§7. Dimension and Entropy.

This section will first describe several possible definitions for the concept of "dimension", and then use related methods to define the "topological entropy" of a compact dynamical system.

§7A. Box Dimension. There are several possible real valued invariants for a compact metric space X which can be called the "dimension" of X. Let us first describe two dimension functions which measure how efficiently X can be covered by balls (or boxes) of equal size. Following Falconer, I will call these the upper and lower "box-counting dimension", or briefly *box dimension*, of X. The presentation will depend on some elementary inequalities, which will also be needed in §7C section on topological entropy.

Remarks. This concept of metric dimension has a long history, and has been studied under a variety of names. The oldest form of the definition, based on studying the volume of the ϵ-neighborhood of X as a function of ϵ, was introduced by Bouligand, generalizing earlier work by Steiner and Minkowski. (See Problem 7-d below.) More modern forms of the definition were introduced by Pontryagin and Schnirelman, and later by Kolmogorov and Tihomirov. (Compare Mandelbrot §39.) It should not be confused with Hausdorff dimension or with topological dimension. The *Hausdorff dimension*, introduced by Felix Hausdorff in 1919, is also a real valued metric invariant, but measures how efficiently X can be covered by balls of varying size. It coincides with the box dimension in many special cases, but is better behaved in general. The *topological dimension*, introduced by Lebesgue, Brouwer, Menger and Urysohn, is quite different, being integer valued, and invariant under arbitrary homeomorphisms. These other concepts of dimension will be discussed in §7B. (For further information, see Edgar, and Falconer, as well as the classical book of Hurewicz and Wallman.)

Let X be compact metric. Suppose that we can only distinguish points of X to an accuracy of ϵ. How many distinct points can we distinguish? Here are three possible definitions. The distance between points of X will be written as $\mathbf{d}(x, x')$.

By the *ϵ-covering number* $C_\epsilon(X)$ will be meant the smallest number $k \geq 0$ so that X can be covered by sets $X_1, \ldots, X_k$ of diameter $\mathrm{diam}(X_i) \leq \epsilon$.

By the *ϵ-separated number* $S_\epsilon(X)$ will be meant the largest cardinality of a subset $\{x_1, \ldots, x_\ell\} \subset X$ such that $\mathbf{d}(x_h, x_j) > \epsilon$ for $h \neq j$.

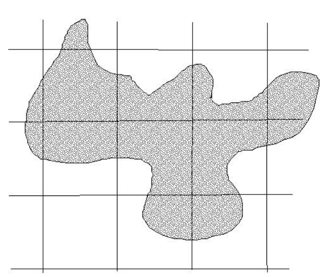

Figure 33. A set $X \subset {}^2$, covered by $B_\epsilon(X) = 13$ ϵ-boxes. This ϵ-box number $B_\epsilon(X)$ is relatively easy to compute, while the numbers $S_\epsilon(X)$ and $C_\epsilon(X)$ are more difficult to compute.

Now suppose that X is contained in the euclidean space $\mathbb{R}^n$. Then the *ϵ-box number* $B_\epsilon(X)$ is defined as follows. Cover $\mathbb{R}^n$ by countably many disjoint "half-open boxes"

$$\mathrm{box}_\epsilon(i_1, \ldots, i_n) \;=\; [i_1\epsilon, (i_1+1)\epsilon) \;\times\; \cdots \;\times\; [i_n\epsilon, (i_n+1)\epsilon)$$

of edge length ϵ, indexed by n-tuples of integers. Define $B_\epsilon(X)$ to be the number of these boxes $\mathrm{box}_\epsilon(i_1, \ldots, i_n)$ which intersect X. Evidently this is a quantity which is particularly well adapted to computer calculations.

An easy compactness argument shows that the number $C_\epsilon(X)$ is always finite. The number $S_\epsilon(X)$ is also finite, with

$$C_{2\epsilon}(X) \;\le\; S_\epsilon(X) \;\le\; C_\epsilon(X)\,. \tag{7:1}$$

For given sets $X_1 \cup \cdots \cup X_k = X$ and $\{x_1, \ldots, x_\ell\} \subset X$ as above, we have $\ell \le k$ since each X_i can contain at most one x_j, and the left hand inequality is true since the closed ϵ-neighborhoods of the x_j are sets of diameter $\le 2\epsilon$ covering X. Similarly, note that

$$C_{\epsilon\sqrt{n}}(X) \;\le\; B_\epsilon(X) \;\le 2^n\, C_\epsilon(X)\,. \tag{7:2}$$

The left hand inequality is true since each such ϵ-box has diameter $\epsilon\sqrt{n}$, and the right hand inequality since a set of diameter $\le \epsilon$ can intersect at most 2^n such ϵ-boxes.

How rapidly do the numbers $C_\epsilon(X)$ or $S_\epsilon(X)$ or $B_\epsilon(X)$ grow as ϵ tends to zero? For many interesting examples, these numbers grow roughly like a power of $1/\epsilon$. In other words, the logarithms $\log C_\epsilon(X) \approx \log S_\epsilon(X) \approx \log B_\epsilon(X)$ are roughly linear functions of $\log(1/\epsilon)$, so that the ratios

$$\frac{\log C_\epsilon(X)}{\log(1/\epsilon)} \;\approx\; \frac{\log S_\epsilon(X)}{\log(1/\epsilon)} \;\approx\; \frac{\log B_\epsilon(X)}{\log(1/\epsilon)}$$

converge to a common limit as $\epsilon \to 0$. This limit does not always exist. (See Lemma 7.2 below. ????) However we can always consider the *lim inf*, which we write as

$$\dim_B^-(X) \;=\; \liminf_{\epsilon\to 0} \frac{\log C_\epsilon(X)}{\log(1/\epsilon)} \;=\; \liminf_{\epsilon\to 0} \frac{\log S_\epsilon(X)}{\log(1/\epsilon)}\;,$$

as well as the *lim sup*

$$\dim_B^+(X) \;=\; \limsup_{\epsilon\to 0} \frac{\log C_\epsilon(X)}{\log(1/\epsilon)} \;=\; \limsup_{\epsilon\to 0} \frac{\log S_\epsilon(X)}{\log(1/\epsilon)}\;.$$

Here the equalities on the right follow easily from formula (7:1) above. Similarly, if X is contained in the Euclidean space $\mathbb{R}^n$, then it follows from (7:2) that

$$\dim_B^-(X) \;=\; \liminf_{\epsilon\to 0} \frac{\log B_\epsilon(X)}{\log(1/\epsilon)}\quad,\qquad \dim_B^+(X) \;=\; \limsup_{\epsilon\to 0} \frac{\log B_\epsilon(X)}{\log(1/\epsilon)}\;.$$

Furthermore, in this case it is easy to check that $\dim_B^+(X) \le n$. In all cases, note that

$$0 \;\le\; \dim_B^-(X) \;\le\; \dim_B^+(X) \;\le\; \infty$$

provided that $X \ne \emptyset$.

Definition. Whether or not X is contained in $\mathbb{R}^n$, we will call $\dim_B^-(X)$ the *lower box dimension* and $\dim_B^+(X)$ the *upper box dimension* of X. If $\dim_B^-(X) = \dim_B^+(X)$, then we will call this common value simply the *box dimension* $\dim_B(X)$.

(For a well defined invariant intermediate between the upper and lower box dimensions, see Problem 7-e.) In terms of information theory, the number $\log B_\epsilon(X)$ can be described as the quantity of *information* needed to specify the coordinates a point of X to within an accuracy of ϵ . If we use logarithms to the base two, so as to measure information in *bits*, and set $\epsilon = 1/2^k$, then the statement that $\log_2 B_\epsilon(X)/\log_2(1/\epsilon) \approx \dim_B(X)$, means that we need roughly $k \dim_B(X)$ bits of information in order to specify a point of X to an accuracy of $1/2^k$.

---

§**7B. Other Definitions of Dimension.** The box dimension is a rather crude invariant, well suited to empirical studies. The Hausdorff dimension is a more sophisticated invariant. For our purposes, it can be defined as follows.

Definition. The *Hausdorff dimension* $\dim_H(X)$ of a (not necessarily compact) metric space X is the infimum of real numbers $\delta \geq 0$ with the following property: *For any* $\epsilon > 0$ *,* X *can be covered by at most countably many sets* $\{A_i\}$ *with* $\operatorname{diam}(A_i) < \epsilon$ *and with*

$$\sum_i \operatorname{diam}(A_i)^\delta \ < \ \epsilon \ .$$

If there is no such number δ , then by definition $\dim_H(X) = +\infty$.

Closely associated is the concept of Hausdorff measure. For an arbitrary subset $S \subset X$ and an arbitrary real number $\delta \geq 0$, the δ-dimensional outer *Hausdorff measure* $\eta_\delta(S)$ is defined by the formula

$$\eta_\delta(S) \ = \ \lim_{\epsilon \to 0} \ \inf \left\{ \sum_i \operatorname{diam}(A_i)^\delta \right\} \ \in \ [0, \infty] \ .$$

Here $\{A_i\}$ is to range over all finite or countably infinite coverings of S by sets of diameter less than ϵ . It is not difficult to check that $\eta_\delta(X) = 0$ whenever $\delta > \dim_H(X)$, and that $\eta_\delta(X) = \infty$ whenever $\delta < \dim_H(X)$. Thus this concept of measure is possibly non-trivial only when δ is precisely equal to the Hausdorff dimension of X .

Definition. Following Lebesgue, the *topological dimension* $\dim_{\rm top}(X)$ can be defined as the smallest integer n such that X can be covered by arbitrarily small open sets with the property that no point belongs to more than $n+1$ of these sets. (This is sometimes called the *covering dimension*, to distinguish it from an alternative inductive definition of topological dimension, due to Menger and Urysohn. However, these two definitions of topological dimension always agree, provided that X is metrizable with a countable dense subset. Compare [Hurewicz-Wallman].)

