Real Analysis Exchange
Vol. 29(1), 2003/2004, pp. 9–43

Zbigniew H. Nitecki,* Department of Mathematics, Tufts University, Medford, MA 02155, USA. email: `znitecki@tufts.edu`

TOPOLOGICAL ENTROPY AND THE PREIMAGE STRUCTURE OF MAPS

Abstract

My aim in this article is to provide an accessible introduction to the notion of topological entropy and (for context) its measure theoretic analogue, and then to present some recent work applying related ideas to the structure of iterated preimages for a continuous (in general non-invertible) map of a compact metric space to itself. These ideas will be illustrated by two classes of examples, from circle maps and symbolic dynamics. My focus is on motivating and explaining definitions; most results are stated with at most a sketch of the proof. The informed reader will recognize imagery from Bowen's exposition of topological entropy [Bow78] which I have freely adopted for motivation.

1 Measure-Theoretic Entropy

How much can we learn from observations using an instrument with finite resolution?

A simple model of a single observation on a "state space" X is a finite partition $\mathcal{P} = \{A_1, \ldots, A_N\}$ of X into atoms, grouping the points (states) in X according to the reading they induce on our instrument. A measure μ on X with total measure $\mu(X) = 1$ defines the probability of a given reading as

$$p_i = \mu(A_i), \quad i = 1, \ldots, N.$$

Shannon [Sha63] (see also [Khi57]) noted that the "entropy" of the partition

$$H(\mathcal{P}) := -\sum_{i=0}^{N} p_i \log p_i$$

Key Words: entropy, topological entropy, preimage entropy, topological pressure, subshift
Mathematical Reviews subject classification: primary: 37-02; secondary: 37B40, 37A35, 37B10
Received by the editors August 28, 2003
Communicated by: Paul D. Humke

*Based on a talk given June 23, 2003 at Summer Symposium in Real Analysis XXVII, Hradec nad Moravicí, Czech Republic (hosted by the Silesian University in Opava).

measures the *a priori* uncertainty about the outcome of an observation—or conversely the information we obtain from performing the observation. The extreme values of entropy among partitions with a fixed number N of atoms are $H(\mathcal{P}) = 0$, when the outcome is completely determined (some $p_i = 1$, all others $= 0$), and $H(\mathcal{P}) = \log N$, when all outcomes are equally likely ($p_i = \frac{1}{N}$, $i = 1, \ldots, N$).

To model a sequence of observations at different times, we imagine a dynamical system generated by the (μ-measurable) map $f : X \to X$, so the state initially at $x \in X$ evolves, after k time intervals, to the state located at $f^k(x)$, where

$$f^n := \underbrace{f \circ \ldots \circ f}_{n \text{ times}}.$$

An observation made after k time intervals is modelled by the partition $f^{-k}[\mathcal{P}] = \{f^{-k}[A_1], \ldots, f^{-k}[A_N]\}$, where the k^{th} iterated preimage of $A \subset X$ is

$$f^{-k}[A] := \{x \in X \mid f^k(x) \in A\}.$$

Assuming that μ is an f-invariant measure ($f^{-1}[A] = A$), the outcomes of observations made at different times are identically distributed. The joint distribution of n successive observations performed one time unit apart is modelled by the mutual refinement

$$\mathcal{P}_n := \mathcal{P} \vee f^{-1}[\mathcal{P}] \vee \ldots f^{-(n-1)}[\mathcal{P}]$$

whose typical atom, $A_{i_0} \cap f^{-1}[A_{i_1}] \cap \cdots \cap f^{-(n-1)}[A_{i_{n-1}}]$, consists of the points with a given *itinerary* of length n with respect to $\mathcal{P}$ (*i.e.*, $f^j(x) \in A_{i_j}$, $j = 0, \ldots, n-1$). The asymptotic average information per observation for a sequence of successive observations

$$H(f, \mathcal{P}) := \lim_{n \to \infty} \frac{1}{n} H(\mathcal{P}_n)$$

is the *entropy of f relative to $\mathcal{P}$*.

For example, suppose $f : X \to X$ is the restriction to the unit circle $S^1 := \{x \in \mathbb{C} \mid |x| = 1\}$ of $x \mapsto x^2$. If we parametrize S^1 by $\theta \in \mathbb{R}$ using $\exp(\theta) := e^{2\pi i \theta} \in S^1$, our map corresponds to $\theta \mapsto 2\theta \pmod{\mathbb{Z}}$, the **angle-doubling map**. (Lebesgue) arclength measure is invariant under this map, and if $\mathcal{P}$ is a partition into two semicircles, say $A_1 = \{0 \le \theta \le \frac{1}{2}\}$, $A_2 = \{\frac{1}{2} \le \theta \le 1\}$, then $\mathcal{P}_n$ is a partition into 2^n intervals of equal arclength. Thus $H(\mathcal{P}_n) = n \log 2$, so

$$H(f, \mathcal{P}) = \log 2.$$

Note that in this case the observations at different times are (probabilistically) independent: knowing the itinerary of length n does not help us predict the next position of a random point.

An equivalent model of this situation comes from expressing the angle in binary notation:

$$\theta = \sum_{i=0}^{\infty} \frac{x_i}{2^{i+1}}, \quad x_i \in \{0,1\},\ i = 0, 1, \dots$$

which is ambiguous only on the Lebesgue-null set of dyadic rational values for θ. Up to this ambiguity, we have a bijection with the set $\{0,1\}^{\mathbb{N}}$ of sequences $x = x_0, x_1, \dots$ in $\{0,1\}$. For any finite sequence $w = w_0, \dots, w_{n-1} \in \{0,1\}^n$, the *cylinder set*

$$\mathcal{C}(w) := \{x \in \{0,1\}^{\mathbb{N}} \mid x_i = w_i \text{ for } i = 0, \dots, n-1\}$$

of sequences which begin with w corresponds to an arc in S^1 of length 2^{-n}, and we can define a measure μ on $\{0,1\}^{\mathbb{N}}$ via

$$\mu(\mathcal{C}(w)) = 2^{-n} \text{ for all } w \text{ of length } n,$$

which is equivalent to arclength measure on S^1. The angle-doubling map corresponds to the **shift map** on sequences

$$s(x_0 x_1 x_2 \dots) = x_1 x_2 \dots.$$

More generally, if $\mathfrak{A}$ is a finite set ("alphabet") and we assign a "weight" $p(a) \geq 0$ to each "letter" $a \in \mathfrak{A}$ so that $\sum_{a \in \mathfrak{A}} p(a) = 1$, then the formula

$$\mu(\mathcal{C}(w_0 \dots w_{n-1})) = p(w_0) p(w_1) \cdots p(w_{n-1})$$

defines a probability measure on the space of sequences [1]

$$\mathfrak{A}^{\mathbb{N}} := \{x = x_0 x_1 \dots \mid x_i \in \mathfrak{A},\ i = 0, 1, \dots\}$$

and the natural shift map on $\mathfrak{A}^{\mathbb{N}}$ with this measure is called a **Bernoulli shift**. The partition $\mathcal{P} = \{\mathcal{C}(a) \mid a \in \mathfrak{A}\}$ has entropy

$$H(\mathcal{P}) = -\sum_{a \in \mathfrak{A}} p(a) \log p(a).$$

[1] It will be convenient to abuse notation and include 0 in $\mathbb{N}$.

The refinement $\mathcal{P}_n$ consists of all cylinder sets $\mathcal{C}(w)$ as w ranges over "words" $w = w_0 \dots w_{n-1} \in \mathfrak{A}^n$ of length $|w| = n$, and a straightforward calculation shows that successive observations are independent, and

$$H(\mathcal{P}_n) = nH(\mathcal{P}), \quad H(s, \mathcal{P}) = H(\mathcal{P}).$$

The quantity $H(f, \mathcal{P})$ depends on our observational device. We obtain a device-independent measurement of the predictability of the dynamics of the measure-theoretic model $f:(X,\mu)\to(X,\mu)$ by maximizing over all finite partitions: this is the **entropy of f with respect to $\boldsymbol{\mu}$**:

$$h_\mu(f) := \sup\{H(f, \mathcal{P}) \mid \mathcal{P} \text{ a finite measurable partition of } X\}.$$

It can be shown that the partition $\mathcal{P}$ of S^1 into semicircles maximizes $H(f, \mathcal{P})$ for the angle-doubling map, so $h_\mu(f) = \log 2$ in this case. For the general Bernoulli shift (determined by the weights $p(a),\ a \in \mathfrak{A}$), the partition $\mathcal{P} = \{\mathcal{C}(a) \mid a \in \mathfrak{A}\}$ into cylinder sets again maximizes entropy, so in this case

$$h_\mu(f) = -\sum_{a\in\mathfrak{A}} p(a) \log p(a).$$

For example, the Bernoulli shift corresponding to a biased coin flip, say $p(0) = \frac{1}{3}$, $p(1) = \frac{2}{3}$, has entropy $h_\mu(f) = \log 3 - \frac{2}{3}\log 2$.

