a minimal cover is not always the same as the minimal cardinality of a cover.

Exercise 8.1.2 Compute the box dimension of $\mathbb{Q} \cap [0, 1]$.

Exercise 8.1.3 For the Smale horseshoe show that $0 < \text{bdim}(\Lambda) < 2$.

Exercise 8.1.4 For an S-shaped horseshoe with three crossings compute the box dimension of the invariant set and prove that it lies between 0 and 2.

Exercise 8.1.5 Find the dimension of the metric d''_λ (7.3.10) on Ω_N and Ω_N^R.

Exercise 8.1.6 Show that the dimension $\text{bdim}(\Omega_N^R, d_\lambda)$ of the one-sided shift space Ω_N^R with the metric d_λ is $\log N/\log\lambda$.

Exercise 8.1.7 Show that the triangular Sierpinski carpet has box dimension $\log 3/\log 2$.

Exercise 8.1.8 Construct Cantor sets on the interval with box dimension 0 and 1.

Exercise 8.1.9 Determine the box dimension of the set of points in [0, 1] that have a binary expansion with no consecutive 0's.

8.2 TOPOLOGICAL ENTROPY

8.2.1 Measures of Complexity and Invariants

We have encountered several indicators of the complexity of a dynamical system: topological transitivity, minimality, density of the set of periodic points, chaos, and topological mixing. Especially the latter indicate the presence of intertwining and separation of different orbits. These are all qualitative ("yes–no") measures of complexity. So far the only quantitative measure of complexity is the growth rate of periodic orbits. While the otherwise simple rational rotations have infinitely many periodic points, it is chaotic examples that are distinguished by the exponential growth of finite numbers of periodic points.

1. Entropy. A step beyond the periodic orbit growth is to measure the growth of all orbits in some sense. This is done by the most important numerical invariant related to the orbit growth, the topological entropy. It represents the exponential growth rate for the number of orbit segments distinguishable with arbitrarily fine but finite precision. In a sense, the topological entropy describes in a crude but suggestive way the total exponential complexity of the orbit structure with a single number. Indeed, we will see that the chaotic systems from among our examples are distinguished by having positive entropy, and the topological entropy is no less than the growth rate of periodic orbits. Therefore it is appropriate to view entropy as a quantitative measure of the amount of chaos in a dynamical system.

2. Invariants. At this point it might be useful to give another motivation for studying invariants of dynamical systems. Invariants are quantities associated with a dynamical system that agree for two dynamical systems that are equivalent in the sense of conjugacy (Definition 7.3.3). When one encounters a new dynamical system it is natural to wonder whether it is equivalent to a previously studied one, which would save a lot of work; or one may try to see whether certain collections of dynamical systems are pairwise equivalent or can be subdivided neatly into equivalence classes (under topological conjugacy). Either way, one needs to decide whether there is a conjugacy between two given systems. If one is unable, after much trying, to find one, the need becomes apparent for methods to show that there can be no conjugacy. Invariants provide a means to do this: If one system is transitive and the other one is not, then they cannot be conjugate. If one circle homeomorphism has rotation number α and another has rotation number $\beta \neq \alpha$, then these two homeomorphisms are not topologically conjugate. Similarly, entropy is an attractive invariant (Corollary 8.2.3) not least for the reason that it takes on real values (as opposed to "yes–no" only) and hence gives a finer distinction between different dynamical systems than transitivity, mixing, and so on. On the other hand, it is defined for a broad class of dynamical systems rather than only circle maps.

8.2.2 First Definition of Entropy

To define entropy we measure the rate of increase of the capacity $S_d(r)$ for fixed r as the metric is refined in a dynamically significant way. This is different from the definition of box dimension, where we study the change in capacity as a function of r for a fixed metric. Suppose $f\colon X \to X$ is a continuous map of a compact metric space X with distance function d and define an increasing sequence of metrics d_n^f, $n = 1, 2, \dots,$ starting from $d_1^f = d$ by

$$d_n^f(x, y) = \max_{0 \le i \le n-1} d(f^i(x), f^i(y)). \tag{8.2.1}$$

In other words, d_n^f is the distance between the orbit segments $\mathcal{O}_n(x) = \{x, \dots, f^{n-1}x\}$ and $\mathcal{O}_n(y)$. We denote the open ball $\{y \in X \mid d_n^f(x, y) < r\}$ by $B_f(x, r, n)$.