As an example, if X is an n-dimensional simplicial complex, then the "open star neighborhoods" of the vertices form a covering with this property, and by subdividing the complex we can make this covering arbitrarily fine. In fact the topological dimension is precisely equal to n in this case. (Compare Problem 7-g, as well as 7.1 below.)

Note that all four definitions of dimension clearly have the following monotonicity property:

$$X \ \subset \ Y \qquad \Longrightarrow \qquad \dim(X) \ \leq \ \dim(Y) \ . \tag{7:3}$$

The relationship between the four definitions can be described as follows.

Lemma 7.1. *The inequalities*

$$0 \;\le\; \dim_{\,\mathrm{top}}(X) \;\le\; \dim_H(X) \;\le\; \dim_B^-(X) \;\le\; \dim_B^+(X) \;\le\; \infty$$

are valid whenever they make sense. Furthermore, if X *is a compact region in Euclidean* n*-space, then all four dimensions take the value* n *.*

In fact the inequality $\dim_H(X) \le \dim_B^-(X)$ follows easily from the definitions. (Compare Problem 7-h below.) The proof that $\dim_{\,\mathrm{top}}(X) \le \dim_H(X)$ is more difficult, and will not be given here. It is due to Nöbeling for subsets of Euclidean space and to Szpilrajn in general. (See [Hurewicz and Wallman] for further information.) For the proof that $\dim_{\,\mathrm{top}}(X) \ge n$ when X contains an open subset of ^{n} , see Problem 7-g. The rest of the proof is straightforward. □

If X can be described as the union of finitely many smooth compact pieces, then it is not hard to show that these four definitions of dimension all coincide. Thus, whenever

$$\dim_{\,\mathrm{top}}(X) \;<\; \dim_B^+(X) \;,$$

the space X cannot be such a finite union. Following Mandelbrot, it is then described as a "*fractal*". If we want to exclude countable unions of smooth pieces, then we need the sharper restriction that

$$\dim_{\,\mathrm{top}}(X) \;<\; \dim_H(X) \;.$$

However, there do exist interesting examples, such as Cantor sets with $\dim_B^+(X) = 0$, which seem quite wild and yet would not be fractal by any such test. For such reasons, Mandelbrot suggests that the term "fractal" is best left without any precise definition.

For fractal sets which occur in dynamics, there is sometimes enough similarity between different regions and between the same region at different scales to guarantee that

$$\dim_H \;=\; \dim_B^- \;=\; \dim_B^+ \;.$$

(See for example [Bishop and Jones], [Pzrytyki].) General theorems of this nature are rare. However, here is one simple but non-trivial example.

Example 7.2. A Generalization of the Cantor Middle Third Set. Let I be a closed interval of real numbers, and let

$$\phi_0\,,\,\phi_1 \;:\; I \;\to\; I$$

be two linear embeddings of I into itself with disjoint images $I_j = \phi_j(I) \subset I$. Then for each finite sequence $j_1\,,\,\ldots\,,\,j_m$ of zeros and ones we can form the image

$$I_{j_1\cdots j_m} \;=\; \phi_{j_1}\circ\cdots\circ\phi_{j_m}(I) \;\subset\; I_{j_1\cdots j_{m-1}} \;.$$

If $a_j = |\phi_j'| = \ell(I_j)/\ell(I)$ is the length ratio, note that the length

$$\ell(I_{j_1\cdots j_m}) \;=\; a_{j_1}\cdots a_{j_m}\,\ell(I)$$

tends uniformly to zero as $m\to\infty$. Hence, to any infinite sequence $j_1\,,\,j_2\,,\,j_3\,,\,\ldots$ of zeros and ones there corresponds a single point $x(j_1\,,\,j_2\,,\,\cdots)$, where

$$\{x(j_1\,,\,j_2\,,\,\cdots)\} \;=\; \bigcap_{m\ge1} I_{j_1\cdots j_m} \;.$$

The compact totally disconnected set consisting of all of these points $x(j_1, j_2, \cdots)$ is the required generalized middle third set $K = K(\phi_0, \phi_1)$. It is not difficult to check that the topological dimension $\dim_{\rm top}(K)$ is zero.

Equivalently, we can consider an inverse map

$$\psi : I_0 \cup I_1 \to I ,$$

where ψ coincides with ϕ_0^{-1} on I_0, and with ϕ_1^{-1} on I_1. Then K can be identified with the set of all points $x \in I_0 \cup I_1$ such that the m-fold composition $\psi^{\circ m}(x)$ is defined and belongs to $I_0 \cup I_1$ for every $m \geq 1$.

Since the lenth ratios $a_j = |\phi_j'|$ satisfy $a_j > 0$ with $a_0 + a_1 < 1$, it follows easily that there is a unique number in the interval $0 < \delta < 1$ which satisfies the equation

$$a_0^\delta + a_1^\delta = 1 . \tag{7:4}$$

We will prove the following.

> **Lemma 7.3.** *The box dimension* $\dim_B(K)$ *is well defined and equal to the Hausdorff dimension* $\dim_H(K)$ *, with*
>
> $$0 < \dim_B(K) = \dim_H(K) = \delta < 1 ,$$
>
> *where* δ *is defined by equation* (7 : 4).

As an example, choosing $a_0 = a_1$ to be any number in the open interval $(0, 1/2)$, we get

$$\dim_H(K) = \dim_B(K) = \delta = \log(2)/\log(1/a_j) ,$$

which can take any value strictly between 0 and 1. For $a_0 = a_1 = 1/3$, we obtain the classical Cantor middle third set, with $\dim_H(K) = \dim_B(K) = \log(2)/\log(3) = 0.6309\cdots$.

Proof of 7.3. Without loss of generality, we may assume that I is the smallest closed interval which contains K. (The endpoints of this minimal interval I are just the fixed points of ϕ_0 and ϕ_1.)

We will prove that the Hausdorff outer measure of K is given by

$$\eta_\delta(K) = \ell(I)^\delta , \tag{7 : 5}$$

and more generally $\eta_\delta(K \cap I_\sigma) = \ell(I_\sigma)^\delta$, where $\sigma = (j_1, \ldots, j_m)$ is any finite sequence of bits. In particular, this will show that $0 < \eta_\delta(K) < \infty$, and hence prove that δ is precisely the Hausdorff dimension of K.

I will call each $I_{j_1 \cdots j_m}$ an interval of *level* m. Clearly it suffices to consider coverings of K by open or closed intervals J. For any interval J, we will prove that

$$\ell(J)^\delta \geq \sum_\sigma \ell(J \cap I_\sigma)^\delta , \qquad \text{with equality when} \qquad J = I , \tag{7 : 6}$$

where I_σ ranges over the 2^m intervals of level m.

First consider the case $m = 1$. Without loss of generality, we may replace J by $J \cap I$, and hence assume that $J \subset I$. Furthermore, we may assume that J intersects both I_0 and I_1, since otherwise the required inequality

$$\ell(J)^\delta \geq \ell(J \cap I_0)^\delta + \ell(J \cap I_1)^\delta$$

would be clear.

If $J = [a, b]$ with $a < b$, then we can compute the partial derivative

$$\partial\,\ell(J)^\delta/\partial b \;=\; \partial(b-a)^\delta/\partial b \;=\; \delta/\ell(J)^{1-\delta} \;>\; 0.$$

Note that this is monotone decreasing as a function of $\ell(J)$. Similarly the partial derivative $\partial\ell(J)^\delta/\partial a \;=\; -\delta/(\ell(J)^{1-\delta} < 0$ is monotone increasing as a function of $\ell(J)$. It follows easily that the difference

$$\ell(J)^\delta \,-\, \ell(J\cap I_0)^\delta \,-\, \ell(J\cap I_1)^\delta$$

decreases monotonically as the endpoint b increases through I_1. Similarly, it decreases monotonically as the endpoint a *decreases* through I_0. Thus this difference attains its minimum value precisely when $J = I$. Since this minimum value is zero by (7 : 4), this proves (7 : 6) for $m = 1$.

For any finite bit sequence σ, a similar argument shows that

$$\ell(J\cap I_\sigma)^\delta \;\geq\; \ell(J\cap I_{\sigma 0}) \,+\, \ell(J\cap I_{\sigma 1})\;.$$

The conclusion (7 : 6) then follows by a straightforward induction.

Given $\epsilon > 0$ we can cover K by finitely many arbitrarily small open intervals J_h so that

$$\sum_{J_h}\ell(J_h)^\delta \;<\; \eta_\delta(K) \,+\, \epsilon.$$

Without loss of generality, we may assume the each J_h has both endpoints outside of the set K, and hence outside of all intervals I_σ of sufficiently high level m. Thus, for m large, each $J_h \cap I_\sigma$ is either equal to I_σ or empty. Considering only the 2^m intervals I_σ of some fixed high level m, it then follows from (7 : 6) that

$$\sum_{J_h}\ell(J_h)^\delta \;\geq\; \sum_{J_h}\;\sum_{I_\sigma\subset J_h}\ell(I_\sigma)^\delta \;\geq\; \sum_{I_\sigma}\ell(I_\sigma)^\delta \;=\; \ell(I)^\delta\;.$$

Since ϵ can be arbitrarily small, this proves that $\eta_\delta(K) \geq \ell(I)^\delta$. On the other hand, the upper bound $\eta_\delta(K) \leq \sum \ell(I_\sigma)^\delta = \ell(I)^\delta$ can be obtained by covering K by the intervals I_σ of some arbitrarily high level m. This completes the proof that $\dim_H(K) = \delta$.