The idea of using Shannon's entropy in this way was suggested by Kolmogorov [Kol58] (and refined by Sinai [Sin59]), who showed that $h_\mu(f)$ is invariant under measure-theoretic equivalence of dynamical systems, and used this to prove the existence of non-equivalent Bernoulli shifts. Subsequently Ornstein [Orn74] showed that for a large class of ergodic systems (including Bernoulli shifts [Orn70]) $h_\mu(f)$ is a complete invariant: two systems from this class are equivalent precisely if they have the same (measure-theoretic) entropy.

2 Topological Entropy

Adler, Konheim and McAndrew [AKM65] formulated an analogue of $h_\mu(f)$ when the measure space (X,μ) is replaced by a compact topological space and f is assumed continuous. They replaced the partition $\mathcal{P}$ with an open cover and the entropy $H(\mathcal{P})$ with the logarithm of the minimum cardinality of a subcover. The resulting *topological entropy*, $h_{top}(f)$, is an invariant of topological conjugacy between continuous maps on compact spaces.

A more intuitive formulation of $h_{top}(f)$, given independently by Bowen [Bow71] and Dinaburg [Din70], uses separated sets in a (compact) metric space.

2.1 Separated Sets

Let us again model observations via instruments with finite resolution, but this time using a (compact) metric d on our space X. We assume that our instrument can distinguish points $x, x' \in X$ precisely if $d(x, x') \geq \varepsilon$ for some positive constant ε. A subset $E \subset X$ is **ε-separated**[2] if our instrument can distinguish the points of E. Compactness puts a finite upper bound on the cardinality of any ε-separated set in X, and we can define

$$\text{maxsep}[d, \varepsilon, X] := \max\{\text{card}[E] \mid E \subset X \text{ is } \varepsilon\text{-separated with respect to } d\}.$$

On the circle, using d the normalized arclength

$$d(exp(\theta), exp(\theta')) = \min_{j \in \mathbb{Z}} |\theta - \theta' + j|$$

any set of N equally spaced points

$$E_N(exp(\theta)) := \{exp(\theta + \frac{j}{N}) \mid j = 0, \ldots, N-1\}$$

is a maximal ε-separated set whenever $\frac{1}{N+1} < \varepsilon \leq \frac{1}{N}$, so

$$\text{maxsep}[d, \frac{1}{N}, S^1] = \text{card}[E_N(x)] = N.$$

The sequence space $\mathfrak{A}^{\mathbb{N}}$ has a natural topology as the countable product of copies of the alphabet $\mathfrak{A}$ (which is given the discrete topology); this is captured in the metric

$$d(\mathbf{x}, \mathbf{x}') := 2^{-\delta(\mathbf{x}, \mathbf{x}')}$$

where

$$\delta(\mathbf{x}, \mathbf{x}') := 1 + \min\{i \mid x_i \neq x'_i\}.$$

Note that if two sequences $\mathbf{x}$, $\mathbf{x}'$ have different initial words w, w' of length n (*i.e.*, $\mathbf{x} \in \mathcal{C}(w)$, $\mathbf{x}' \in \mathcal{C}(w')$, $|w| = |w'| = n$ and $w \neq w'$), then $\delta(\mathbf{x}, \mathbf{x}') \leq n$, so $\mathcal{C}(w)$ and $\mathcal{C}(w')$ are at mutual distance at least 2^{-n}, and each such cylinder has diameter $2^{-(n+1)}$. It follows that a set consisting of one representative from each cylinder set $\mathcal{C}(w)$, $w \in \mathfrak{A}^n$, is a maximal 2^{-n}-separated set, and since there are $(\text{card}[\mathfrak{A}])^n$ words of length n,

$$\text{maxsep}[d, 2^{-n}, \mathfrak{A}^{\mathbb{N}}] = (\text{card}[\mathfrak{A}])^n.$$

---

[2]In §6, we mention the complementary notion of *ε-spanning*, and note that all of the definitions which follow can be reformulated in terms of this notion.

2.2 Bowen-Dinaburg Definition of Topological Entropy

Now we introduce dynamics via a continuous map $f: X \to X$, and ask about the resolution of n successive observations separated by unit time intervals. This is captured in the **Bowen-Dinaburg metrics**, defined for $n = 1, 2, \ldots$ by

$$d_n^f(x, x') := \max_{0 \leq i < n} d(f^i(x), f^i(x')).$$

Two points $x, x' \in X$ cannot be distinguished by our sequence of measurements if they **(n,ε)-shadow** each other (*i.e.*, $d(f^i(x), f^i(x')) < \varepsilon$ for $i = 0, \ldots, n-1$), so the points of $E \subset X$ are distinguished precisely if any two $x \neq x' \in E$ have $d_n^f(x, x') \geq \varepsilon$–that is, E is ε-separated with respect to d_n^f, or **($\mathbf{n}, \boldsymbol{\varepsilon}$)-separated**. The number of distinguishable orbit segments of length n is thus

$$\text{maxsep}[d_n^f, \varepsilon, X] = \max\{\text{card}[E] \mid E \subset X \text{ is } (n, \varepsilon) - \text{separated}\}.$$

For the angle-doubling map, note that if $d(x, x') \leq \frac{1}{4}$ then $d(f(x), f(x')) = 2d(x, x')$. In particular, if

$$d(x, x') = 2^{-k}$$

for some $k \geq 1$ then

$$d(f^j(x), f^j(x')) = \begin{cases} 2^{j-k} & \text{for } j < k \\ 0 & \text{for } j \geq k. \end{cases}$$

and, noting that $f(E_{2^k}(x)) = E_{2^{k-1}}(f(x))$, we see that $E_{2^k}(x)$ is

- 2^{-k}-separated with respect to d, and
- (n, ε)-separated for any $\varepsilon \leq \frac{1}{2}$ if $n \geq k$.

In particular, for $0 < \varepsilon < \frac{1}{2}$ and $n > log_{\frac{1}{2}} \varepsilon$,

$$\text{maxsep}[d_n^f, \varepsilon, S^1] = \text{card}[E_{2^n}(x)] = 2^n.$$

An (n, ε)-separated set is analogous to a collection of different itineraries of length n (with respect to some partition whose atoms have diameter ε). Since the number of conceivable itineraries grows exponentially with n, it is natural to look at the exponential growth rate of the cardinalities above. For any sequence $\{c_n\}$ of positive real numbers, we write

$$GR\{c_n\} := \limsup_n \frac{1}{n} \log c_n.$$

The complexity of the dynamics emanating from any subset $K \subset X$ is reflected in

$$h_{top}(f, K) := \lim_{\varepsilon \to 0} GR\{\text{maxsep}[d_n^f, \varepsilon, K]\}.$$

Our primary interest is when $K = X$: the **topological entropy** of $f : X \to X$ is

$$h_{top}(f) := h_{top}(f, X).$$

We have seen that the angle doubling map has topological entropy $\log 2$; in fact the analogous *angle-stretching maps* $\zeta_k : x \mapsto x^k$ $(k \geq 2)$ satisfy

$$h_{top}(\zeta_k) = \log k.$$

A beautiful general relation between measure-theoretic and topological entropy was established through the work of Goodwyn [Goo69], Dinaburg [Din70] and Goodman [Goo71]:

Theorem 1 (Restricted Variational Principle for Entropy) *For $f : X \to X$ any continuous map on a compact metric space,*

$$h_{top}(f) = \sup\{h_\mu(f) \mid \mu \text{ is an } f\text{-invariant Borel probability measure on } X\}.$$

2.3 One-sided Subshifts

The shift map on the sequence space $\mathfrak{A}^{\mathbb{N}}$

$$s(x_0 x_1 x_2 \dots) = x_1 x_2 \dots$$

is a card$[\mathfrak{A}]$-to-one map, continuous with respect to the product topology. By a (one-sided[3]) **subshift** we mean the restriction $f : X \to X$ of the shift to a closed invariant[4] subset $X \subset \mathfrak{A}^{\mathbb{N}}$. Such a set is determined by its admissible words: for $n = 1, 2, \dots,$ let

$$W_n(X) := \{w = w_0 \dots w_{n-1} \in \mathfrak{A}^n \mid \exists x \in X,\ i \in \mathbb{N} \text{ with } x_{i+j} = w_j,\ j = 0, \dots, n-1\}.$$

Note that a word which appears starting at position i in $x \in X$ appears as the *initial* subword of $f^i(x)$, which also belongs to X if X is shift-invariant. Thus $W_n(X)$ equals the set of words $w \in \mathfrak{A}^n$ with $X \cap \mathcal{C}(w)$ nonempty, and it follows that a maximal 2^{-n}-separated set $E_n \subset X$ results from picking one

[3]The space $\mathfrak{A}^{\mathbb{Z}}$ of bisequences also has a natural shift map, and invariant subsets are called *two-sided* subshifts. (*cf* §5.3)

[4]we require only $f(X) \subseteq X$

representative from each such nonempty intersection. Thus for $2^{-(k+1)} < \varepsilon < 2^{-k}$, E_{n+k} is a maximal (n, ε)-separated set, and

$$\text{maxsep}[d_n^f, \varepsilon, X] = \text{card}[W_{n+k}(X)]$$

giving us for any subshift $f : X \to X$

$$h_{top}(f) = \lim_{k\to\infty} GR\{\text{card}[W_{n+k+1}(X)]\} = GR\{\text{card}[W_n(X)]\}.$$

We spell out the results of this calculation for several examples.