Definition 8.2.1 Let $S_d(f, r, n)$ be the r-capacity of d_n^f. Explicitly, a set $E \subset X$ is r-dense with respect to d_n^f, or (n, r)*-dense,* if $X \subset \bigcup_{x \in E} B_f(x, r, n)$. Then $S_d(f, r, n)$ is the minimal cardinality of an (n, r)-dense set or, equivalently, the cardinality of a *minimal* (n, r)*-dense set.*This is the minimal number of initial conditions whose behavior up to time n approximates the behavior of *any* initial condition up to r.

Consider the exponential growth rate

$$h_d(f, r) := \overline{\lim_{n\to\infty}} \frac{1}{n} \log S_d(f, r, n) \tag{8.2.2}$$

of $S_d(f, r, n)$. Obviously $h_d(f, r)$ does not decrease with r, so we can define

$$h_d(f) := \lim_{r\to 0} h_d(f, r). \tag{8.2.3}$$

Call $h(f) := h_{\text{top}}(f) := h_d(f)$ the *topological entropy* of f.

Note that we take a double limit, first with respect to n and then with respect to r. The important limit is the one in n, because it is there that the dynamics enters. In many interesting cases the limit in r is, in fact, trivial, because $h_d(f, r)$ is independent of r (for small r) to begin with.

A priori, $h_d(f)$ might depend on the metric d. Actually it does not, so long as one changes to a homeomorphic metric (Definition A.1.17). This justifies dropping the reference to the metric in (8.2.3).

Proposition 8.2.2 *If d' is a metric on X equivalent to d, then $h_{d'}(f) = h_d(f)$.*

Proof The identity map $\text{Id}\colon (X, d) \to (X, d')$ is a homeomorphism by assumption and uniformly continuous in both directions by the compactness of X. Thus, given $r > 0$, there exists a $\delta(r) > 0$ such that, if $d'(x_1, x_2) < \delta$, then $d(x_1, x_2) < r$, that is, any δ-ball in the metric d' is contained in an r-ball in the metric d. By (8.2.1) this also holds for ${d'}^f_n$ and d^f_n. Thus $S_{d'}(f, \delta, n) \geq S_d(f, r, n)$ for every n, so $h_{d'}(f, \delta) \geq h_d(f, r)$ and $h_{d'}(f) \geq \lim_{\delta\to 0} h_{d'}(f, \delta) \geq \lim_{r\to 0} h_d(f, r) = h_d(f)$. Interchanging d and d' one obtains $h_d(f) \geq h_{d'}(f)$, and hence equality. □

Corollary 8.2.3 *Topological entropy is an invariant of topological conjugacy.*

Proof Let $f\colon X \to X$, $g\colon Y \to Y$ be topologically conjugate via a homeomorphism $h\colon X \to Y$ (see Definition 7.3.3). Fix a metric d on X and define d' on Y as the pullback of d, that is, $d'(y_1, y_2) = d(h^{-1}(y_1), h^{-1}(y_2))$ (Section 2.6.1). Then h becomes an isometry, so $h_d(f) = h_{d'}(g)$. □

8.2.3 Subexponential Growth

As a first example of how to apply this concept, consider situations with relatively simple dynamics.

Proposition 8.2.4 *The topological entropy of contractions and isometries is zero. In particular, any translation T_γ of the torus or any linear flow T^t_ω on the torus (see Section 5.1) has zero entropy.*

Proof If X is a compact metric space and $f\colon X \to X$ is 1-Lipschitz, then $d^f_n = d$ for all n and consequently $S_n(f, r, n)$ does not depend on n; so $h(f) = 0$. The situation with isometric flows is completely similar to that of maps. □

This absence of any growth is most removed from the case of positive topological entropy. Between these two extreme cases there is a variety of situations of "moderate", that is, subexponential, growth for those quantities. An example is given by the linear twist $T\colon S^1 \times [0,1] \to S^1 \times [0,1]$, $T(x, y) = (x + y, y)$ in Section 6.1.1. In this case we can give a d_n^f-r-dense set of nr^2 balls with centers spaced uniformly r apart along the horizontal and uniformly nr apart on the vertical. The centers are then also $r/2$-separated.

8.2.4 Entropy via Covers

Topological entropy is not always easy to calculate, and it helps to have alternative definitions in order to be able to choose a convenient one as the situation requires (this comes in handy already in Proposition 8.2.9).

There are several quantities similar to $S_d(f, r, n)$ that can be used to define topological entropy. Let $D_d(f, r, n)$ be the minimal number of sets whose diameter in the metric d_n^f is less than r and whose union covers X.

Lemma 8.2.5 $\tilde{h}_d(f, r) := \lim_{n\to\infty}(1/n)\log D_d(f, r, n)$ *exists for any* $r > 0$.