In order to compute the box dimension of K, choose some small mesh ϵ, and consider all intervals $I_{j_1\cdots j_m}$ such that

$$\ell(I_{j_1\cdots j_m}) \;<\; \epsilon \;\leq\; \ell(I_{j_1\cdots j_{m-1}})\;.$$

Evidently these form a covering of K by finitely many intervals of length satisfying

$$\epsilon/c \;\leq\; \ell(I_{j_1\cdots j_m}) \;<\; \epsilon\;.$$

where c is the larger of a_0 and a_1. Thus the Hausdorff measure $\eta_\delta(I_{j_1\cdots j_m})$ of each such interval lies between $(\epsilon/c)^\delta$ and ϵ^δ. However, the sum of $\eta_\delta(I_{j_1\cdots j_m})$ over all such intervals is equal to $\eta_\delta(I) = \ell(I)^\delta$. Therefore the number N of such intervals satisfies

$$(\epsilon/c)^\delta N \;\leq\; \ell(I)^\delta \;<\; \epsilon^\delta N\;.$$

In particular, the ϵ-covering number satisfies

$$C_\epsilon(K) \;\leq\; N \;\leq\; c'/\epsilon^\delta$$

where $c' = c\,\ell(I)$ is constant. It follows immediately that $\dim_B^+(K) \le \delta$. Since

$$\delta \;=\; \dim_H(K) \;\le\; \dim_B^-(K) \;\le\; \dim_B^+(K) \;,$$

this completes the proof of 7.3. $\square$

However, it is certainly not true that these various definitions of dimension always coincide. For an example with $\dim_H(X) < \dim_B(X)$ see Problem 7-i below. The following result emphasizes the difference between the upper and lower box dimensions. Let K be any infinite compact metrizable space with $\dim_{\rm top}(X) = d < \infty$. Note that the set $C^0(K, \mathbb{R}^n)$ consisting of all continuous maps from K to $\mathbb{R}^n$ has a natural topology, associated with the "uniform metric" $\mathbf{d}(f,g) = \max_{\pi x \in K} \mathbf{d}(f(x)\,,\, g(x))$. (Compare §8E.)

Lemma 7.4. *If* $n \ge 2d+1$ *, then a generic mapping* $f : X \to \mathbb{R}^n$ *embeds* K *homeomorphically as a subset* $f(K) = X \subset \mathbb{R}^n$ *which satisfies*

$$\dim_{\rm top}(X) \;=\; \dim_H(X) \;=\; \dim_B^-(X) \;=\; d$$

but $\dim_B^+(X) \;=\; n$ *.*

It follows from 7.1 and 7.4 that the topological dimension can be defined as the minimum over all compatible metrics of the values of $\dim_H$ or of $\dim_B^-$.

Proof of 7.4. Start with any number $k > 0$ and form the set V_k consisting of all maps $f : K \to \mathbb{R}^n$ such that, for some $\epsilon < 1/k$, we have

$$\frac{\log S_\epsilon(f(K))}{\log(1/\epsilon)} \;\ge\; n - \frac{1}{k}\;. \tag{7:7}$$

Then it is easy to check that V_k is an open subset of $C^0(K, \mathbb{R}^n)$. We claim that V_k is also a dense subset. In fact, given any neighborhood of a map $f_0 : K \to \mathbb{R}^n$, choose any sequence of p^n distinct points $x_i \in K$ which map close to some single point of $\mathbb{R}^n$. Now deform f_0 to a map f which carries the x_i to the lattice points of a very small $p \times p \times \cdots \times p$ grid. If 2ϵ is the distance between adjacent grid points, then $\log S_\epsilon(F(K)) \ge n \log p$. More precisely, let us choose the grid size so that

$$\epsilon \;=\; p^{-n/(n-k^{-1})}\;.$$

Then $n\,\log(p) \;=\; (n-k^{-1})\,\log(1/\epsilon)$, so the required inequality (7:7) will certainly be satisfied. We can carry out this construction so that the distance from f to f_0 is at most a constant times the product $p\,\epsilon = p^{-1/(nk-1)}$. Evidently we can make this product as small as we wish by choosing p large. Thus each V_k is dense and open. If f belongs to the intersection of the V_k , it follows easily that the upper box dimension $\dim_B^+(f(K))$ is equal to n .

Now let U_k be the set of all $f : K \to \mathbb{R}^n$ such that

$$\frac{\log S_\epsilon(f(K))}{\log(1/\epsilon)} \;<\; d + \frac{1}{k} \tag{7:8}$$

for some $\epsilon < 1/k$, where d is the topological dimension. We will show that this is a dense open subset of $C^0(K, \mathbb{R}^n)$. In fact, if $f_0 \in U_k$ and if $\mathbf{d}(f\,,\,f_0) < \eta/2$, note that

$$S_{\epsilon+\eta}(f) \;\le\; S_\epsilon(f_0)\;.$$

If η is sufficiently small, it follows that $f \in U_k$. To prove density, given any $f_0 \in C^0(K, \ ^n)$, choose a covering of K by small open sets $W_1 , \dots , W_p$ and choose base points $b_i \in \ ^n$ so that each b_i is close to the image $f_0(W_i)$. Let $1 = \phi_1 + \cdots + \phi_p$ be a partition of unity, where each $\phi_i : K \to$ vanishes outside of W_i . Then the function

$$f(x) \ = \ \phi_1(x)\, b_1 + \cdots + \phi_p(x)\, b_p$$

approximates f_0 and maps K into a d-dimensional simplicial complex which is linearly immersed in $\ ^n$. Thus $\dim_B^{+}(f(K)) \le d$, and it follows that $f(K)$ belongs to U_k for every k . Finally, for any f in the intersection of the U_k , we clearly have $\dim_B^{-}(K) \le d$. Since $n \ge 2d+1$, a similar argument show that a generic map from K to $\ ^n$ is an embedding. (Compare [Hurewicz and Wallman].) Together with 7.1, this completes the proof. □

Remarks. Presumably, there also exist embeddings with $\dim_H(f(K)) = n$. In many cases, there also exists an embedding with $\dim_B^{+}(f(K)) = d$. (Compare Problem 7-f.) However, I don't know whether such an embedding always exists.

Further development of these concepts will be formulated as a series of exercises at the end of this section.

---

§**7C. Topological Entropy.** This section will define and study the *topological entropy* of a compact dynamical system (X, f) . To begin the exposition, we assume that X is a metric space, although we will see later that the metric is not really needed. The distance between two points $x, y \in X$ will be written as $\mathbf{d}(x, y)$.

Fix some integer $\ell \ge 1$, and define the ℓ*-shadowing metric* $\mathbf{d}_\ell$ on X as follows. Given x , $y \in X$, let us follow the orbits

$$x = x_0 \mapsto x_1 \mapsto x_2 \mapsto \cdots \mapsto x_{\ell-1} \quad \text{and} \quad y = y_0 \mapsto y_1 \mapsto y_2 \mapsto \cdots \mapsto y_{\ell-1}$$

under f through $\ell - 1$ iterations, and set

$$\mathbf{d}_\ell(x, y) \ = \ \max_{0 \le i < \ell} \mathbf{d}(x_i \, , \, y_i) \ .$$

Thus two points of X are ϵ-close in this new metric if and only if the orbits of x and y remain ϵ-close in the original metric $\mathbf{d} = \mathbf{d}_1$ through ℓ successive points of the orbit. We will sometimes use the more precise notation $\mathbf{d}_\ell^f$ to emphasize the dependence of this new metric on the mapping f .

As in the §7A, let us suppose that points of X can only be distinguished to accuracy $\epsilon > 0$. We ask: how many distinct orbits can we distinguish if we follow the orbits for time ℓ ? Using notations from the §7A, this means that we want to compute the ϵ-covering number $C_\epsilon(X \, , \, \mathbf{d}_\ell)$ or the ϵ-separated number $S_\epsilon(X \, , \, \mathbf{d}_\ell)$ for this new metric space $(X \, , \, \mathbf{d}_\ell)$. We always assume that X is non-vacuous, so that these numbers are strictly positive. In order to study asymptotic behavior as $\ell \to \infty$, we first note the following.

Lemma 7.5. *The inequality*

$$C_\epsilon(X \, , \, \mathbf{d}_{k+\ell}) \ \le \ C_\epsilon(X \, , \, \mathbf{d}_k)\, C_\epsilon(X \, , \, \mathbf{d}_\ell)$$

is satisfied for all integers k , $\ell \ge 1$.