Full shift: When $W_n(X) = \mathfrak{A}^n$, so $X = \mathfrak{A}^{\mathbb{N}}$, we have

$$h_{top}(f) = GR\{\text{card}[\mathfrak{A}]^n\} = \log \text{card}[\mathfrak{A}]$$

"Golden Mean" Shift: Define X as the set of all sequences of 0's and 1's in which 1 is never followed immediately by itself, so $W_2(X) = \{00, 01, 10\}$. If we list all words of length n, then the words of length $n+1$ come from either following an arbitrary word of length n with 0, or following a word of length n that ends in 0 with a 1. If we set

$$w_n := \text{card}[W_n(X)],$$

we see that there are w_n words of length $n+1$ which end in 0, and hence w_{n-1} words of length $n+1$ which end in 1: this gives the recursive relation

$$w_{n+1} = w_n + w_{n-1}$$

showing that w_n grows at the same rate as the Fibonacci numbers F_n (in fact, $w_n = F_{n+3}$). This rate is known [LM95, p. 101] to be the logarithm of the golden mean, so

$$h_{top}(f) = GR\{w_n\} = GR\{F_n\} = \log\left(\frac{1+\sqrt{5}}{2}\right).$$

A generalization of this example arises from any finite alphabet $\mathfrak{A} = \{a_1, \ldots, a_N\}$ and a list $\mathcal{W}_a \subset \mathfrak{A}^2$ of allowed pairs: X is then defined as the set of all sequences in $\mathfrak{A}^{\mathbb{N}}$ for which every subword of length 2 belongs to $\mathcal{W}_a$. This information can be encoded in a square *transition matrix* A of size $N = \text{card}[\mathfrak{A}]$ whose (i, j) entry is 1 (*resp.* 0) if the word $a_i a_j$ belongs (*resp.* does not belong) to $\mathcal{W}_a$. Note that the (i, j) entry of a power A^k of A equals the number of admissible words of length $k+1$

which begin with a_i and end with a_j, so $w_n := \mathrm{card}[W_n(X)]$ equals the sum $\| A^{n-1} \|$ of the entries of A^{n-1}, and we have

$$h_{top}(f) = GR\{w_n\} = GR\{\| A^{n-1} \|\} = \log(\text{spectral radius of } A).$$

In the special case of the "golden mean" shift, we have

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

whose characteristic polynomial, $t^2 - t - 1$, has the golden mean as its larger root.

Even Shift: Let X be the set of sequences of 0's and 1's in which two successive appearances of 1 are separated by a block of consecutive 0's of *even* length (which may be the empty block, of length zero). This is most easily described by giving a list $\mathcal{W}_d$ of *disallowed* words, in this case

$$\mathcal{W}_d = \{1(0)^{2n+1}1 \,|\, n = 0, 1, \dots\}$$

and specifying that X consists of all sequences in which no word from $\mathcal{W}_d$ appears (anywhere).

In general, such a description essentially specifies a basis of open subsets of the complement $\mathfrak{A}^{\mathbb{N}} \setminus X$. When such a list is (or can be made) finite, a recoding allows us to construct X as a subshift on more symbols, but specified as in the previous case by the allowed (or disallowed) *pairs*. This is called a **subshift of finite type** (or **topological Markov chain**).

The "even" shift is clearly *not* of finite type, as no test on words of bounded length can detect long forbidden words. However, it can be shown [LM95, p. 103] that in this case $\mathrm{card}[W_n] = F_{n+3} - 1$ (where again F_n is the n^{th} Fibonacci number), so the even shift has

$$h_{top}(f) = GR\{F_{n+3} - 1\} = \log\left(\frac{1+\sqrt{5}}{2}\right).$$

Dyck Shift: This beautiful example, first suggested by Krieger [Kri72] and named after an early contributor to the study of free groups and formal languages, codifies the rules of matching parentheses. As it is not readily accessible in the literature, I give a detailed account[5] based on ideas I learned from Doris and Ulf Fiebig.

[5] I thank Eugen Mihailescu for pointing out a substantial error in an earlier version of this account.

The alphabet consists of N pairs of matching left and right delimiters

$$\mathfrak{A} = \{\ell_1, r_1, \ldots, \ell_N, r_N\}.$$

For example, when $N = 2$, we can think of

$$\ell_1 = \text{“(”},\ r_1 = \text{“)”},\ \ell_2 = \text{“\{”},\ r_2 = \text{“\}”}.$$

Call a word $w = w_0, ..., w_{2k-1}$ of even length **balanced** if its entries can be paired subject to

- *a pair of entries consists of a left delimiter to the left of a matching right delimiter:* if w_α is paired with w_β, where $0 \leq \alpha < \beta \leq 2k-1$, then $w_\alpha = \ell_i$ for some index i and $w_\beta = r_i$ for the *same* index;
- *distinct pairs are nested or disjoint:* given $\alpha < \beta$ as above, every intermediate w_γ ($\alpha < \gamma < \beta$) is paired with some other intermediate w_δ ($\alpha < \delta < \beta$).

Note that a pairing of this type is unique if it exists. We regard the empty word as balanced.

Now we specify the (infinite) list of disallowed words as

$$\mathcal{W}_d = \{\ell_i b r_j \mid b \text{ is a balanced word and } i \neq j\}.$$

The subshift on the set of sequences $\mathfrak{D}_N \subset \mathfrak{A}^{\mathbb{N}}$ in which no element of $\mathcal{W}_d$ appears is the (one-sided) **Dyck shift** on N pairs. When $N = 1$, $\mathcal{W}_d$ is empty, so $\mathfrak{D}_1$ is the full shift on two symbols; we will tacitly assume that $N \geq 2$.

Proposition 1 *The Dyck shift $f : \mathfrak{D}_N \to \mathfrak{D}_N$ on N pairs has*

$$h_{top}(f) = \log(N+1).$$

PROOF. An admissible word has the general form

$$w = b_0 r_{i_1} b_1 r_{i_2} \ldots b_{k-1} r_{i_k} b_k \ell_{j_1} b_{k+1} \ldots \ell_{j_m} b_{k+m}$$

where each b_α, $\alpha = 0, \ldots, k+m$, is a (possibly empty) balanced subword, and the $k \geq 0$ *right* delimiters which are not matched in w all occur to the left of the $m \geq 0$ unmatched *left* delimiters in w. This leads to a natural decomposition of any admissible word as a concatenation of three (possibly empty) subwords

$$w = ABC$$

where $B = b_k$ is balanced, while $A = b_0 \dots r_{i_k}$ (*resp.* $C = \ell_{j_1} \dots b_{k+m}$) ends (*resp.* starts) with an unmatched right (*resp.* left) delimiter.

To calculate the topological entropy, note first that every admissible word w is the initial subword of at least $N + 1$ admissible words of length $|w|+1$: the N words $w\ell_i$, $i = 1, \dots, N$ are always admissible, and wr_{j_m} is admissible if $m \geq 0$ while *all* words wr_i are admissible if $m = 0$. Thus

$$\mathrm{card}[W_{n+1}] \geq (N+1)\mathrm{card}[W_n]$$

for all n, and so

$$h_{top}(f) = GR\{\mathrm{card}[W_n]\} \geq \log(N+1).$$

To handle the opposite inequality, we first estimate the cardinality of the sets A_n, B_n, C_n of admissible words of length n whose decomposition has only one nonempty factor, of the type indicated by the letter.

We begin with balanced words: since $\mathsf{B}_n = \emptyset$ for n odd, assume $n = 2p$. To estimate $\mathrm{card}[\mathsf{B}_n]$, we note that the number of possible configurations of p "ℓ"'s and p "r"'s in a balanced word of length n is bounded above by $\binom{n}{p}$, and for each such configuration, once we have assigned an index to each ℓ (which we can do in N^p ways), the uniqueness of the pairing insures that the word has been determined. Thus,

$$\mathrm{card}[\mathsf{B}_n] \leq \binom{n}{p} N^p < (N+1)^n,$$

where the last inequality is a consequence of the binomial theorem.