Proof If A is a set of d_n^f-diameter less than r and B is a set of d_m^f-diameter less than r, then $A \cap f^{-n}(B)$ has d_{m+n}^f-diameter less than r. Thus if $\mathfrak{A}$ is a cover of X by $D_d(f, r, n)$ sets of d_n^f-diameter less than r and $\mathfrak{B}$ is a cover of X by $D_d(f, r, m)$ sets of d_m^f-diameter less than r, then the cover by all sets $A \cap f^{-n}(B)$, where $A \in \mathfrak{A}$, $B \in \mathfrak{B}$, contains at most $D_d(f, r, n) \cdot D_d(f, r, m)$ sets and is a cover by sets of d_{m+n}^f-diameter less than r. Thus

$$D_d(f, r, m+n) \le D_d(f, r, n) \cdot D_d(f, r, m)$$

for all m, n. For $a_n = \log D_d(f, r, n)$, this means $a_{m+n} \le a_n + a_m$ and hence $\lim_{n\to\infty} a_n/n$ exists by Lemma 4.3.7. □

Proposition 8.2.6 *If* $\underline{h}_d(f, r) := \underline{\lim}_{n\to\infty}(1/n)\log S_d(f, r, n)$, *then*

$$\lim_{r\to 0}\tilde{h}_d(f, r) = \lim_{r\to 0}\underline{h}_d(f, r) = \lim_{r\to 0} h_d(f, r) = h(f). \tag{8.2.4}$$

Proof The diameter of an r-ball is at most $2r$, so every covering by r-balls is a covering by sets of diameter $\le 2r$, that is,

$$D_d(f, 2r, n) \le S_d(f, r, n). \tag{8.2.5}$$

On the other hand, any set of diameter $\le r$ is contained in the r-ball around each of its points, so

$$S_d(f, r, n) \le D_d(f, r, n). \tag{8.2.6}$$

Thus

$$\tilde{h}_d(f, 2r) \le \underline{h}_d(f, r) \le h_d(f, r) \le \tilde{h}_d(f, r). \ \square$$

8.2.5 Topological Entropy via Separated Sets

Another way to define topological entropy is via the maximal number $N_d(f, r, n)$ of points in X with pairwise d_n^f-distances at least r. We say that such a set of points is (n, r)-*separated*. (See Figure 8.2.1.) Such points generate the maximal number of orbit segments of length n that are distinguishable with precision r.

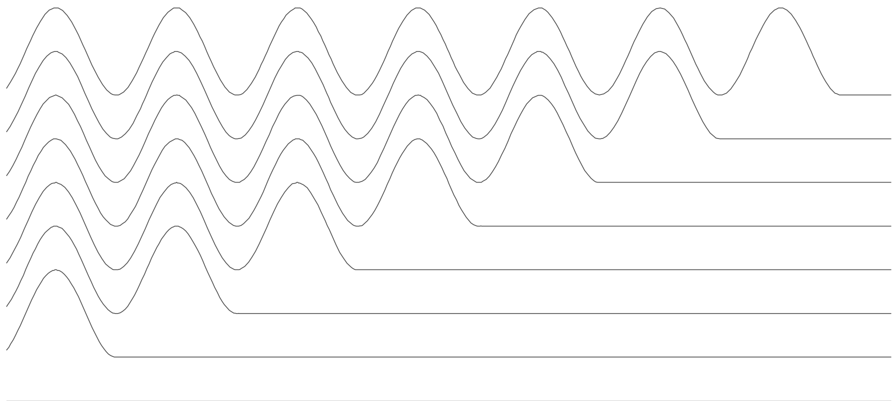

Figure 8.2.1. A separated set.

Proposition 8.2.7

$$h_{top}(f) = \lim_{r\to 0} \overline{\lim_{n\to\infty}} \frac{1}{n} \log N_d(f, r, n) = \lim_{r\to 0} \underline{\lim_{n\to\infty}} \frac{1}{n} \log N_d(f, r, n). \tag{8.2.7}$$

Remark 8.2.8 This justifies the verbal description of entropy as the exponential growth rate for the number of orbit segments distinguishable with arbitrarily fine but finite precision that we gave at the beginning of this section.

Proof A maximal (n, r)-separated set is (n, r)-dense, that is, for any such set of points the r-balls around them cover X, because otherwise it would be possible to increase the set by adding any point not covered. Thus

$$S_d(f, r, n) \le N_d(f, r, n). \tag{8.2.8}$$

On the other hand, no $r/2$-ball can contain two points r apart. Thus

$$N_d(f, r, n) \le S_d\left(f, \frac{r}{2}, n\right). \tag{8.2.9}$$

Using (8.2.8) and (8.2.9) we obtain

$$\underline{h}_d(f, r) \le \underline{\lim_{n\to\infty}} \frac{1}{n} \log N_d(f, r, n) \le \overline{\lim_{n\to\infty}} \frac{1}{n} \log N_d(f, r, n) \le h_d\left(f, \frac{r}{2}\right). \tag{8.2.10}$$

The result follows by Proposition 8.2.6. □

8.2.6 Some Properties of Entropy

The following proposition contains some standard elementary properties of topological entropy. The proofs demonstrate the usefulness of switching back and forth from one of the three definitions to another.