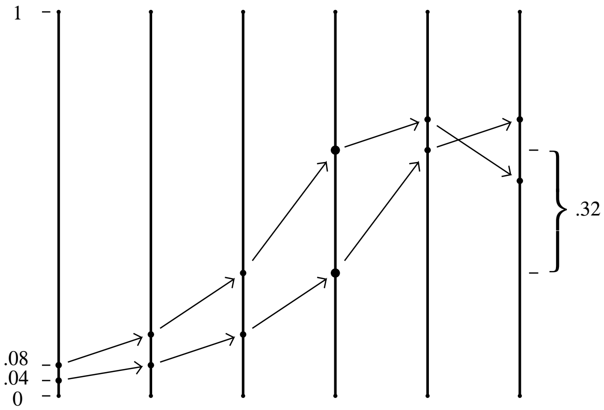

Figure 34. An example: For the tent map $f(x) = \min(2x\,,\,2-2x)$ *on the unit interval, the 6-shadowing distance* $\mathbf{d}_6(.04\,,\,.08)$ *is equal to* $.32$ *. (Six copies of the unit interval are shown.)*

For if $X = A_1 \cup \ldots \cup A_p$ is a covering of X by p subsets of $\mathbf{d}_k$-diameter $\le \epsilon$ and $X = B_1 \cup \cdots \cup B_q$ is a covering by q subsets of $\mathbf{d}_\ell$-diameter $\le \epsilon$, then

$$X \;=\; \bigcup_{i=1}^{p}\bigcup_{j=1}^{q} A_i \cap f^{-k}B_j$$

is a covering of X by $p\,q$ subsets of $\mathbf{d}_{k+\ell}$-diameter $\le \epsilon$. □

Setting $a_k = \log C_\epsilon(X\,,\,\mathbf{d}_k)$, it follows that the sequence $\{a_k\}$ is *sub-additive*, that is

$$a_{k+\ell} \;\le\; a_k + a_\ell \;.$$

Lemma 7.6. *For any sub-additive sequence* $\{a_k\}$ *of real numbers, the ratios* a_k/k *tend to a finite or negative-infinite limit as* $k\to\infty$ *. Furthermore, this limit is equal to the infimum of the numbers* a_k/k *.*

Proof. Using sub-additivity, we see by a straightforward induction on k that $a_k \le k\,a_1$. Similarly, for any fixed m we have $a_{m\,k} \le k\,a_m$. It follows that

$$a_{mk+i} \;\;\le\;\; a_{mk} + a_i \;\;\le\;\; k\,a_m \,+\, a_i \;\;.$$

Still fixing m , note that any positive integer n can be written uniquely as $n = mk + i$ with $0 \le i < m$. Thus

$$\frac{a_n}{n} \;\;=\;\; \frac{a_{mk+i}}{n} \;\;\le\;\; k\,\frac{a_m}{n} \;+\; \frac{a_i}{n} \;\;.$$

Now let us take the lim sup of both sides as n tends to infinity with m fixed. Note that the ratio k/n converges to $1/m$, while a_i/n converges to zero. Hence, in the limit, we have

$$\limsup_{n\to\infty} \frac{a_n}{n} \;\le\; \frac{a_m}{m} \;.$$

Taking the infimum over m , it follows that

$$\limsup_{n\to\infty} \frac{a_n}{n} \;\;\le\;\; \inf_{m\ge1} \frac{a_m}{m} \;\;\le\;\; \liminf_{m\to\infty} \frac{a_m}{m} \;\;\le\;\; \limsup_{m\to\infty} \frac{a_m}{m} \;\;.$$

Thus these three quantitites are all equal, which completes the proof. $\square$

In particular, taking $a_k = \log C_\epsilon(X\,,\,\mathbf{d}_k)$, it follows that the ratios $\log C_\epsilon(X\,,\,\mathbf{d}_k)/k$ necessarily converge to a finite limit. Now define $h(f,\epsilon)$ to be this limit:

$$h(f,\epsilon) \;=\; \lim_{k\to\infty} \frac{\log C_\epsilon(X\,,\,\mathbf{d}_k)}{k} \;=\; \inf_{k\ge 1} \left\{ \frac{\log C_\epsilon(X\,,\,\mathbf{d}_k)}{k} \right\} \quad .$$

Since X is non-vacuous, we have $0 \le h(f,\epsilon) \le \log C_\epsilon(X\,,\,\mathbf{d}_1)$. Note that $h(f,\epsilon) > 0$ if and only if the numbers $C_\epsilon(X\,,\,\mathbf{d}_k)$ grow at least exponentially with k . If we think of $\log C_\epsilon(X)$ as the *quantity of information* needed to specify a point of X to accuracy ϵ , then $h(f,\epsilon)$ can be described as the *average quantity of information per timestep* which is needed to specify a long orbit under f to accuracy ϵ .

As ϵ tends to zero, note that $h(f,\epsilon)$ cannot decrease. Hence the finite or infinite limit

$$\lim_{\epsilon\to 0} h(f,\epsilon) \;=\; \sup_{\epsilon>0} h(f,\epsilon)$$

necessarily exists. By definition, this limit is called the *topological entropy*

$$h(f) \quad=\quad h_{\rm top}(f) \quad\in\quad [0\,,\,\infty] \quad .$$

It is of course essential that we first take the limit as $k\to\infty$, and only then take the limit as $\epsilon\to 0$.

Remark. We can also define topological entropy using the " ϵ-separated numbers" $S_\epsilon(X\,,\,\mathbf{d}_k)$ of the previous section. I don't know whether the ratios $\log S_\epsilon(X\,,\,\mathbf{d}_k)/k$ necessarily converge as $k\to\infty$. However, using the inequality (1) of §D we certainly have

$$h(f,\epsilon) \quad\le\quad \liminf_{k\to\infty} \frac{\log S_\epsilon(X\,,\,\mathbf{d}_k)}{k} \quad\le\quad \limsup_{k\to\infty} \frac{\log S_\epsilon(X\,,\,\mathbf{d}_k)}{k} \quad\le\quad h(f\,,\,\epsilon/2)$$

and it follows that

$$\lim_{\epsilon\to 0}\, \liminf_{k\to\infty} \frac{\log S_\epsilon(X\,,\,\mathbf{d}_k)}{k} \;=\; \lim_{\epsilon\to 0}\, \limsup_{k\to\infty} \frac{\log S_\epsilon(X\,,\,\mathbf{d}_k)}{k} \;=\; h(f) \quad .$$

An Easy Example. For the doubling map $m_2(x) \;=\; 2\,x \bmod 1$ on the circle $T \;=\; .\;/\;.$, in order to specify an orbit through k timesteps to an accuracy of ϵ , we must specify the initial point to an accuracy of $\epsilon/2^k$. Hence $C_\epsilon(T_k) \approx 2^k/\epsilon$, and it follows that

$$h(m_2,\epsilon) \quad=\quad \lim_{k\to\infty} \frac{\log(2^k/\epsilon)}{k} \quad=\quad \log 2 \quad .$$

Therefore $h(m_2) = \log 2$. (In this example, it is noteworthy that we do not need to pass to the limit as $\epsilon\to 0$. Any small ϵ will do.) Similarly, for the p-tupling map $m_p(x) = p\,x \bmod 1$ on the circle, we get

$$h(m_p) \;=\; h(m_p,\epsilon) \;=\; \log p$$

for any small ϵ .

Here are four basic properties of topological entropy

Lemma 7.7:
(**a**) *If there is a topological semi-conjugacy from* (X,f) *onto* (Y,g) *, then*

$h(f) \ge h(g)$. It follows in particular that h is a topological invariant, independent of the choice of metric.

(**b**) *Similarly, if $Y \subset X$ with $f(Y) \subset Y$, then $h(f) \ge h(f|Y)$.*

(**c**) *The entropy of a cartesian product is given by*

$$h_{\rm top}(f \times g) \quad = \quad h_{\rm top}(f) + h_{\rm top}(g) \quad .$$

(**d**) *The entropy of the ℓ-fold iterate of a map from X onto itself is given by*

$$h_{\rm top}(f^{\circ \ell}) \quad = \quad \ell \, h_{\rm top}(f)$$

for any $\ell \ge 0$. Furthermore, if f is a homeomorphism from X onto itself, then $h_{\rm top}(f^{-1}) = h_{\rm top}(f)$.

Proof (a): Recall that a *semi-conjugacy* ϕ from the dynamical system (X, f) onto (Y, g) is a continuous map from X onto Y which satisfies $\phi \circ f = g \circ \phi$, so that the diagram

$$\begin{array}{ccc} X & \xrightarrow{f} & X \\ \phi \downarrow & & \phi \downarrow \\ Y & \xrightarrow{g} & Y \end{array}$$

is commutative. If there exists such a semi-conjugacy ϕ from the compact metric dynamical systems (X, f) onto the compact metric dynamical system (Y, g) , then by uniform continuity, given $\epsilon > 0$ we can find $\delta > 0$ so that

$$\mathbf{d}_X(x, x') < \delta \qquad \Longrightarrow \qquad \mathbf{d}_Y(\phi(x), \phi(x')) < \epsilon \quad .$$

It follows easily that

$$C_\delta(X \, , \mathbf{d}_\ell) \quad \ge \quad C_\epsilon(Y \, , \mathbf{d}_\ell) \qquad \text{hence} \qquad h(f, \delta) \; \ge \; h(g, \epsilon) \quad .$$

Taking the limit as both ϵ and δ tend to zero, it follows that $h(f) \ge h(g)$. Now if (X, f) is topologically conjugate to (Y, g) , then there exist semi-conjugacies in both directions, hence $h(f) = h(g)$. Thus topological entropy is indeed a topological invariant.

(b): This is clear.

(c): As in Problem D-3 of §D, we can use the maximum metric

$$\mathbf{d}((x, y) \, , (x', y')) \quad = \quad \max\left(\mathbf{d}(x, x') \, , \mathbf{d}(y, y')\right)$$

on the product $X \times Y$, and note that

$$C_\epsilon(X \times Y) \le C_\epsilon(X) \, C_\epsilon(Y) \quad , \qquad S_\epsilon(X \times Y) \ge S_\epsilon(X) \, S_\epsilon(Y) \quad .$$

The same inequalities hold for the spaces $(X \, , \mathbf{d}_\ell)$ and $(Y \, , \mathbf{d}_\ell)$ with the ℓ-shadowing metrics, and the conclusion follows.