We now consider the set C_n of words beginning with an unmatched left delimiter, noting that the initial length k subword of any $w \in \mathsf{C}_n$ itself belongs to C_k. Given $w \in \mathsf{C}_n$, we immediately have $w\ell_i \in \mathsf{C}_{n+1}$ for $i = 1, \dots, N$, and $wr_i \in \mathsf{C}_{n+1}$ provided that w has at least two unmatched left delimiters, the *last* of which is ℓ_i. This gives us

$$\mathrm{card}[\mathsf{C}_{n+1}] \leq (N+1)\mathrm{card}[\mathsf{C}_n]$$

and since $\mathrm{card}[\mathsf{C}_1] = N$,

$$\mathrm{card}[\mathsf{C}_n] \leq (N+1)^n.$$

A similar estimate can be obtained for $\mathrm{card}[\mathsf{A}_n]$, either by repeating the argument or by noting the bijection between A_n and C_n obtained by reversing letter order and interchanging ℓ with r (keeping indices).

Finally, to estimate card$[W_n]$ we consider, for each ordered triple (i, j, k) of nonnegative integers summing to n, the set of words of the form $w = ABC$ with $|A| = i$, $|B| = j$, and $|C| = k$. Since an arbitrary factoring is possible, the number of such words is

$$\text{card}[\mathsf{A}_i] \cdot \text{card}[\mathsf{B}_j] \cdot \text{card}[\mathsf{C}_k] \leq (N+1)^{i+j+k} = (N+1)^n.$$

But the number of possible triples (i, j, k) summing to n is less than $(n+1)^3$, so

$$\text{card}[W_n] \leq (n+1)^3 (N+1)^n.$$

The growth rate of the right-hand quantity is $\log(N+1)$, so

$$h_{top}(f) = \log(N+1).$$

□

Square-Free Sequences: An even more complicated subshift is defined by forbidding any subword to immediately follow a copy of itself:

$$\mathcal{W}_d = \{w^2 := ww \mid w \in \mathfrak{A}^+ := \bigcup_{k=1}^{\infty} \mathfrak{A}^k\}.$$

An elementary argument shows that $\mathfrak{A}$ must have at least three letters for this to give a nonempty subshift. For three (or more) letters, there exist square-free sequences[6] and it is known [Bri63] that $h_{top}(f) > 0$. Although there are some bounds for the entropy [Gri01, She81a, She81b, SS82], a precise value has not been determined.

3 Pointwise Preimage Entropy

There is a curious asymmetry in the definitions of entropy in §§1-2, which look only at the future behavior of points. When f is invertible, it turns out that the inverse map f^{-1} has the same entropy: for $h_{top}(f)$, this follows from the observation that x and x' (n, ε)-shadow each other under a homeomorphism f precisely if their $f^{(n-1)}$-images (n, ε)-shadow each other under f^{-1}.

However, when f is not invertible the iterated preimages $f^{-n}[x]$ of a point are in general *sets* rather than points, so the formulations in §2 cannot be "reversed" in time. In 1991, Langevin and Walczak [LW91] built on ideas from their earlier work with Ghys (on the "entropy" of a foliation) to formulate an

[6] For experts, one example is the sequence $x_i = m_{i+1} - m_i \in \{-1, 0, 1\}$, where $\{m_i\}$ is the Morse-Thue sequence of 0's and 1's—I believe this was first observed by Thue.

invariant based on the behavior of preimages. We direct the interested reader to their original paper or to [NP99] for more details on this invariant, whose definition is rather involved; it is related to, and often equals, the branch preimage entropy which we present in §5.

Instead we begin with a more accessible pair of invariants defined by Hurley [Hur95] in 1995, looking at the growth rate of the size of iterated preimages of a point, measured via the Bowen-Dinaburg metrics. The two invariants differ in the stage at which one globalizes the pointwise measurements by maximizing over $x \in X$:

$$\begin{aligned} h_p(f) &:= \sup_{x \in X} \lim_{\varepsilon \to 0} GR\{ \qquad \operatorname{maxsep}[d_n^f, \varepsilon, f^{-n}[x]]\} \\ h_m(f) &:= \qquad \lim_{\varepsilon \to 0} GR\{ \max_{x \in X} \operatorname{maxsep}[d_n^f, \varepsilon, f^{-n}[x]]\} \end{aligned}$$

We refer to h_p and h_m collectively as **pointwise preimage entropies**; both are invariants of topological conjugacy [NP99] and we have the trivial inequalities

$$h_p(f) \leq h_m(f) \leq h_{top}(f).$$

There are examples for which either of these inequalities is strict: any homeomorphism with $h_{top}(f) > 0$ works for the second inequality (since $f^{-n}[x]$ is a single point, both pointwise preimage entropies are zero) and an example for the first is given in [FFN03]. However, the thrust of our discussion in this section and the next is that there are many cases when the three invariants agree. (We will also see this from a different perspective in §5.2.)

For the angle-doubling map, we note that the n^{th} iterated preimage of a point consists of 2^n equally spaced points:

$$f^{-n}[x] = E_{2^n}(x_n)$$

where x_n is any n^{th} preimage of x: for example if $x = exp(\theta)$ we can take $x_n = exp(2^{-n}\theta)$. Since this set is (n, ε)-separated if $\varepsilon \leq 2^{-n}$ (or $n \geq \log_{\frac{1}{2}} \varepsilon$), we have, independent of $x \in S^1$,

$$\operatorname{maxsep}[d_n^f, \varepsilon, f^{-n}[x]] = \operatorname{card}[f^{-n}[x]] = 2^n$$

so

$$h_p(f) = h_m(f) = \log 2.$$

A similar argument gives the common value $\log k$ for $h_p(\zeta_k)$ and $h_m(\zeta_k)$ where ζ_k is the angle-stretching map $x \mapsto x^k$, $k = 3, 4, \ldots$.

3.1 Pointwise Preimage Entropy for Subshifts

If $x \in X \subset \mathfrak{A}^{\mathbb{N}}$ is a point in the shift-invariant set X, its n^{th} **predecessor set** (in X) consists of all the words $w \in \mathfrak{A}^n$ of length n such that the concatenation wx also belongs to X:

$$P_n(x) = P_n(x, X) := \{w \in \mathfrak{A}^n \mid wx \in X\}.$$

Note that by definition $P_n(x, X) \subset W_n(X)$. Clearly, the n^{th} iterated preimage of x under the subshift $f : X \to X$ is the set of all concatentations wx, $w \in P_n(x, X)$, so from our earlier calculations, when $0 < \varepsilon \leq \frac{1}{2}$ and $x \in X$

$$\mathrm{maxsep}[d_n^f, \varepsilon, f^{-n}[x]] = \mathrm{card}[P_n(x)].$$

This immediately gives

$$\begin{aligned} h_p(f) &:= \sup_{x \in X} GR\{\phantom{\max_{x\in X}} \mathrm{card}[P_n(x)]\} \\ h_m(f) &:= \phantom{\sup_{x \in X}} GR\{\max_{x \in X} \mathrm{card}[P_n(x)]\} \end{aligned}$$

Again, we trace the application of this through our examples of subshifts:

Full Shift: Clearly, $P_n(x, \mathfrak{A}^{\mathbb{N}}) = \mathfrak{A}^n$ for all $x \in \mathfrak{A}^{\mathbb{N}}$, so

$$h_p(f) = h_m(f) = \log \mathrm{card}[\mathfrak{A}].$$

Subshifts of Finite Type: When X is defined by the transition matrix A, the predecessor set of any $x \in X$ is determined by its initial entry, x_0:

$$P_n(x, X) = \{w \in W_n(X) \mid wx \in W_{n+1}(X)\}$$

and the cardinality of this is the column sum in A^n corresponding to x_0. If we pick x_0 so that this column sum grows (with n) at least as fast as all the other columns, then any $x \in X$ beginning with x_0 has a maximal growth rate, and this equals the growth rate of $\| A^n \|$, so

$$h_p(f) = h_m(f) = GR\{\| A^n \|\} = \log(\text{spectral radius of } A).$$

Even Shift: The predecessor set of a sequence in the even shift is determined by the parity of the location of the first 1 in the sequence: if $x = 0^\infty$ then $P_n(x) = W_n(X)$, while if $x_k = 1$ and $x_i = 0$ for all $i < k$, then $w \in W_n(X)$ belongs to $P_n(x)$ if either $w = 0^n$ or w ends with 10^ℓ, where ℓ has the same parity as k. Thus $P_n(x)$ is in one-to-one correspondence with the set of admissible words of length $n+2$ (*resp.* $n+1$) ending with 01 (*resp.* 1) if k is odd (*resp.* if k is even or $x = 0^\infty$), and our earlier considerations show that all of these sets grow at the rate

$$h_p(f) = h_m(f) = \log\left(\frac{1+\sqrt{5}}{2}\right).$$

Dyck Shift: If x is a sequence formed by concatenating infinitely many balanced words, then

$$P_n(x, \mathfrak{D}_N) = W_n(\mathfrak{D}_N)$$

so

$$h_p(f) = h_m(f) = GR\{\text{card}[W_n(\mathfrak{D}_N)]\} = \log(N+1).$$

Square-Free Sequences: The predecessor sets in this subshift vary wildly from point to point (*cf* §5.1) and the tools used in the other cases tell us nothing about pointwise preimage entropy in this case.