Proposition 8.2.9

(1) *If Λ is a closed f-invariant set, then $h_{top}(f\restriction_\Lambda) \le h_{top}(f)$.*
(2) *If $X = \bigcup_{i=1}^m \Lambda_i$, where Λ_i, $(i = 1, \dots, m)$ are closed f-invariant sets, then $h_{top}(f) = \max_{1\le i\le m} h_{top}(f\restriction_{\Lambda_i})$.*
(3) *$h_{top}(f^m) = |m| h_{top}(f)$.*
(4) *If g is a factor of f, then $h_{top}(g) \le h_{top}(f)$.*

(5) $h_{top}(f \times g) = h_{top}(f) + h_{top}(g)$, *where* $f\colon X \to X$, $g\colon Y \to Y$ *and* $f \times g\colon X \times Y \to X \times Y$ *is defined by* $(f \times g)(x, y) = (f(x), g(y))$.

Proof Statement (1) is obvious since every cover of X by sets of d_n^f-diameter less than r is at the same time a cover of Λ.

To prove (2) note that $D_d(f, r, n) \le \sum_{i=1}^m D_d(f\restriction_{\Lambda_i}, r, n)$, because the union of covers of $\Lambda_1, \dots, \Lambda_m$ by sets of diameter less than r is a cover of X. Thus

$$D_d\big(f\restriction_{\Lambda_i}, r, n\big) \ge \frac{1}{m} D_d(f, r, n)$$

for at least one i. Since there are only finitely many i's, at least one i works for infinitely many n. For this $i \in \{1, \dots, m\}$

$$\overline{\lim_{n\to\infty}} \frac{\log D_d\big(f\restriction_{\Lambda_i}, r, n\big)}{n} \ge \overline{\lim_{n\to\infty}} \frac{\log D_d(f, r, n) - \log m}{n} = \tilde{h}_d(f, r).$$

Together with (1) this proves (2).

If m is positive, then (3) follows from two remarks. First

$$d_n^{f^m}(x, y) = \max_{0\le i\le n-1} d(f^{im}(x), f^{im}(y)) \le \max_{0\le i\le mn-1} d(f^i(x), f^i(y)) = d_{nm}^f(x, y),$$

so any $d_n^{f^m}$ r-ball contains a d_{mn}^f r-ball and

(8.2.11) $$S_d(f^m, r, n) \le S_d(f, r, mn).$$

Hence $h_{\mathrm{top}}(f^m) \le m h_{\mathrm{top}}(f)$. On the other hand, for every $r > 0$ there is a $\delta(r) > 0$ such that $B(x, \delta(r)) \subset B_f(x, r, m)$ for all $x \in X$. Thus

$$\begin{aligned} B_{f^m}(x, \delta(r), n) &= \bigcap_{i=0}^{n-1} f^{-im} B(f^{im}(x), \delta(r)) \\ &\subset \bigcap_{i=0}^{n-1} f^{-im} B_f(f^{im}(x), r, m) = B_f(x, r, mn). \end{aligned}$$

Consequently,

$$S_d(f, r, mn) \le S_d(f^m, \delta(r), n)$$

and $m h_{\mathrm{top}}(f) \le h_{\mathrm{top}}(f^m)$. If f is invertible, then $B_f(x, r, n) = B_{f^{-1}}(f^{n-1}(x), r, n)$ and $S_d(f, r, n) = S_d(f^{-1}, r, n)$; so $h_{\mathrm{top}}(f) = h_{\mathrm{top}}(f^{-1})$.

If m is negative, then (3) follows from the statement for $m > 0$ and $n = -1$.

Statement (4) deals with $f\colon X \to X$, $g\colon Y \to Y$, $h\colon X \to Y$ such that $h \circ f = g \circ h$ and $h(X) = Y$ (Definition 7.3.3). Denote by d_X, d_Y the distance functions in X and Y, correspondingly.

h is uniformly continuous, so for any $r > 0$ there is $\delta(r) > 0$ such that, if $d_X(x_1, x_2) < \delta(r)$, then $d_Y(h(x_1), h(x_2)) < r$. Thus the image of any $(d_X)_n^f$ ball of radius $\delta(r)$ lies inside a $(d_Y)_n^f$ ball of radius r, that is,

$$S_{d_X}(f, \delta(r), n) \ge S_{d_Y}(g, r, n).$$

Taking logarithms and limits, we obtain (4).