(d): If f is a homeomorphism, then the identity $h_{\rm top}(f) = h_{\rm top}(f^{-1})$ follows easily from the definition. For the computation of $h_{\rm top}(f^{\circ \ell})$, we will use the more explicit notation $\mathbf{d}_\ell^f$ in order to indicate the dependence of the ℓ-shadowing metric both on the mapping f and on the original metric $\mathbf{d} = \mathbf{d}_1$. With this notation, suppose that we consider $f^{\circ \ell}$ as a map from the metric space $(X \, , \mathbf{d}_\ell^f)$ to itself. Evidently two points are ϵ-close for k successive points of the orbit of $f^{\circ \ell}$ under this metric if and only if they are ϵ-close for

$k\,\ell$ successive points under f for the original metric $\mathbf{d}_1$. Since we know that $h_{\rm top}$ does not depend on the choice of metric, the required identity $h_{\rm top}(f^{\circ\ell}) = \ell\, h_{\rm top}(f)$ now follows easily. □

Examples. To illustrate these statements, first consider the quadratic map $q(x) = 2x^2 - 1$ on the interval $[-1\,,\,1]$. The projection map $x + iy \mapsto x$ from the unit circle to $[-1\,,\,1]$ semi-conjugates the squaring map $s(z) = z^2$ to this quadratic map q , hence $h(s) \ge h(q)$. Since s is topologically conjugate to the doubling map on $\mathbb{R}/\mathbb{Z}$, it follows that $h(s) = \log 2$. In fact we will see later that $h(q)$ is also equal to $\log 2$.

The doubling map $f(\mathbf{x}) = 2\,\mathbf{x}$ on the n-dimensional torus $T^n = \mathbb{R}^n/\mathbb{Z}^n$ can be considered as the n-fold Cartesian product of the doubling map on the circle with itself, hence $h_{\rm top}(f) = n\,\log 2$. If we iterate this map k-times, then we get the map $\mathbf{x} \mapsto 2^k\,\mathbf{x}$ on T^n , with entropy $k\,n\log 2$. Note that there is a semi-conjugacy from the doubling map on T^n onto the doubling map on T^{n-1} .

---

§7D. A Dimension Inequality. By definition, a map f between metric spaces is *Lipschitz* if there is some constant $c \ge 0$, called a *Lipschitz constant*, so that

$$\mathbf{d}(f(x), f(y)) \quad \le c\,\mathbf{d}(x, y)$$

for every pair of points x and y in the domain of f . If f is a map from a complete metric space to itself with Lipschitz constant $c < 1$, then it is easy to show that every orbit converges to a unique attracting fixed point. Thus, for a dynamically interesting example, we must have $c \ge 1$. Recall that $\dim_B^-(X)$ denotes the lower box dimension of X .

Lemma 7.8. *If f is a map from a compact metric space X to itself with Lipschitz constant $c \ge 1$, then*

$$h_{\rm top}(f) \quad \le \quad \dim_B^-(X)\ \log c \quad .$$

As an immediate corollary, we see that any smooth map from a compact manifold to itself necessarily has finite topological entropy.

Proof of 7.8. Since any larger value of c will also serve as Lipschitz constant, it suffices to consider the case $c > 1$. Recall the definitions

$$h(f, \epsilon) \;=\; \lim_{k\to\infty} \frac{1}{k} \log C_\epsilon(X_k)$$

and

$$\dim_B^-(X) \;=\; \liminf_{\eta\to 0} \frac{\log C_\eta(X)}{\log(1/\eta)} \quad .$$

For fixed ϵ and arbitrarily small η we can choose an integer k so that

$$\epsilon/c^k \;\ge\; \eta \;\ge\; \epsilon/c^{k+1} \quad .$$

(Here we make use of the hypothesis that $c > 1$.) Note that k tends to infinity as $\eta \to 0$. Now if two points of X satisfy $\mathbf{d}(x, y) \le \eta$, then it follows that

$$\mathbf{d}(f^{\circ i}(x)\,,\, f^{\circ i}(y)) \;\le\; c^i\,\eta \;\le\; \epsilon$$

for $i \le k$, hence $\mathbf{d}_k(x,y) \le \epsilon$. Therefore

$$C_\eta(X) \quad \ge \quad C_\epsilon(X_k) \quad ,$$

and it follows that

$$\frac{\log C_\eta(X)}{\log(1/\eta)} \quad \ge \quad \frac{\log C_\epsilon(X_k)}{\log(1/\eta)} \quad \sim \quad \frac{\log C_\epsilon(X_k)}{\log(c^k/\epsilon)} \quad \sim \quad \frac{\log C_\epsilon(X_k)}{k\,\log c} \quad .$$

Now as $\eta \to 0$ and $k \to \infty$, the *lim inf* of the left hand side is equal to $\dim_B^-(X)$, while the right hand side converges to $h(f,\epsilon)/\log c$. Thus $h(f,\epsilon) \le \dim_B^-(X)\,\log c$, and taking the limit as $\epsilon \to 0$ we obtain the required inequality $h(f) \le \dim_B^-(X)\,\log c$. □

Examples. Here is an example to show that this inequality is sharp. Let d_n be the doubling map $\mathbf{x} \mapsto 2\,\mathbf{x}$ on the torus $T^n = {}^n/{}^n$. Then the dimension $\dim_B(T^n)$ is n and the best Lipschitz constant is $c = 2$. It follows from 7.7(b) that the entropy $h_{\rm top}(d_n)$ is precisely equal to the product $n\,\log 2$.

More generally, consider the n-dimensional torus $T^n = {}^n/{}^n$ and a linear map

$$f \;:\; \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \;\mapsto\; M \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \bmod {}^n$$

where M is an $n \times n$ integer matrix. Suppose, to simplify the discussion, that M has n linearly independent real eigenvalues, or equivalently that M is conjugate to a diagonal matrix,

$$A\,M\,A^{-1} \quad = \quad \mathrm{diag}(\lambda_1\,,\,\ldots\,,\,\lambda_n)$$

where A is some non-singular real matrix. Then, setting $Y = AX$, we see that f is topologically conjugate to the mapping $Y \mapsto (AMA^{-1})Y \bmod A\,{}^n$, or in other words

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \mapsto \quad \begin{pmatrix} \lambda_1 y_1 \\ \vdots \\ \lambda_n y_n \end{pmatrix} \qquad \bmod A\,{}^n \quad .$$

It will be convenient to number the eigenvalues so that

$$|\lambda_1| \;\ge\; \cdots \;\ge\; |\lambda_m| \;>\; 1 \;\ge |\lambda_{m+1}| \;\ge\; \cdots \;\ge\; |\lambda_n| \quad ,$$

so that only the first m lie outside the unit disk. Then to specify an orbit for k timesteps to an accuracy of ϵ , we must specify the initial coordinate y_i to an accuracy of $\epsilon/|\lambda_i|^k$ whenever $i \le m$. However, an accuracy of ϵ will suffice whenever $i > m$. From this, it follows that the ϵ-covering numbers are given very roughly by

$$C_\epsilon(T^n_k) \quad \approx \quad \frac{|\lambda_1|^k}{\epsilon} \cdots \frac{|\lambda_m|^k}{\epsilon}\,\frac{1}{\epsilon} \cdots \frac{1}{\epsilon} \quad = \quad \frac{|\lambda_1 \cdots \lambda_m|^k}{\epsilon^n} \quad .$$

More precisely, the ratio of the left and right hand expressions remains bounded and bounded away from zero as $k \to \infty$. In fact, some constant multiple of this number of suitably small n-dimensional rectangles clearly suffices to cover the torus, and an n-dimensional volume computation shows that no smaller number will suffice. Hence

$$h_{\rm top}(f) \quad = \quad h(f,\epsilon) \quad = \quad \log|\lambda_1 \cdots \lambda_m| \quad = \quad \sum_{|\lambda_i|>1} \log|\lambda_i|$$

for any small ϵ. (We can also write the expression on the right as $\sum_1^n \log^+|\lambda_i|$, where $\log^+(x) = \max(\log x\,,\,0)$.) Thus the eigenvalues with $|\lambda_i| \le 1$ make no contribution to topological entropy.

Remark. This same formula remains true even when M does not have n linearly independent real eigenvectors. See for example Walters, §8.4.

We can compare this exact formula with the estimate 7.8. For a suitably chosen metric, the best Lipschitz constant is equal to the spectral radius $|\lambda|_{\max}$. Hence the upper bound of 7.8 is $n \log|\lambda|_{\max}$, which is clearly greater than or equal to $h_{\rm top}(f) = \sum \log^+|\lambda_i|$.

---

§**7E. Defining Entropy by Coverings.** We next define the entropy of f with respect to some covering of X by subsets. We will be particularly interested in coverings by open subsets. However, in Chapter III we will also want to consider coverings by certain closed subsets.

Let $\mathcal{A} = \{A_\alpha\}$ be some collection of subsets with union equal to X. We define $N(\mathcal{A})$ to be the smallest number of sets from $\mathcal{A}$ which suffice to cover X. We always suppose that X is non-vacuous so that $N(\mathcal{A}) \ge 1$, and in practice we will always assume also that $N(\mathcal{A}) < \infty$. As an example, suppose that $\mathcal{C}_\epsilon$ is the collection consisting of all sets $A \subset X$ with diameter $\mathrm{diam}(A) \le \epsilon$. Thus $N(\mathcal{C}_\epsilon)$ is equal to the minimum number of sets of diameter $\le \epsilon$ needed to cover X. Evidently this coincides with the number $C_\epsilon(X)$, as defined earlier.

Fixing some map $f : X \to X$, let $\mathcal{A}^k = \mathcal{A}^k_f$ be the collection consisting of all k-fold intersections

$$A_{\alpha_0} \cap f^{-1}(A_{\alpha_1}) \cap \cdots \cap f^{-(k-1)}(A_{\alpha_{k-1}})$$

with $A_{\alpha_i} \in \mathcal{A}$. Note that a point x belongs to such an intersection if and only if its orbit $x = x_0 \mapsto x_1 \mapsto \cdots$ satisfies $x_i \in A_{\alpha_i}$ for $0 \le i < k$. We are interested in the number $N(\mathcal{A}^k)$ of such interesections needed to cover X. It is easy to check that

$$N(\mathcal{A}^{k+\ell}) \quad \le \quad N(\mathcal{A}^k)\, N(\mathcal{A}^\ell) \quad .$$

Therefore, according to 7.6, the limit

$$\lim_{k\to\infty} \frac{\log N(\mathcal{A}^k)}{k}$$

always exists and is equal to

$$\inf_{k\ge 1} \frac{\log N(\mathcal{A}^k)}{k} \quad .$$

By definition, this limit is called the *entropy* $h(f\,,\,\mathcal{A})$ of f with respect to the covering $\mathcal{A}$. Note that $0 \le h(f\,,\,\mathcal{A}) \le \log N(\mathcal{A})$.