The alert reader will have noted that in all the cases except the last, the pointwise preimage entropies $h_p(f)$ and $h_m(f)$ agree not only with each other but also with the topological entropy $h_{top}(f)$. This is no accident:

Theorem 2 ([FFN03]) *For any one-sided subshift* $f : X \to X$*, if*

$$GR\{W_n X\} = \log\lambda$$

then there exists a point $p \in X$ *such that*

$$card[P_n(p, X)] \geq \lambda^n \quad \text{for all } n = 1, 2, \ldots.$$

The argument for this rests on a combinatorial lemma[7] concerning the growth of branches in a tree, saying roughly that if we pick a "root" vertex and have, for some N, more than λ^N vertices at distance N from the root, then for some k (depending on λ, N, and the maximum valence of vertices in the tree) there exists a vertex v such that for $i = 1, \ldots, k$ the number of vertices at distance i from v, in a direction away from the root, is at least λ^i.

[7] A related result was apparently obtained by Furstenberg and Ledrappier & Peres.

4 Entropy Points

The phenomenon described for one-sided subshifts in the preceding section—that the preimages of some point determine the topological entropy—never occurs for homeomorphisms with positive topological entropy (*e.g.*, most two-sided subshifts), since any preimage of a point is still a single point. However, it is possible to resolve this cognitive dissonance via a calculation of topological entropy in the spirit of pointwise preimage entropy—looking at preimages of *local stable sets* instead of points.

For $\varepsilon > 0$, the **ε-stable set** of $x \in X$ under the map $f: X \to X$ is

$$S(x, \varepsilon, f) := \{y \in X \mid d(f^i(x), f^i(y)) < \varepsilon \text{ for all } i \geq 0\}.$$

(This is just the intersection of ε-balls with respect to the various Bowen-Dinaburg metrics.) We can define a kind of "ε-local preimage entropy" by

$$h_s(f, x, \varepsilon) := \lim_{\delta \to 0} GR\{\text{maxsep}[d_n^f, \delta, f^{-n}[S(x, \varepsilon, f)]]\}.$$

Recall that a map $f: X \to X$ is **forward-expansive** if for some *expansiveness constant* $c > 0$, every ε-stable set for $0 < \varepsilon \leq c$ is a single point (*i.e.*, $S(x, \varepsilon, f) = \{x\}$ whenever $\varepsilon \leq c$ and $x \in X$). Every one-sided shift, as well as each of the angle-stretching maps on S^1, is forward-expansive. Clearly, for forward-expansive maps,

$$h_p(f) = \sup_{x \in X} h_s(f, x, \varepsilon)$$

whenever $0 < \varepsilon \leq c$. More generally, though, we have

Theorem 3 ([FFN03]) *If X is a compact metric space of finite covering dimension, then for every continuous map $f: X \to X$ and every $\varepsilon > 0$,*

$$\sup_{x \in X} h_s(f, x, \varepsilon) = h_{top}(f).$$

It is possible, adapting an argument of Mañé [Mañ79], to show [FFN03] that forward-expansiveness of $f: X \to X$ implies finite covering dimension for X (if it is compact metric), immediately implying the equality $h_p(f) = h_m(f) = h_{top}(f)$ in this case. Theorem 2 shows that for one-sided shifts, the supremum in Theorem 3 is actually a maximum. This leads us to consider the set of **entropy points** of a continuous map $f: X \to X$, defined as

$$\mathcal{E}(f) := \{x \in X \mid \lim_{\varepsilon \to 0} h_s(f, x, \varepsilon) = h_{top}(f)\}.$$

Points of $\mathcal{E}(f)$ are those near which the local "backward" behavior reflects the topological entropy of f.

How big is the set $\mathcal{E}(f)$ of entropy points for a general map $f:X\to X$? For one-sided subshifts, $\mathcal{E}(f)$ is always nonempty, but there are examples where it is nowhere dense in X, and there are examples of other continuous maps with $\mathcal{E}(f)=\emptyset$ [FFN03]. A number of conditions, given in [FFN03], imply $\mathcal{E}(f)\neq\emptyset$; the most general of these was defined by Misiurewicz (modifying a notion due to Bowen): a continuous map $f:X\to X$ is **asymptotically h-expansive** if

$$\lim_{\varepsilon\to 0}\sup_{x\in X} h_{top}(f, S(x,\varepsilon,f)) = 0.$$

In effect, this says that ε-stable sets for small $\varepsilon>0$ look almost like points from the perspective of topological entropy. We have

Theorem 4 ([FFN03]) *Every asymptotically h-expansive map on a compact metric space has*

$$\mathcal{E}(f)\neq\emptyset.$$

Forward-expansive maps are automatically asymptotically h-expansive, but the latter class is far larger; in particular

Theorem 5 ([Buz97]) *Every C^∞ diffeomorphism of a compact manifold is asymptotically h-expansive.*

5 Branch Preimage Entropy

In formulating the pointwise preimage entropies, one focuses on the preimage sets $f^{-n}[x]$ of individual points. These sets have a natural tree-like structure, with preimage points as "vertices" and an "edge" from $z\in f^{-n}[x]$ to $f(z)\in f^{-(n-1)}[x]$, and one can try to examine the structure of *branches* in this tree—sequences $\{z_i\}$ with $z_0=x$ and $f(z_{i+1})=z_i$ for all i. The idea of the Langevin-Walczak invariant [LW91], which is to compare points $x,x'\in X$ by means of their respective branch structures, was used by Hurley [Hur95] to formulate an invariant that fits our general context and in many natural cases[8] equals that defined by Langevin and Walczak [LW91].

A complication for both formulations is that, if a map is not surjective, some branches may terminate at points with no preimage; to avoid this largely technical distraction, we will assume tacitly that $f:X\to X$ is a surjection.

[8](but not all—see [NP99] for an example)

Recall that for any compact metric space (X, d), there is an associated **Hausdorff metric** $\mathfrak{H}d$ which makes the collection $\mathfrak{H}(X)$ of nonempty closed subsets of X into a compact metric space: for $K_0, K_1 \in \mathfrak{H}(X)$,

$$\mathfrak{H}d(K_0, K_1) := \max_{i=0,1}\{\sup_{x \in K_i}[\inf_{x' \in K_{1-i}} d(x, x')]\}.$$

Given $f : X \to X$ a continuous surjection, we can apply the Hausdorff extension to the Bowen-Dinaburg metrics d_n^f to define a sequence of **branch metrics** on X via

$$d_n^b(x, x') := \mathfrak{H}d_n^f(f^{-n}[x], f^{-n}[x']).$$

That is, $x \in X$ is "branch close" to $x' \in X$ if every branch at x is shadowed by some branch at x', and vice-versa. Applying the usual mechanism to these metrics yields the **branch preimage entropy**

$$h_b(f) := \lim_{\varepsilon \to 0} GR\{\text{maxsep}[d_n^b, \varepsilon, X]\}.$$

Standard arguments apply to show that topologically conjugate maps have equal branch preimage entropy. When f is a homeomorphism, this equals the topological entropy, but in general $h_b(f)$ acts very differently from $h_{top}(f)$—a number of general equalities for $h_{top}(f)$ become inequalities (sometimes strict) for $h_b(f)$ [NP99].

One can think of $h_b(f)$ as measuring the homogeneity of the preimage structure of f. For example, the preimage sets of two points $x, x' \in S^1$ under the angle-doubling map are rotations of each other, yielding $d_n^b(x, x') = d(x, x')$ and hence $h_b(f) = 0$; this argument has a natural extension to any self-covering map $f : X \to X$.

5.1 Branch Preimage Entropy for Subshifts

Suppose that $f : X \to X$ is the restriction of the shift map to some (shift-invariant) closed subset $X \subset \mathfrak{A}^{\mathbb{N}}$. We have already seen that preimage sets can be identified with predecessor sets

$$f^{-n}[x] = \{wx \mid w \in P_n(x, X)\}.$$

Suppose now that $x, x' \in X$ have different $(n+k)^{th}$ predecessor sets, say $w = w_0 ... w_{n+k-1} \in P_{n+k}(x) \setminus P_{n+k}(x')$, which means that $z = wx$ belongs to $f^{-(n+k)}[x]$, but for any $z' \in f^{-(n+k)}[x']$ we have $z' = w'x'$, where $w' = w'_0 ... w'_{n+k-1}$ and $w'_j \neq w_j$ for some $j < n + k$. If we let $i = \min(j, n)$, then the initial k-words of $f^i(z)$ and $f^i(z')$ are distinct, so

$$d_n^f(z, z') \geq 2^{-k},$$

and this shows that whenever $P_{n+k}(x) \neq P_{n+k}(x')$ as sets,

$$d_n^b(x, x') \geq 2^{-k}.$$

But if $w \in P_{n+k}(x) \cap P_{n+k}(x')$ then $z = wx$ and $z' = w'x'$ satisfy $d_n^f(x, x') \leq 2^{-k}$; it follows that

$$\operatorname{maxsep}[d_n^b, 2^{-k}, X] = NP_{n+k}[X].$$

where $NP_m[X]$ denotes the number of distinct m^{th} predecessor sets $P_m(x)$ (as x ranges over X). So we have, for any one-sided subshift $f\colon X \to X$,

$$h_b(f) = \lim_{k\to\infty} GR\{NP_{n+k}[X]\} = GR\{NP_n[X]\}.$$

Here are the details of this calculation for our earlier examples:

Full Shift: Since $P_n(x, \mathfrak{A}^{\mathbb{N}}) = \mathfrak{A}^n$ for all $x \in \mathfrak{A}^{\mathbb{N}}$, $NP_n[\mathfrak{A}^{\mathbb{N}}] = 1$ for all n and

$$h_b(f) = 0.$$

Subshifts of Finite Type: We saw earlier that $P_n(x)$ is determined by x_0, so $NP_n[X] \leq \operatorname{card}[\mathfrak{A}]$ for all n, and

$$h_b(f) = 0.$$

Sofic Subshifts: We saw that the even shift has precisely two distinct n^{th} predecessor sets for each n, so $NP_n[X] = 2$ for all n and $h_b(f) = 0$. In general, a subshift $f\colon X \to X$ is called **sofic** if $NP_n[X]$ has a finite upper bound as $n \to \infty$; Benjamin Weiss [Wei73] showed that $f\colon X \to X$ is sofic precisely if there is a subshift of finite type $g\colon Y \to Y$ and a continuous surjection $p\colon Y \to X$ such that $p \circ g = f \circ p$ (*i.e.*, f is a **factor** of g). All sofic subshifts clearly have

$$h_b(f) = 0.$$

Dyck Shift: Any balanced word can precede any sequence in $\mathfrak{D}_N$: more generally, if $w = ABC \in W_n$ (as in §2.2.3) then, if C is empty, $w \in P_n(x, \mathfrak{D}_N)$ for all $x \in \mathfrak{D}_N$. If $C \neq \emptyset$, the unmatched left delimiters in C must match the first unmatched right delimiters (if any) in x. To be precise, suppose $w \in W_n$ has $m \geq 0$ unmatched left delimiters, $\ell_{j_1}, \ldots, \ell_{j_m}$ (reading left-to-right in w) and $x \in \mathfrak{D}_N$ has $0 \leq p \leq \infty$ unmatched delimiters; let $q = \min(m, p) \leq n$ and suppose the first q unmatched right delimiters in x are $r_{s_0},...,r_{s_q}$ (reading left-to-right in x). Then $w \in P_n(x)$ precisely if the indices match, moving in opposite directions in x and w:

$$s_i = j_{m-i} \text{ for } 0 \leq i < q.$$

This shows that the predecessor set $P_n(x)$ is determined by the indices of the first n (or fewer, if x has fewer) unmatched right delimiters in x. $NP_n[\mathfrak{D}_N]$ thus equals the number of sequences of length n or less of indices from $\{1,\ldots,N\}$, or

$$NP_n[\mathfrak{D}_N] = \sum_{i=0}^{n} N^i \leq (n+1)N^n$$

which has growth rate

$$h_b(f,\mathfrak{D}_N) = GR\{(n+1)N^n\} = \log N.$$

(For comparison, recall that $h_{top}(f,\mathfrak{D}_N) = \log(N+1)$.)

Square-Free Sequences We show, as in [NP99], that if $\mathfrak{A}$ is an alphabet on six or more letters then the shift $f: X \to X$ on square-free sequences in $\mathfrak{A}$ has infinite branch preimage entropy.

Pick three distinguished letters from $\mathfrak{A}$, and

$$\beta = b_0 b_1 b_2 ...$$

a square-free sequence in just these three letters. The complement $\mathfrak{A}^*$ of these letters in $\mathfrak{A}$ still has at least three letters, so we have the nonempty subset $X^* \subset X$ of square-free sequences which have no letter in common with β.

We will produce, for every subset $E \subset W_n(X^*)$ of square-free words in $\mathfrak{A}^*$, a sequence $x_E \in X$ whose predecessor set in X intersects $W_n(X^*)$ precisely in E:

$$P_n(x_E, X) \cap (\mathfrak{A}^*)^n = P_n(x_E, X) \cap W_n(X^*) = E.$$

When $E = W_n(X^*)$, $x_E = \beta$ works, since for $A \in W_n(X^*)$ the sequence $A\beta$ is square-free. Otherwise, we work with the complementary set of words

$$F := W_n(X^*) \setminus E = \{A_0, A_1, \ldots, A_k\}.$$

Our sequence will have the form

$$x_E = W_E \beta$$

where the initial word W_E is designed so that $W_E b_0$ has no squares, but $AW_E b_0$ has a square if $A \in F$ and not if $A \in E$. If $W_E b_0$ has no squares, then $AW_E b_0$ has a square if $W_E b_0$ has an initial word of the form $\mathtt{w}A\mathtt{w}$, where $\mathtt{w}$ is some (possibly empty) word. We construct W_E using induction on the cardinality of F. Set

$$\mathtt{w}_0 = b_0$$

and note that

$$W_0 = b_0 A_0$$

leads to $W_0 b_0 = \mathtt{w}_0 A \mathtt{w}_0$, so any sequence beginning with $W_0 b_0$ cannot be preceded by A_0. If $F = \{A_0\}$, then $W_E = W_0$ gives us the desired sequence in the form

$$x_E = W_E \beta = W_0 \beta = b_0 A_0 b_0 b_1 \ldots .$$

To also exclude a second word A_1, we use

$$\mathtt{w}_1 = W_0 b_0$$

and

$$W_1 = \mathtt{w}_1 A_1 W_0 = b_0 A_0 b_0 A_1 b_0 A_0 .$$

If $F = \{A_0, A_1\}$, then $x_E = W_1 \beta$ has $\mathtt{w}_0 A_0 \mathtt{w}_0$ and $\mathtt{w}_1 A_1 \mathtt{w}_1$ as initial words, but no other word of $W_n(X^*)$ appears (anywhere) in x_E; furthermore, it is easy to check that x_E is square-free.

Inductively, if for $j = 1, \ldots, k-1$ we set

$$\begin{array}{rcl} \mathtt{w}_j & := & W_{j-1} b_0 \\ W_j & := & \mathtt{w}_j A_j W_{j-1} \end{array}$$

it is easy to check that each word $W_j b_0$ is square-free, has $x_i A_i \mathtt{w}_i$ as an initial word for $i = 0, \ldots, j$, and contains no word in $W_n(X^*) \setminus \{A_0, ..., A_j\}$. It follows that

$$x_E := W_k \beta$$

has the required properties.

This shows that the number $NP_n[X]$ of distinct n^{th} predecessor sets for X is at least the number of distinct *subsets* of $W_n(X^*)$, or 2^{w_n} (where $w_n = \text{card}[W_n(X^*)]$). But we know that w_n has positive exponential growth rate (since X^* has positive topological entropy), and hence

$$h_b(f) = GR\{NP_n[X]\} \geq GR\{2^{w_n}\} = \left(\limsup_{n \to \infty} \frac{w_n}{n} \right) \cdot \log 2 = \infty .$$

5.2 Hurley's Inequalities

The main result of Hurley's paper [Hur95] is a beautiful inequality relating pointwise, branch and topological entropy:

Theorem 6 ([Hur95]) *For any continuous map $f : X \to X$ on a compact metric space,*

$$h_m(f) \leq h_{top}(f) \leq h_m(f) + h_b(f).$$

In particular, for any map with branch preimage entropy zero, pointwise preimage entropy automatically agrees with topological entropy. We have seen that this occurs for subshifts of finite type and more generally for sofic subshifts, but for other subshifts Theorem 2 appears to provide the only proof that $h_m(f) = h_{top}(f)$.

Several other classes of maps are known to have $\mathbf{h_b(f) = 0}$ (and hence $h_m(f) = h_{top}(f)$):

- A **forward-expansive** map on a compact **manifold** is automatically a self-covering map [HR69] and so has branch entropy zero (as noted earlier in this section).
- Any **rational map** $f(z) = \frac{p(z)}{q(z)}$ (p, q polynomials) on the Riemann sphere has zero branch preimage entropy [LP92].
- If X is homeomorphic to a **finite graph** (including the interval and circle) then every continuous map $f : X \to X$ has branch preimage entropy zero [NP99].

5.3 Natural Extensions

Given $f : X \to X$ a continuous map on a compact space, define the space

$$\hat{X} = \hat{X}_f := \{\hat{x} = ...x_{-1}x_0x_1... \in X^{\mathbb{Z}} \mid f(x_i) = x_{i+1} \text{ for all } i \in \mathbb{Z}\}$$

(with the induced product topology) and the projection $\pi : \hat{X} \to X$ via

$$\pi(\hat{x}) = x_0.$$

The image of the projection is the eventual range of f

$$\pi[\hat{X}] = \bigcap_{i=0}^{\infty} f^i[X]$$

which is homeomorphic to the quotient space $\hat{X}/\pi$. The shift map $\hat{f}:\hat{X}\to\hat{X}$

$$[\hat{f}(\hat{x})]_i = x_{i+1}, \quad i \in \mathbb{Z}$$

is a homeomorphism called the **natural extension** (or *inverse limit*) of $f:X\to X$. In effect, $\hat{X}_f$ separates the various prehistories of points; note that for $\hat{x}\in\hat{X}$, $x_0=\pi(\hat{x})$ determines all x_i with $i\geq 0$.