To prove (5) use the product metric $d((x_1, y_1), (x_2, y_2)) = \max(d_X(x_1, x_2), d_Y(y_1, y_2))$ in $X \times Y$. Balls in the product metric are products of balls on X and Y. The same is true for balls in $d_n^{f\times g}$. Thus $S_d(f \times g, r, n) \le S_{d_X}(f, r, n) S_{d_Y}(g, r, n)$

and $h_{\text{top}}(f \times g) \le h_{\text{top}}(f) + h_{\text{top}}(g)$. On the other hand, the product of any (n, r)-separated set in X for f and any (n, r)-separated set in Y for g is an (n, r)-separated set for $f \times g$. Thus

$$N_d(f \times g, r, n) \ge N_{d_X}(f, r, n) \times N_{d_Y}(g, r, n)$$

and hence $h_{\text{top}}(f \times g) \ge h_{\text{top}}(f) + h_{\text{top}}(g)$. □

EXERCISES

Exercise 8.2.1 Compute the topological entropy of $f(x) = x(1 - x)$ on $[0, 1]$.

Exercise 8.2.2 Compute the topological entropy of the linear horseshoe.

Exercise 8.2.3 Suppose $f: S^1 \to S^1$ is an orientation-preserving C^2-diffeomorphism without periodic points. Find $h_{\text{top}}(f)$.

Exercise 8.2.4 Let $f: \mathbb{T}^3 \to \mathbb{T}^3$, $f(x, y, z) = (x, x + y, y + z)$. Find $h_{\text{top}}(f)$.

Exercise 8.2.5 Suppose $X = \bigcup_i X_i$ is compact, $f: X \to X$ such that each X_i is closed and f-invariant. Show that $h_{\text{top}}(f) = \sup h_{\text{top}}(f\restriction_{X_i})$.

PROBLEMS FOR FURTHER STUDY

Problem 8.2.6 Given $f: X \to X, g: Y \to Y$, suppose $h \circ f = g \circ h$, where $h: X \to Y$ is a continuous surjective map such that every $y \in Y$ has finitely many preimages. Show that $h_{\text{top}}(f) = h_{\text{top}}(g)$.

8.3 APPLICATIONS AND EXTENSIONS

8.3.1 Expanding Maps

The expanding maps E_m represent the first situation in our survey where a really complicated orbit structure appears. Since one of the features of this structure is the exponential growth of periodic orbits (Proposition 7.1.2), it is natural to expect the total exponential orbit complexity, measured by the topological entropy, to be positive too.

Proposition 8.3.1 *If $m \in \mathbb{N}$, $|m| \ge 2$, then $h_{top}(E_m) = \log|m| = p(E_m)$.*

Proof For the map E_m, and in fact for any expanding map, the distance between iterates of any two points grows until it becomes greater than a certain constant depending on the map ($1/2|m|$ for the map E_m). To simplify notations, assume $m > 0$. If $d(x, y) < m^{-n}/2$, then $d_n^{E_m}(x, y) = d(E_m^{n-1}(x), E_m^{n-1}(y))$; so if $d_n^{E_m}(x, y) \ge r$, then $d(x, y) \ge rm^{-n}$. Taking $r = m^{-k}$, this shows that $\{im^{-n-k} \mid i = 0, \dots, m^{n+k} - 1\}$ is a maximal set of points whose pairwise $d_n^{E_m}$-distances are at least m^{-k}, that is,

$$N_d(E_m, m^{-k}, n) = m^{n+k},$$

and consequently

$$h(E_m) = \lim_{k\to\infty} \overline{\lim_{n\to\infty}} \frac{\log N_d(E_m, m^{-k}, n)}{n} = \lim_{k\to\infty} \lim_{n\to\infty} \frac{n+k}{n} \log m = \log m.$$

The case $m < 0$ is completely parallel. □

Since topological entropy is invariant under topological conjugacy (Corollary 8.2.3) and every expanding map of degree m is topologically conjugate to the map E_m (Theorem 7.4.3), we obtain from Proposition 8.3.1

Corollary 8.3.2 *If $f: S^1 \to S^1$ is an expanding map of degree m, then*

$$h_{top}(f) = p(f) = \log|m|.$$

8.3.2 Shifts and Topological Markov Chains

Proposition 8.3.3 $h_{top}(\sigma_A) = p(\sigma_A) = \log|\lambda_A^{\max}|$ *for any topological Markov chain* σ_A.