As an example, if $\mathcal{C}_\epsilon$ is the covering of X by all sets of diameter $\le \epsilon$, then $\mathcal{A}^k = {\mathcal{C}_\epsilon}^k$ is the covering by all sets which have diameter $\le \epsilon$ in the k-shadowing metric. Hence

$$N({\mathcal{C}_\epsilon}^k) \quad = \quad C_\epsilon(X_k) \quad ,$$

and it follows that $h(f\,,\mathcal{C}_\epsilon)$ is equal to

$$\lim \frac{\log C_\epsilon(X_k)}{k} \quad = \quad h(f,\epsilon) \quad .$$

Using this fact, we obtain an alternative definition of topological entropy which is clearly topologically invariant.

Lemma 7.9 *The topological entropy* $h_{\rm top}(f) = \lim_{\epsilon\to 0} h(f,\epsilon)$ *is equal to the supremum of* $h(f\,,\mathcal{U})$ *as* $\mathcal{U}$ *ranges over all coverings of* X *by open sets.*

The proof will depend on the following.

Definition. Given coverings $\mathcal{A}$ and $\mathcal{B}$ of X, we say that $\mathcal{A}$ is a *refinement* of $\mathcal{B}$ if every set $A\in\mathcal{A}$ is contained in some set $B\in\mathcal{B}$.

If $\mathcal{A}$ is a refinement of $\mathcal{B}$, then clearly $N(\mathcal{A})\ge N(\mathcal{B})$. Furthermore, $\mathcal{A}^k$ is a refinement of $\mathcal{B}^k$, hence $N(\mathcal{A}^k)\ge N(\mathcal{B}^k)$, and it follows that $h(f\,,\mathcal{A})\ge h(f\,,\mathcal{B})$. As an example, if $\epsilon<\eta$ then clearly $\mathcal{C}_\epsilon$ is a refinement of $\mathcal{C}_\eta$, and it follows that $h(f,\epsilon)\ge h(f,\eta)$.

Proof of 7.9. Let $\mathcal{O}_\epsilon$ be the covering of X by all open sets of diameter $\le\epsilon$. Then $\mathcal{C}_{\epsilon/2}$ is a refinement of $\mathcal{O}_\epsilon$, which in turn is a refinement of $\mathcal{C}_\epsilon$. Hence

$$h(f\,,\mathcal{C}_{\epsilon/2}) \quad\ge\quad h(f\,,\mathcal{O}_\epsilon) \quad\ge\quad h(f\,,\mathcal{C}_\epsilon) \quad .$$

Taking the limit as $\epsilon\to 0$, it follows that

$$h_{\rm top}(f) \quad = \quad \lim_{\epsilon\to 0} h(f\,,\mathcal{O}_\epsilon) \quad\le\quad \sup_{\mathcal{U}} h(f\,,\mathcal{U}) \quad .$$

On the other hand, any covering $\mathcal{U}$ of the compact metric space X by open sets has a *Lebesgue number* $\epsilon>0$ with the property that every set of diameter $\le\epsilon$ is contained in some $U\in\mathcal{U}$. If we exclude the trivial case where $X\in\mathcal{U}$, then the real valued function

$$\phi(x) \quad = \quad \sup_{U\in\mathcal{U}} \mathbf{d}(x\,,\,X\smallsetminus U)$$

is continuous and strictly positive on X. Any ϵ less than the minimum value of ϕ will serve as the required Lebesgue number. Now it follows that $\mathcal{C}_\epsilon$ is a refinement of $\mathcal{U}$, hence $h_{\rm top}(f)\ge h(f,\epsilon)=h(f,\mathcal{C}_\epsilon)\ge h(f\,,\mathcal{U})$. Taking the supremum over all coverings by open sets, we obtain

$$h_{\rm top}(f) \quad\ge\quad \sup_{\mathcal{U}} h(f\,,\mathcal{U}) \quad ,$$

which completes the proof. □

---

§7F. Piecewise Monotone Maps of the Interval.

Let $I=[a,b]$ be a closed interval of real numbers. By definition, a map $f:I\to I$ is *piecewise monotone* if there are points

$$a=c_0 \;<\; c_1 \;<\; c_2 \;<\; \cdots \;<c_n=b$$

such that f is strictly monotone, either increasing or decreasing, on each interval $I_j=[c_{j-1}\,,\,c_j]$. [Misiurewicz and Szlenk] have provided a formula for computing the topological entropy of such a map, which can be described as follows. Let us define a finite sequence $(j_0\,,\,j_1\,,\,\cdots\,,\,j_k)$ of integers between 1 and n, to be *admissible* (for the partition

$I = I_1 \cup \cdots \cup I_n$) if there exists an orbit $x_0 \mapsto x_1 \mapsto x_2 \cdots$ such that x_i belongs to the interior of the interval I_{j_i} for $0 \le i \le k$. Equivalently this sequence, of length $k+1$, is admissible if and only if the intersection

$$I(j_0\,,\,j_1\,,\,\ldots\,,\,j_k) \;=\; I_{j_0} \,\cap\, f^{-1}(I_{j_1}) \,\cap\, f^{-2}(I_{j_2}) \,\cap\, \cdots \,\cap\, f^{-k)}(I_{j_k}) \qquad (7:9)$$

contains an interior point. Whenever this condition is satisfied, a straightforward induction shows that the k-fold composition

$$I(j_0\,,\,j_1\,,\,\ldots\,,\,j_k) \xrightarrow{f} I(j_1\,,\,j_2\,,\,\ldots\,,\,j_k) \xrightarrow{f} \cdots \xrightarrow{f} I(j_k) = I_{j_k} \qquad (7:10)$$

embeds $I(j_0\,,\,j_1\,,\,\ldots\,,\,j_k)$ homeomorphically as a non-degenerate subinterval of I_{j_k} .

Definition. Let $\mathrm{Admis}(k) \le n^k$ be the number of admissible sequences $(j_0\,,\,j_1\,,\,\ldots\,,\,j_{k-1})$ of length k . Then we can state their result as follows.

Theorem 9.10 (Misiurewicz and Szlenk). *The topological entropy* $h_{\mathrm{top}}(f)$ *of a piecewise-monotone map can be computed as* $\lim_{k\to\infty} \frac{1}{k} \log \mathrm{Admis}(k)$.

In particular, since $1 \le \mathrm{Admis}(k) \le n^k$, it follows that $0 \le h_{\mathrm{top}}(f) \le \log n$. I will not give the proof of 9.10, but will derive several consequences.

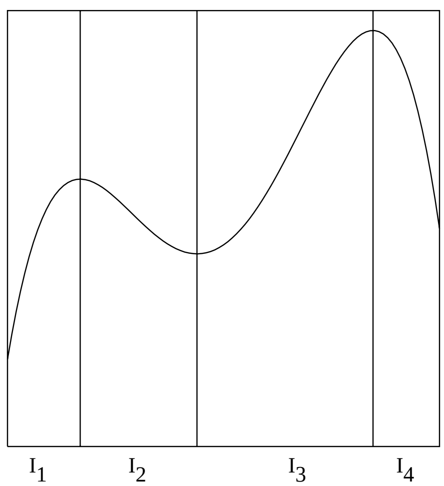

Figure 35. Graph of a piecewise monotone map with 4 laps.

For any piecewise monotone map, we can choose the unique minimal collection of points c_j which decompose I into intervals on which f is monotone. In fact these c_j must be precisely the points at which f attains a local maximum or minimum. The corresponding intervals $I_1\,,\,\ldots\,,\,I_n$ are called the *laps* of f . I will write $n = \mathrm{Lap}\#(f)$.

Corollary 7.11. *The entropy* $h_{\mathrm{top}}(f)$ *of a piecewise monotone map can be characterized as the limit of* $\frac{1}{k} \log \mathrm{Lap}\#(f^{\circ k})$ *as* $k \to \infty$.

Proof. If $I_1\,,\,\ldots\,,\,I_n$ are the laps of f , then it is not difficult to see that the various non-trivial intersections (7 : 9) are precisely the various laps of the iterate $f^{\circ(k+1)}$. $\square$

By definition, a piecewise monotone map f is a *Markov map* if the subdivsion points $c_0 < c_1 < \cdots < c_n$ can be chosen so that each $f(c_j)$ is itself one of these points

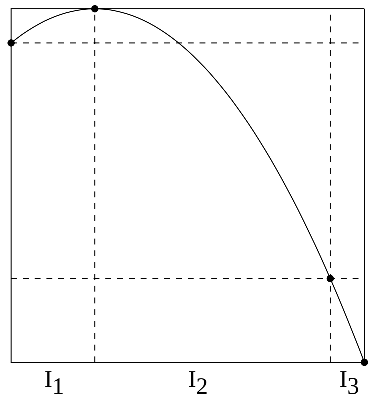

Figure 36. Graph of a Markov map with $I_1 \xrightarrow{\cong} I_3$, $I_2 \xrightarrow{\cong} I_2 \cup I_3$, $I_3 \xrightarrow{\cong} I_1$.

c_0 , c_1 , ... , c_n . An equivalent condition would be that each I_j maps homeomorphically onto some union of consecutive $I_{j'}$. Define the associated $n \times n$ ***transition matrix*** $M = [M_{ij}]$ by the requirement that

$$M_{ij} \;=\; \begin{cases} +1 & \text{if} \quad f(I_i) \supset I_j\,, \\ 0 & \text{otherwise}\,. \end{cases}$$

Thus when $M_{ij} = 0$ the intervals $f(I_i)$ and I_j have at most a single endpoint in common.