The natural extension of the angle-doubling map can be identified with the "solenoid" of Smale [Shu86, 4.9], [KH95, 17.1], while the natural extension of a one-sided subshift $X\subset\mathfrak{A}^{\mathbb{N}}$ is the two-sided subshift $\hat{X}\subset\mathfrak{A}^{\mathbb{Z}}$ specified by the same list of disallowed words. In general, $h_{top}(\hat{f})=h_{top}(f)$.

Of course, topologically conjugate maps have topologically conjugate natural extensions, but the converse is not always true. The following example was shown to me by Bob Burton.

Consider the *coding* $\varphi:\mathfrak{A}^2\to\mathfrak{B}$ which assigns to each word $w\in\mathfrak{A}^2$ of length 2 in the alphabet $\mathfrak{A}=\{0,1\}$ a letter $\varphi(w)\in\mathfrak{B}$ in the alphabet $\mathfrak{B}=\{1,2,3\}$ via

$$\begin{aligned}\varphi(01)&=1\\ \varphi(11)&=2\\ \varphi(00)=\varphi(10)&=3.\end{aligned}$$

Any such coding induces a continuous map $\hat{h}:\mathfrak{A}^{\mathbb{Z}}\to\mathfrak{B}^{\mathbb{Z}}$ via $\hat{h}(\hat{x})=\hat{y}$, where

$$y_i=\varphi(x_{i-1}x_i).$$

The image $\hat{h}[\mathfrak{A}^{\mathbb{Z}}]$ is the subshift $\hat{X}\subset\mathfrak{B}^{\mathbb{Z}}$ with the transition matrix

$$A=\begin{bmatrix}0&1&1\\0&1&1\\1&0&1\end{bmatrix}.$$

Furthermore, y_i determines x_i, so $\hat{h}$ is a homeomorphism between $\mathfrak{A}^{\mathbb{Z}}$ and $\hat{X}\subset\mathfrak{B}^{\mathbb{Z}}$ which conjugates the shift maps on these spaces.

However, the one-sided subshift $f:X\to X$ defined by the transition matrix A cannot be conjugated to the (full) shift on $\mathfrak{A}^{\mathbb{N}}$, because for $y=y_0y_1\ldots\in X$, $f^{-1}[y]$ has cardinality numerically equal to $y_0\in\{1,2,3\}$, while every $x\in\mathfrak{A}^{\mathbb{N}}$ has precisely two preimages.

The two one-sided subshifts are both of finite type, so automatically satisfy $h_b(f)=0$. But more generally, the following is true:

Theorem 7 *If $f : X \to X$ and $g : Y \to Y$ are both forward-expansive with topologically conjugate natural extensions $\hat{f} : \hat{X} \to \hat{X}$, $\hat{g} : \hat{Y} \to \hat{Y}$, then*

$$h_b(f) = h_b(g).$$

This theorem was first conjectured by Bob Burton, with whom I unsuccessfully sought a proof several years ago. I know of two arguments for this fact, both unpublished. One proceeds by analyzing the structure of conjugacies between natural extensions (which for forward-expansive maps come from a kind of generalized coding) and using it to estimate the growth rate of maxsep$[d_n^b, \varepsilon, X]$ for $\varepsilon < c$. The other is based on "lifting" $h_b(f)$ to $\hat{f}$ by a trick similar to our replacement of points with local stable sets in §4. Unlike the situation there, the resulting quantity has not been shown invariant under conjugacy of $\hat{f}$, except when f is forward-expansive. Both arguments are due to Doris and Ulf Fiebig, with some contribution on my part to the first one.

6 Pressure and Hausdorff Dimension

In the context of an abstract "thermodynamic formalism" for dynamical systems, Ruelle [Rue73, Rue78] modified the concept of topological entropy, replacing the number maxsep$[d_n^f, \varepsilon, X]$ of n-orbit segments with a "weighted" count, the weights coming from a function φ, to get the *topological pressure* of φ with respect to f. To be precise[9] given $f : X \to X$ a continuous map and $\varphi : X \to \mathbb{R}$ a continuous real-valued function, the sum of φ along the n-orbit segment starting at $x \in X$ is denoted

$$S_n\varphi(x) := \sum_{i=0}^{n-1} \varphi(f^i(x))$$

and for $\varepsilon > 0$ we consider

$$N(f, \varphi, \varepsilon, n) := \sup_E \sum_{x \in E} e^{S_n\varphi}$$

the supremum taken over all (n, ε)-separated sets in X. The **topological pressure** of φ with respect to f is then

$$P_f(\varphi) := \lim_{\varepsilon \to 0} GR\{N(f, \varphi, \varepsilon, n)\}.$$

[9]We loosely follow [KH95, §20.2], which together with [Wal82, Chap. 9] is a good reference for details.

It can be shown that $P_f(\varphi)$ is either always finite or always infinite for all $\varphi \in \mathcal{C}(X)$, the space of continuous real-valued functions on X, and when finite $P_f : \mathcal{C}(X) \to \mathbb{R}$ is monotone, convex and continuous. It is also clear that the topological pressure of the constant zero function is the topological entropy:

$$P_f(0) = h_{top}(f).$$

There is a fascinating connection between topological pressure and the Hausdorff dimension of certain invariant sets. This connection was first noted, in the context of Fuchsian groups, in Bowen's last paper [Bow79] (published posthumously), and is generally referred to as *Bowen's formula*. For any strictly negative $\varphi \in \mathcal{C}(X)$, the function $t \mapsto P_f(t \cdot \varphi)$ has a unique zero t_φ. Ruelle showed [Rue82] that if f is $C^{1+\alpha}$ and J is a conformal repellor (J is the closure of some recurrent f-orbit, and the derivative multiplies the length of all vectors at $x \in J$ by a factor $\alpha(x)$, where $\alpha(x) > 1$ for all $x \in J$) then the Hausdorff dimension $HD(J)$ of J equals t_φ, where $\varphi(x) = -\log \alpha(x)$.

Analogous results for saddle sets of surface diffeomorphisms were obtained by Manning *et al* [Man81, MM83]. A *saddle set* for a diffeomorphism of a surface is an invariant set Λ such that at each $x \in \Lambda$ there exist two independent vectors $v_+, v_- \in T_x\Lambda$ with $\| Df^n(v_\pm) \|$ going to zero at a (uniform) exponential rate as $n \to \pm\infty$. Every point $x \in \Lambda$ then has an invariant curve $W^s(x)$ (its *stable manifold*) which goes through x tangent to v_+. The prototype of this is the Smale "horseshoe" ([Shu86, KH95]), where $v_\pm$ are coordinate vectors. The **stable dimension** at $x \in \Lambda$ of a saddle set Λ is the Hausdorff dimension of the intersection of Λ with the stable manifold of x:

$$sd(\Lambda, x) := HD(\Lambda \cap W^s(x)).$$

If we define $\phi^s \in \mathcal{C}(X)$ by

$$\phi^s(x) := \log \| Df(v_+) \|$$

then, under a few mild technical assumptions[10] we again have [MM83] Bowen's formula

$$sd(\Lambda, x) = t_{\phi^s}.$$

The same formula was obtained for the $\mathbb{C}^2$ version of the Hénon map by Verjovsky and Wu [VW96].

When the map is not invertible, the situation becomes more complicated. Mihailescu [Mih01] showed that in a complex two-dimensional setting, the stable dimension of a saddle set for a holomorphic endomorphism (with no

[10] Λ is a basic set

critical points in the set) has t_{ϕ^s} as an upper bound, but the inequality can be strict. By taking account of the minimum number of preimages of points in Λ, Mihailescu and Urbański [MU01] obtained a better upper bound on $sd(\Lambda)$.

In the same paper [MU01], Mihailescu and Urbański also obtained a lower bound, using a new "entropy" invariant $h_-(f)$ which we shall sketch below; they showed that this invariant, for the restriction of f to Λ, is a lower bound for the stable dimension times the supremum of $|\phi^s|$ on Λ. Subsequently [MU02] they defined two new notions of pressure, $P_f^-(\varphi)$ and $P_{f,-}(\varphi)$ and used Bowen type formulas to obtain lower and upper bounds for stable dimension.