Proof Analogously to Section 7.3.4, any cylinder

$$C^{-m,\dots,n+m}_{\alpha_{-m},\dots,\alpha_{n+m}} := \{\omega \in \Omega_N \mid \omega_i = \alpha_i \text{ for } -m \le i \le m+n\} \tag{8.3.1}$$

is at the same time the ball of radius $r_m = \lambda^{-m}/2$ around each of its points with respect to the metric $d_n^{\sigma_N}$ associated with the shift σ_N (because $\lambda > 3$). Thus, any two $d_n^{\sigma_N}$ balls of radius r_m are either identical or disjoint, and there are exactly N^{n+2m+1} different ones of the form (8.3.1); so $S_{d_\lambda}(\sigma_N, r_m, n) = N^{n+2m+1}$ and

$$h(\sigma_N) = \lim_{m\to\infty} \lim_{n\to\infty} \frac{1}{n} \log N^{n+2m+1} = \log N.$$

Similarly, if σ_A is a topological Markov chain, then $S_d(\sigma_A, r_m, n)$ is the number of those cylinders (8.3.4) that have nonempty intersection with Ω_A. Assume that each row of A contains at least one 1. Since the number of admissible paths of length n that begin with i and end with j is the entry a_{ij}^n of A^n (see Lemma 7.3.5), the number of nonempty cylinders of rank $n+1$ in Ω_A is $\sum_{i,j=0}^{N-1} a_{ij}^n < C \cdot \|A^n\|$ for some constant C. On the other hand, $\sum_{i,j=0}^{N-1} a_{ij}^n > c\|A^n\|$ for another constant $c > 0$ because all numbers a_{ij}^n are nonnegative and hence the left-hand side is the norm $\sum_{i,j=0}^{N-1} a_{ij}^n$ of A^n, which is equivalent to the usual norm because all norms on $\mathbb{R}^{N^2}$ are equivalent. Thus, we have

$$S_{d_\lambda}(\sigma_A, r_m, n) = \sum_{i,j=0}^{N-1} a_{ij}^{n+2m} \tag{8.3.2}$$

and

$$\begin{aligned} h(\sigma_A) &= \lim_{m\to\infty} \overline{\lim_{n\to\infty}} \frac{1}{n} \log S_{d_\lambda}(\sigma_A, r_m, n) \\ &= \lim_{m\to\infty} \lim_{n\to\infty} \frac{1}{n} \log\|A^{n+2m}\| = \lim_{n\to\infty} \frac{1}{n} \log\|A^n\| = \log r(A) = \log\left|\lambda_A^{\max}\right|, \end{aligned} \tag{8.3.3}$$

where $r(A)$ is the spectral radius of the matrix A (Definition 3.3.1). Equation (8.3.3) and Proposition 7.3.7 now give the claim. □

8.3.3 The Hyperbolic Toral Automorphism

In calculating the entropy of the toral automorphism we use both coding and our knowledge of the growth rate of periodic points.

Proposition 8.3.4 *If $F_L\colon \mathbb{T}^2 \to \mathbb{T}^2$ is given by $F_L(x, y) = (2x + y, x + y) \pmod 1$, then*

$$h(F_L) = p(F_L) = \frac{3+\sqrt{5}}{2}.$$

Proof In Section 7.4.5 we showed that

$$F_L(x, y) = (2x + y, x + y) \pmod 1$$

is a factor of the topological Markov chain σ_A, where

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

and that the maximal eigenvalue of A is $\lambda_A^{\max} = 3 + \sqrt{5}/2$. Proposition 8.2.9(4) shows that

$$h(F_L) \le h(\sigma_A) = \log \frac{3+\sqrt{5}}{2}. \tag{8.3.4}$$

On the other hand, we next show that the set of n-periodic points of F_L is $(n, 1/4)$-separated for any $n \in \mathbb{N}$. This implies $N_d(F_L, 1/4, n) \ge P_n(F_L)$ and

$$h(F_L) \ge p(F_L) = \log \frac{3+\sqrt{5}}{2}$$

by Proposition 7.1.10. By (8.3.4), the result then follows.

If p, q are n-periodic points and $d(p, q) < 1/4$, then there is a uniquely defined minimal rectangle R with vertices p, s, q, t formed by segments of expanding and contracting lines passing through p and q. (See Figure 8.3.1.) Under the action of F_L the sides ps and qt expand with coefficient $\lambda_1 = (3 + \sqrt{5}/2) > 2$ while the other two sides contract with coefficient λ_1^{-1}.