Corollary 7.12. *If* f *is a Markov map of the interval with associated transition matrix* $M = [M_{ij}]$ *, and if* $\lambda_{\max}$ *is the largest real eigenvalue of* M *, then the entropy* $h_{\rm top}(f)$ *is equal to* $\log(\lambda_{\max})$ *.*

As examples, for the Markov maps of Figures 36 and 37 the associated matrices are

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}$$

respectively. In the first case, the eigenvalues are $1, 1, -1$, so that $h_{\rm top} = \log 1 = 0$, while in the second case the eigenvalues are $(1 \pm \sqrt{5})/2$, so that $h_{\rm top}(f) = \log\left((1+\sqrt{5})/2\right) = 0.4812 \cdots$.

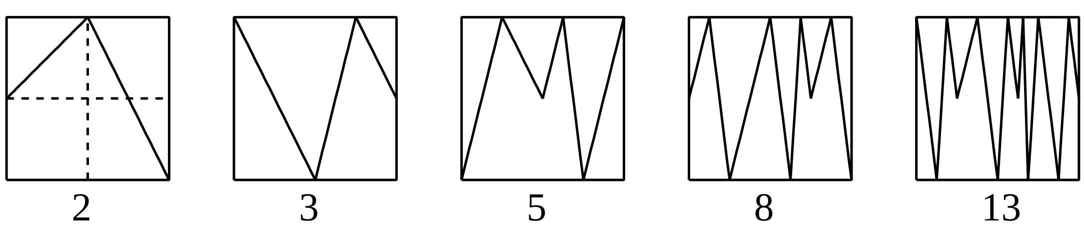

Figure 37. Graphs for the sequence of iterates $f^{\circ k}$ *of the piecewise linear Markov map associated with a period 3 orbit* $0 \mapsto 1/2 \mapsto 1 \mapsto 0$ *. Note that the lap numbers* $\mathrm{Lap}\#(f^{\circ k})$ *, shown under the graphs, are Fibonacci numbers.* (*Compare Problem 7-k.*)

Proof of 7.12. In the case of a Markov partition $I = I_1 \cup \cdots \cup I_n$, note that the k-fold composition $(7:10)$ associated with any admissible sequence maps the interval

$I(j_0\,,\,j_1\,,\,\dots\,,\,j_k)$ homeomorphically onto the interval $I(j_k) = I_{j_k}$. Evidently the sequence $(j_0\,,\,j_1\,,\,\dots\,,\,j_k)$ is admissible if and only if

$$M_{j_0 j_1} \;=\; M_{j_1 j_2} \;=\; \cdots \;=\; M_{j_{k-1} j_k} \;=\; 1\;.$$

Now let M^k be the k-fold matrix product with M with itself. Evidently the (i,j)-th entry $(M^k)_{ij}$ of M^k is equal to the number of admissible sequences of the form

$$i = j_0\,,\,j_1\,,\,\dots\,,\,j_{k-1}\,,\,j_k = j\;.$$

Summing over i and j , this shows that

$$\mathrm{Admis}(k+1) \;=\; \sum_{ij}(M^k)_{ij}\;.$$

For any $n \times n$ real or complex matrix A , let $\lambda_1\,,\,\dots\,,\,\lambda_n$ be the associated eigenvalues, and define the *spectral radius* of A to be the maximum of the absolute values $|\lambda_h|$. If we set $\|A\| = \sum_{ij}|A_{ij}|$, then it is not difficult to show that this spectral radius can be expressed as a limit

$$\max_h |\lambda_h| \;=\; \lim_{k\to\infty}\;\sqrt[k]{\|A^k\|}\;.$$

(Compare A.3 in the Appendix.) In the case of a Markov matrix M , since the entries M_{ij} are all non-negative real numbers, it follows by the Perron-Frobenius Theorem that the spectral radius is equal to the largest real eigenvalue. (Again see the Appendix.) Thus

$$h_{\mathrm{top}}(f) \;=\; \lim_{k\to\infty}\frac{1}{k}\log \mathrm{Admis}(k+1) \;=\; \lim_{k\to\infty}\frac{1}{k}\log\|M^k\| \;=\; \log\lambda_{\mathrm{max}}\;,$$

which completes the proof of 7.12. □

The largest possible entropy associated with an $n \times n$ matrix of zeros and ones occurs when all of the entries are one, so that all of the n^k symbol sequences of length k are admissible. In this case, we have $\mathrm{Admis}(k) = n^k$, hence $h_{\mathrm{top}}(f) = \log n$. As an example, for the Tschebychef map $f(x) = 4x^3 - 3x$ of order 3, as shown in Figure 38, since each of the three laps maps onto the full interval $[-1,1]$, it follows that $h_{\mathrm{top}}(f) = \log 3$.

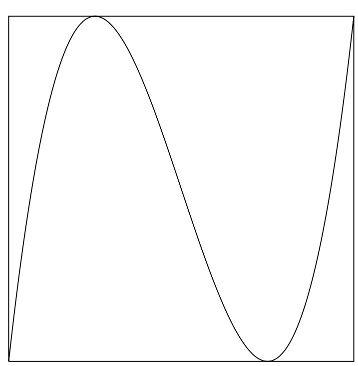

Figure 38. Graph of $x \mapsto 4x^3 - 3x$ *on the interval* $[-1,1]$.

As one application of this observation, we can give an example of an interval map with infinite entropy. Simply map the subinterval $[0,\frac{1}{2}]$ onto itself with entropy $\log 3$, the

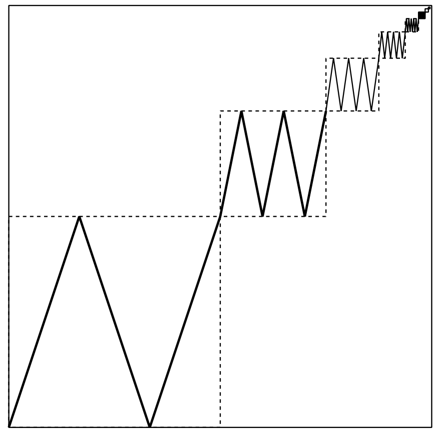

Figure 39. An interval map with infinite topological entropy

interval $[\frac{1}{2}, \frac{3}{4}]$ onto itself with entropy $\log 5$, and so on. Putting all of these together, we obtain a continuous map from the full interval onto itself whose entropy must be infinite.

As another application, we see that entropy need not vary continuously as we deform the map. The lefthand figure below shows a piecewise linear Markov map where both the left half interval L_1 and the right half interval L_2 map homeomorphically onto L_1 . It is easy to check that $h_{\rm top}(f) = 0$. But an arbitrarily small perturbation yields a map g such that a small subinterval J is mapped onto itself with entropy $\log 2$. It follows easily that $h_{\rm top}(g) = \log 2$.

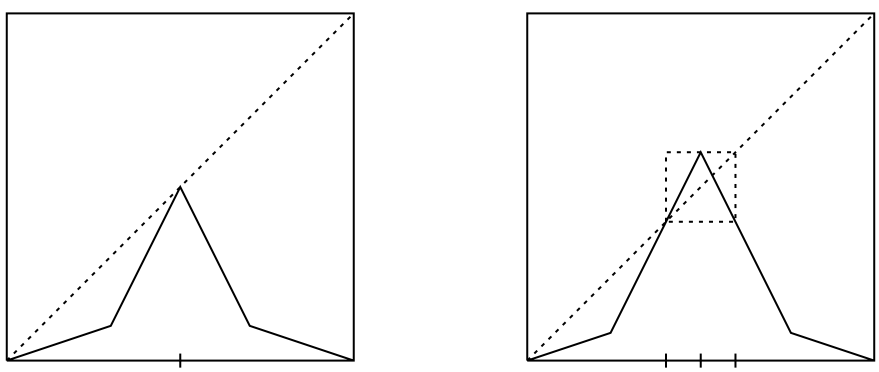

Figure 40. The map on the left has $h_{\rm top} = 0$, but an arbitrarily small perturbation yields a map with $h_{\rm top} = \log 2$.

For further discussion, including an algorithm for effectively computing the topological entropy of a piecewise monotone map, see [Milnor and Tressser]. See also [Milnor and Thurston], and for textbooks on 1-dimensional dynamics see [Alseda, Llibre and Misiurewicz] or [de Melo and van Strien].

§7G. Some Problems.

Problem 7-a. Unions. Show that $\dim_B^+(X \cup Y) = \max(\dim_B^+(X), \dim_B^+(Y))$.