This implies $F_L^n(R) \ne R$ because F_L^n cannot leave all four sides invariant while also expanding and contracting them. Equivalently, $F_L^{-n}(R) \ne R$. Therefore, there is a $k \le n$ for which $F_L^k(R)$ is not a minimal rectangle. For the smallest such k we then

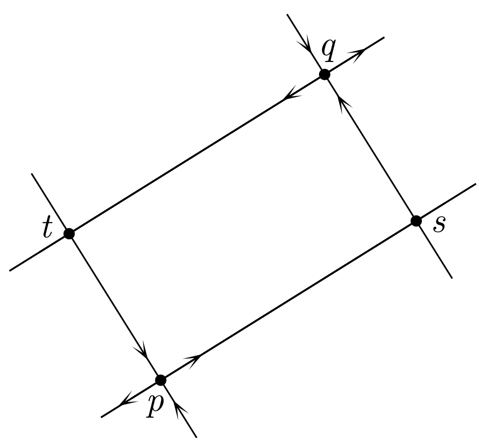

Figure 8.3.1. Heteroclinic points.

have $d(F_L^k(p), F_L^k(q)) > 1/4$ because a rectangle with diagonals shorter than 1/4 is minimal. Thus the periodic points of period n form an $(n, 1/4)$-separated set. □

Remark 8.3.5 In the case of expanding maps E_m and for topological Markov chains σ_A one can also show that periodic points form (n, r_0)-separated sets for some r_0. This allows us to produce the inequality $h_{\text{top}} \geq p$ in a uniform way for all three cases.

8.3.4 Periodic Points and Entropy

Our examples show an interesting pattern. For both smooth examples with complicated exponentially growing orbit structure, namely, expanding maps (Proposition 8.3.1) and hyperbolic toral automorphisms (Proposition 8.3.4), the two natural measures of the exponential orbit growth – the growth rate p of periodic points and the topological entropy h_{top} – coincide. This is a rather widespread phenomenon, although not universal. It is related to the local hyperbolic structure, that is, the stretching and folding common to these examples. (This is systematically introduced in Chapter 10.) For topological Markov chains the growth rate of periodic points and topological entropy also coincide (Proposition 8.3.3). Hyperbolicity is a relevant explanation here, too, since by Proposition 7.4.6 topological Markov chains are topologically conjugate to the restriction of some smooth systems to special invariant sets that possess hyperbolic behavior.

8.3.5 Topological Entropy for Flows

The definition of topological entropy $h_{\text{top}}(\Phi)$ for a flow $\Phi = (\varphi^t)_{t\in\mathbb{R}}$ is completely parallel to that for the discrete-time case. The only change is that the metrics in (8.2.1) are replaced by the nondecreasing family

$$d_T^\Phi(x, y) = \max_{0\leq t\leq T} d(\varphi^t(x), \varphi^t(y))$$

of metrics. This parallelism has a particularly useful consequence analogous to Proposition 8.2.9.(3).

Proposition 8.3.6 $h_{top}(\Phi) = h_{top}(\varphi^1)$.

Proof If $r > 0$, then by compactness and continuity for there is a $\delta(r) > 0$ such that $d(x, y) \leq \delta(r)$ implies $\max_{0\leq t\leq 1} d(\varphi^t(x), \varphi^t(y)) < r$. Then any r-ball in the metric d_T^Φ contains a $\delta(r)$-ball in the metric $d_{\lfloor T\rfloor}^{\varphi^1}$. On the other hand, $d_n^\Phi \geq d_n^{\varphi^1}$. These two remarks imply the statement. □

The topological entropy for a flow is thus invariant under *flow equivalence*, that is, coincides for two flows whose time-t maps are topologically conjugate with the same conjugacy for all t. It changes under time change (Definition 9.4.12) and hence under orbit equivalence (flow equivalence with a time change) in a rather complicated way. One can show that for a flow without fixed points any time change preserves vanishing of the topological entropy, that is, a time change of a flow with zero entropy also has zero entropy. If the topological entropy for a map or a flow vanishes, the subexponential asymptotic of any of the quantities involved in its definition may provide useful insights into the complexity of the orbit structure.

8.3.6 Local Entropy as a Measure of Sensitive Dependence

As we mentioned in the introduction to this section, entropy can be viewed as a measure of the amount of chaos in a system. We now explicitly show how entropy provides a quantitative measure of the amount of sensitive dependence in a dynamical system. To that end we introduce a closely related notion of *local entropy*, explain how it measures sensitive dependence on the one hand, and how it is related to topological entropy on the other.

Fix a point x and a "microscopic" ϵ as well as a "macroscopic" r, and let $N_d(f, r, n, x, \epsilon)$ be the maximal number of points in $B_d(x, \epsilon)$ with pairwise d_n^f-distances at least r. A large such number would certainly indicate rather sensitive dependence on initial conditions.

Definition 8.3.7 If

$$h_{d,x,r}(f) := \lim_{\epsilon\to 0} \overline{\lim_{n\to\infty}} \frac{1}{n} \log N_d(f, r, n, x, \epsilon),$$

then

$$h_{d,x}(f) := \lim_{r\to 0} h_{d,x,r}(f)$$

is called the *local entropy* of f at x.

Remark 8.3.8 The limits exist because the dependence on ϵ is increasing and on r decreasing.

Proposition 8.3.9 $h_{d,x}(f) \leq h_{top}(f)$.

Proof Topological entropy corresponds to the case of leaving ϵ fixed at a size for which $B(x, \epsilon)$ is the entire space. Therefore any point with strong sensitive dependence in this sense necessarily produces large topological entropy. □

On the other hand, there is a relation with $h_d(f, r)$ [see (8.2.2)]:

Proposition 8.3.10 *For $r > 0$ there exists an x such that*

$$h_{d,x,r}(f) \geq h_d(f, r).$$

Proof If $S_d(f, r, n, x, \epsilon)$ is the minimal number of d_n^f-r-balls covering $B_d(x, \epsilon)$, then there is an x such that

$$S_d(f, r, n) \leq S_d(\epsilon) S_d(f, r, n, x, \epsilon) \tag{8.3.5}$$

because we can take a cover of the space by $S_d(\epsilon)$-balls of radius ϵ and, denoting their centers by x_j, we have

$$S_d(f, r, n) \leq \sum_{j=1}^{S_d(\epsilon)} S_d(f, r, n, x_j, \epsilon);$$

hence

$$S_d(f, r, n, x, \epsilon) \geq S_d(f, r, n)/S_d(\epsilon)$$

for x being one of the x_j.

As $n \to \infty$ we obtain a sequence of such points x_n satisfying (8.3.2) for the respective values of n. Take an accumulation point x of this sequence and consider the 2ϵ-ball around it. For sufficiently large n we have $B_d(x_n, \epsilon) \subset B_d(x, 2\epsilon)$ and hence

$$S_d(f, r, n, x, 2\epsilon) \geq S_d(f, r, n)/S_d(\epsilon)$$

for all n, which implies

$$\overline{\lim_{n\to\infty}} \frac{1}{n} \log S_d(f, r, n, x, 2\epsilon) \geq \overline{\lim_{n\to\infty}} \frac{1}{n} \log)(S_d(f, r, n)/S_d(\epsilon)) = h_d(f, r).$$

Using arguments as before we can replace $S_d(f, r, n, x, 2\epsilon)$ by the corresponding number of r-dense points and let $\epsilon \to 0$ to get

$$h_{d,x,r}(f) \geq h_d(f, r)$$

for all r. □

Remark 8.3.11 Since $h_d(f, r) \xrightarrow[r\to 0]{} h(f)$, this shows that there are points with $h_{d,x,r}(f)$ arbitrarily close to the topological entropy. Thus the supremum of local entropies over the space is the topological entropy, and topological entropy indeed measures the amount of sensitive dependence on initial conditions.

EXERCISES

Exercise 8.3.1 Prove Corollary 8.3.2 without reference to topological conjugacy.

Exercise 8.3.2 Construct a map with positive topological entropy that has no periodic points.

Exercise 8.3.3 Construct a topologically transitive map of a compact metric space that has infinite topological entropy.

Exercise 8.3.4 Prove that the local entropy of E_m is independent of the point and equals topological entropy.

Exercise 8.3.5 Prove that the local entropy of the shift on m symbols is independent of the point and equals topological entropy.

Exercise 8.3.6 Prove that the local entropy of the toral automorphism induced by $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ is independent of the point and equals topological entropy.

Exercise 8.3.7 Consider the closed unit disk in $\mathbb{R}^2$ and the map f_λ on it defined in polar coordinates by $f_\lambda(re^{i\theta}) = \lambda r e^{2i\theta}$, where $0 \leq \lambda \leq 1$. Show that $h_{\text{top}}(f_1) \geq \log 2$ and that $h_{\text{top}}(f_\lambda) = 0$ for $\lambda < 1$.

PROBLEMS FOR FURTHER STUDY

Problem 8.3.8 Prove that $h_{\text{top}}(\varphi^t) = |t| h_{\text{top}}(\varphi^1)$ for any flow φ^t.

Problem 8.3.9 Give an example of a topologically transitive map for which local entropy is not constant.