Problem 7-b. Lipschitz Maps. If there is a map from X onto Y satisfying a

Lipschitz condition $\mathbf{d}(f(x)\,,\,f(x')) \le c\,\mathbf{d}(x,x')$, show that

$$\dim_B^-(X) \ge \dim_B^-(Y) \quad , \qquad \dim_B^+(X) \ge \dim_B^+(Y)\ .$$

Problem 7-c. Products. If we use the maximum metric

$$\mathbf{d}((x,y)\,,\,(x',y')) \;=\; \max\left(\mathbf{d}(x,x')\,,\,\mathbf{d}(y,y')\right)$$

on a cartesian product $X \times Y$, show that

$$C_\epsilon(X\times Y) \le C_\epsilon(X)\,C_\epsilon(Y) \quad , \qquad S_\epsilon(X\times Y) \ge S_\epsilon(X)\,S_\epsilon(Y) \quad ,$$

and that

$$\dim_B^+(X\times Y) \le \dim_B^+(X) + \dim_B^+(Y) \quad , \qquad \dim_B^-(X\times Y) \ge \dim_B^-(X) + \dim_B^-(Y)\ .$$

Problem 7-d. "Minkowski" Dimension. For $X \subset {}^n$ let $N_\epsilon(X)$ be the ϵ-*neighborhood*, consisting of all points $y \in {}^n$ for which $\mathbf{d}(y,X) \le \epsilon$. Show that there are constants $\alpha_n\,,\,\beta_n > 0$ so that the n-dimensional volume of this neighborhood satisfies

$$\alpha_n\,\epsilon^n\,S_\epsilon(X) \;\le\; \mathrm{vol}(N_\epsilon(X)) \;\le\; \beta_n\,\epsilon^n\,C_\epsilon(X)\ .$$

Conclude that

$$\dim_B^+(X) \;=\; n \;-\; \liminf_{\epsilon\to 0}\ \frac{\log \mathrm{vol}(N_\epsilon(X))}{\log \epsilon} \quad ,$$

with an analogous formula for $\dim_B^-(X)$.

Problem 7-e. An Intermediate Box Dimension. For X compact metric, show that there exists one and only one number $0 \le \beta(X) \le \infty$ with the following property: *The integral*

$$\int_0^1 S_\epsilon(X)\,d\epsilon^t \qquad or \qquad \int_0^1 C_\epsilon(X)\,d\epsilon^t$$

is finite for $t > \beta(X)$ *and is infinite for* $0 < t < \beta(X)$. Show that

$$\dim_B^-(X) \;\le\; \beta(X) \;\le\; \dim_B^+(X)\ .$$

Problem 7-f. Another Cantor Set. Here is another generalisation of the Cantor middle third set. Given any sequence of numbers $1 = \ell_0 > \ell_1 > \ell_2 > \cdots > 0$, define compact subsets

$$[0,1] \;=\; K_0 \;\supset\; K_1 \;\;\supset\; K_2 \;\supset \cdots$$

inductively as follows. Each K_m will be a union of 2^m disjoint equal intervals, with total length ℓ_m . The inductive step consists of cutting a middle portion out of each of the components of K_{m-1} , so as to leave two end segments, each of the required length $\ell_m/2^m$. If K is the intersection of the K_m , and if $\ell_m/2^m \le \epsilon < \ell_{m-1}/2^{m-1}$, show that the ϵ-covering number $C_\epsilon(K)$ is equal to 2^m . Now define numbers

$$0 \;\le\; L^- \;\le\; L^+ \;\le\; 1$$

by the formulas

$$L^- \;=\; \liminf_{m\to\infty}\ \sqrt[m]{\ell_m}\ , \qquad L^+ \;=\; \limsup_{m\to\infty}\ \sqrt[m]{\ell_m}\ .$$

Show that

$$\dim_H(K) \;=\; \dim_B^-(K) \;=\; \liminf_{\epsilon\to 0}\frac{\log C_\epsilon(K)}{\log(1/\epsilon)} \;=\; \liminf_{m\to\infty}\frac{m\log 2}{\log(2^m/\ell_m)} \;=\; \frac{1}{1-\log_2 L^-}$$

and similarly show that $\dim_B^+(K) = 1/(1-\log_2 L^+)$. By suitable choice of the sequence $\{\ell_i\}$, show that we can construct examples with any values of the box dimensions satisfying

$$0 \;\le\; \dim_B^-(K) \;\le\; \dim_B^+(K) \;\le\; 1 \;.$$

Show that the Lebesgue measure $\lambda(K) = \lim_{m\to\infty}\ell_m$ satisfies $0 \le \lambda(K) < 1$; but if $\lambda(K) > 0$ check that we must have $\dim_H(K) = \dim_B(K) = 1$.

Problem 7-g. Topological Dimension. Show that the topological dimension of a region in ^{n} (defined using coverings, as in §7B) is equal to n , as follows. Let Δ^n be an n-dimensional simplex. Start with the theorem that the identity map of an $(n-1)$-dimensional sphere is not homotopic to a constant map. Assuming this, show that any map from Δ^n into itself which maps each face into itself must be onto. Now consider any open covering $\{U_i\}$ which is fine enough so that no U_i intersects all of the $(n-1)$-dimensional faces. Prove that at least $n+1$ of the U_i must intersect, as follows. Let $\{\,\phi_i : \Delta^n \to [0,1]\,\}$ be an associated partition of unity, so that $\sum_i \phi_i(x) = 1$, with $\phi_i(x) = 0$ outside of U_i . For each U_i choose an $(n-1)$-dimensional face which is disjoint, and let v_i be the opposite vertex. Show that the map $x \mapsto \sum \phi_i(x)\, v_i$ carries each boundary face of Δ^n into itself, and hence is onto. Now if x_0 maps to any interior point of Δ^n then it follows that x_0 belongs to at least $n+1$ of the U_i . This proves that $\dim_{\rm top}(\Delta^n) \ge n$. On the other hand, choosing an arbitrarily fine subdivision of Δ^n , show that the "open star neighborhoods" of the vertices form a covering with the property that at most $n+1$ of these sets can intersect, so that $\dim_{\rm top}(X) \le n$.

Problem 7-h. Hausdorff Dimension. Show that $\dim_H(X) \le \dim_B^-(X)$. In particular, if $X \subset {}^n$ then it follows that $\dim_H(X) \le \dim_B^+(X) \le n$. If $X \subset {}^n$ has Lebesgue measure $\lambda(X) > 0$ show that $\dim_H(X) = \dim_B(X) = n$. (Use the fact that any set A_i of diameter d is contained in a cube of edge d and hence satisfies $\lambda(A_i) \le d^n$.)

Problem 7-i. A "Horseshoe". Now consider a map $f : X \to X$ and suppose that X contains two disjoint non-empty compact subsets with $f(X_i) \supset X_0 \cup X_1$ for $i = 0\,,1$. Given any finite sequence $\alpha_0\,, \alpha_1\,, \ldots\,, \alpha_n$ of zeros and ones, show that f has an orbit $x_0 \mapsto x_1 \mapsto x_2 \mapsto \cdots$ so that $x_j \in X_{\alpha_j}$ for $0 \le j \le n$. Conclude that $h_{\rm top}(f) \ge \log 2$.

Problem 7-j. An Example with Box Dimension > Hausdorff Dimension. Let $X = \{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots\} \subset [0,1]$ consist of the sequence of points $1/n$, $n \ge 1$, together with the limit point zero. Show that the number of ϵ-boxes needed to cover $X \cap [0\,, \sqrt{\epsilon}]$ is roughly $1/\sqrt{\epsilon}$, and the number needed to cover the remaining points of X is also roughly $1/\sqrt{\epsilon}$, so that

$$|\, B_\epsilon(X) \;-\; 2/\sqrt{\epsilon}\,| \;<\; \text{constant} \quad .$$

Conclude that $\dim_B(X)$ is defined and equal to $1/2$. On the other hand, since X is a countable set, show that $\dim_H(X) = 0$.

Problem 7-k. Fibonacci Numbers. For the map f of Figure 37, let

$$\phi(k) \;=\; \text{Lap}\#(f^{\circ k}) \;,$$

with $\phi(0) = 1$, $\phi(1) = 2$, $\phi(2) = 3$, and so on. Show that the number of laps of $f^{\circ k}$ in the left hand half-interval I_1 is $\phi(k-2)$ while the number in the right hand half-interval I_2 is $\phi(k-1)$. Thus

$$\phi(k) \;=\; \phi(k-2) + \phi(k-1) \;,$$

which shows that the $\phi(k)$ are Fibonacci numbers. Now check that the powers of the numbers $\xi_\pm = (1 \pm \sqrt{5})/2$ satisfy a corresponding recursion relation $\xi_\pm^k = \xi_\pm^{k-2} + \xi_\pm^{k-1}$. Conclude that we can write

$$\phi(k) \;=\; a\,\xi_+^{\,k} \,+\, b\,\xi_-^{\,k} \;,$$

where a and b are suitably chosen non-zero coefficients, and thus give a different proof that

$$\lim_{k\to\infty} \frac{1}{k} \log \mathrm{Lap}\#(f^{\circ k}) \;=\; \log \xi_+ \;.$$

Problem 7-l. Constant |Slope|. If the Markov map f is piecewise linear, with slope $f'(x) = \pm s$ everywhere, show that the lengths r_j of the intervals I_j form a right eigenvector

$$\sum_j M_{ij}\, r_j \;=\; s\, r_i \qquad \text{with} \qquad s \;=\; \lambda_{\max} \;.$$

Conversely, suppose that the Markov matrix for some given map g has a right eigenvector with all entries positive

$$\sum_i M_{ij}\, r_j \;=\; s\, r_i \;>\; 0 \qquad \text{for} \qquad 1 \le i \le n \;.$$

(Compare the Appendix.) Modifying g in such a way that each interval I_j is replaced by an interval of length r_j, and so that the mapping restricted to each I_j is linear, show that we can obtain a new map f with the same Markov matrix, but with

$$|\text{slope}| \;=\; s \;=\; \lambda_{\max} \qquad \text{everywhere} \;.$$

Problem 7-m. Periodic Points. If f is a Markov map with $n \times n$ transition matrix M, show that the number of fixed points of $f^{\circ k}$ satisfies

$$\#\mathbf{Fix}(f^{\circ k}) \,+\, n - 1 \;\ge\; \mathrm{trace}(M^k) \;.$$

Using the identity

$$\lambda_{\max} \;=\; \limsup \Big(\mathrm{trace}(M^k)\Big)^{1/k}$$

(see A.3 in the Appendix), conclude that

$$\limsup \; \#\mathbf{Fix}(f^{\circ k})^{1/k} \;\ge\; \lambda_{\max} \;.$$