A notion complementary to that of an ε-separated set is an **ε-spanning**[11] set: $E \subset X$ ε-spans X if every point of X is within distance $< \varepsilon$ of some point of E. A (set-theoretically) maximal ε-separated subset of X automatically ε-spans X, and a minimal ε-spanning set is $\frac{\varepsilon}{3}$-separated, so in all of our definitions of "entropy" we could replace $\mathrm{maxsep}[d, \varepsilon, X]$ with the number

$$\mathrm{minspan}[d, \varepsilon, X] := \min\{\mathrm{card}[E] \mid E \subset X \ \varepsilon\text{-spans } X\}.$$

For the Mihailescu-Urbański invariants it is more natural to work with this number.

The difference between $h_{top}(f)$ and $h_b(f)$, when phrased in terms of spanning sets, can be clarified (at least when f is surjective) by noting that each n-branch $z_0, z_1, ...z_{n-1}$ of f^{-1} has a well-defined "root" $x = z_0$ and "tip" $z = z_{n-1} \in f^{-n}[x]$; the latter determines the branch via $f(z_i) = z_{i-1}$. A set $E \subset X$ ε-spans X in the branch metric d_n^b if the collection of branches rooted at points in E, or in terms of "tips", $E_{f,n} := \{f^{-n}[x] \mid x \in E\} \subset \mathfrak{H}(X)$, ε-spans $X_{f,n}$ in the Hausdorff Bowen-Dinaburg metric $\mathfrak{H}d_n^f$—which is to say for any $x \in X$ we can find $x' \in E$ such that every branch rooted at one of x or x' is (n, ε)-shadowed by at least one branch rooted at the other. However, if we consider branches without regard to their roots, merely asking for a collection of branches which includes an (n, ε)-shadow of every branch, we are simply asking for a collection of tips which ε-spans X in the Bowen-Dinaburg metric d_n^f, and so the usual machinery in this case leads to $h_{top}(f)$.

The Mihailescu-Urbański definitions mix these two notions. Let us say that a collection of n-branches **weakly ε-spans n-branches** in X if for any $x \in X$ we can find at least one n-branch at x which is (n, ε)-shadowed by one from our collection. Looking at "tips", this amounts to saying we have a collection $E' \subset X$ of tips such that the *minimum* Bowen-Dinaburg distance d_n^f of any preimage set $f^{-n}[x]$, $x \in X$ from our set E' is at most ε. Denote the minimum cardinality of a set E' which weakly ε-spans n-branches in X

[11] The phrase *ε-dense* denotes the same idea.

by $w[f,n,\varepsilon,X]$, and let

$$h_w(f) := \lim_{\varepsilon\to 0} GR\{w[f,n,\varepsilon,X]\}.$$

Note that since any set which (n,ε) spans X also weakly ε-spans n-branches in X, we have

$$w[f,n,\varepsilon,X] \le \text{minspan}[d_n^f,\varepsilon,X]$$

so

$$h_w(f) \le h_{top}(f).$$

Going further, we say that a collection $E \subset X$ (of "roots") **very weakly ε-spans n-branches** in X if the collection of all branches rooted at points of E weakly ε-spans n-branches in X. The minimum cardinality of a set which very weakly ε-spans n-branches in X, which we will denote $v[f,n,\varepsilon,X]$, is bounded above by $w[f,n,\varepsilon,X]$, since if E' is the set of "tips" for a weakly ε-spanning set of n-branches, then the corresponding set $E = f^n[E']$ of "roots" is a very weakly ε-spanning set with cardinality less than or equal to card$[E']$. Thus, the "entropy" defined using $v[f,n,\varepsilon,X]$,

$$h_v(f) := \lim_{\varepsilon\to 0} GR\{v[f,n,\varepsilon,X]\}$$

satisfies

$$h_v(f) \le h_w(f) \le h_{top}(f).$$

Furthermore, any set which ε-spans X in the branch metric d_n^b also weakly ε-spans n-branches in X, so

$$v[f,n,\varepsilon,X] \le \text{minspan}[d_n^b,\varepsilon,X]$$

which implies

$$h_v(f) \le h_b(f).$$

To define the corresponding notions of pressure, we set, for $f: X \to X$ and $\varphi \in \mathcal{C}(X)$,

$$P_f^-(\varphi) := \lim_{\varepsilon\to 0} GR\{\inf_{E'} \sum_{z\in E'} e^{S_n\varphi(z)}\}$$

where the infimum is taken over sets E' of "tips" for collections which weakly ε-span n-branches in X, and

$$P_{f,-}(\varphi) := \lim_{\varepsilon\to 0} GR\{\inf_{E} \sum_{x\in E} \min_{z\in f^{-n}[x]} e^{S_n\varphi(z)}\}$$

where the infimum is taken over sets E (of "roots") which very weakly ε-span n-branches in X.

It can be shown [MU02] that these are invariant in the sense that if $f:X\to X$ and $g:Y\to Y$ are maps conjugated by the homeomorphism $h:X\to Y$ ($h\circ f = g\circ h$), then for any $\varphi\in\mathcal{C}(X)$,

$$P_f^-(\varphi) = P_g^-(\varphi\circ h^{-1})$$
$$P_{f,-}(\varphi) = P_{g,-}(\varphi\circ h^{-1})$$

Note that when φ is the constant zero function, then $e^{S_n\varphi(z)} = 1$ for all $z\in X$ and $n\in\mathbb{N}$, so

$$P_f^-(0) = h_w(f)$$
$$P_{f,-}(0) = h_v(f).$$

The invariance of pressure implies the invariance of these "entropies"; in [MU01, MU02] h_v (*resp.* h_w) is denoted h_- (*resp.* h^-).

The bounds on stable dimension given by Mihailescu-Urbański can then be stated as follows:

Theorem 8 ([MU02]) *Suppose f is a holomorphic Axiom A map of $\mathbb{P}^2$ and Λ is a basic saddle set for f with no critical points of f. Let*

$$\phi^s(x) := \log \parallel Df(v_+) \parallel$$

where v_+ is the "contracting" vector at $x\in\Lambda$, and denote by t^s (resp. t^s_-) the (unique) zero of the function $t\mapsto P_f^-(t\cdot\phi^s)$ (resp. $t\mapsto P_{f,-}(t\cdot\phi^s)$).

Then for all $x\in\Lambda$

$$t^s_- \leq sd(\Lambda, x) \leq t^s.$$

7 Other Directions

I would like to close with some brief speculative comments on two other possible directions of study in the spirit of preimage entropy:

Variational Principle: The relation between measure-theoretic and topological entropy given by Theorem 1 has an extension to topological pressure [Rue73, Wal76, Mis76]:

Theorem 9 (Variational Principle) *For any continuous map $f:X\to X$ on a compact metric space and any $\varphi\in\mathcal{C}(X)$,*

$$P_f(\varphi) = \sup_\mu \left\{ h_\mu(f) + \int \varphi \, d\mu \right\}$$

where the supremum is taken over all f-invariant Borel probability measures μ.

It is natural to ask whether there is an analogue of this for preimage entropy: one needs to find an appropriate version of pressure and of measure-theoretic entropy, probably based on the branch structure of preimages. Mihailescu and Urbański have some ideas and results in this direction.

As this survey was going to press, I learned of important new results related to the restricted variational principle (Theorem 1) by Cheng and Newhouse [CN].

Cheng and Newhouse define two new kinds of "preimage entropy" invariants. The first can be regarded as a modification of the pointwise preimage entropy $h_m(f)$ sketched in §3. Instead of looking at (n,ε)-separated sets in the n^{th} preimage sets of points, they look inside all possible k^{th} preimage sets, either for $k \geq n$ or for all $k \geq 1$:

$$\begin{aligned} h_{pre}(f) &:= \lim_{\varepsilon\to 0} GR\{\max_{k\geq n}\max_{x\in X} \text{maxsep}[d_n^f, \varepsilon, f^{-k}[x]]\} \\ h'_{pre}(f) &:= \lim_{\varepsilon\to 0} GR\{\max_{k\geq 1}\max_{x\in X} \text{maxsep}[d_n^f, \varepsilon, f^{-k}[x]]\}. \end{aligned}$$

Clearly for any map

$$h_m(f) \leq h_{pre}(f) \leq h'_{pre}(f) \leq h_{top}(f).$$

A second class of invariants defined in [CN] is measure-theoretic. Denoting by $\mathcal{B}$ the Borel σ-algebra, note that the preimage map $A \mapsto f^{-1}[A]$ is a Boolean endomorphism of $\mathcal{B}$; its eventual range is the "infinite past" σ-algebra

$$\mathcal{B}^- := \bigcap_{k\geq 0} f^{-k}[\mathcal{B}].$$

A standard procedure [Pet83, §5.2] is to condition the entropy on a subalgebra: given a finite partition $\mathcal{P}$, and fixing an f-invariant Borel probability measure μ, denote by p_i^- the conditional probability of the i^{th} atom, given $\mathcal{B}^-$. Then the uncertainty about the position relative to $\mathcal{P}$, given the infinite past $\mathcal{B}^-$, is

$$H(\mathcal{P}, \mathcal{B}^-) := -\sum_{i=0}^{N} p_i \log p_i^-.$$

Using this in place of $H(\mathcal{P})$ as in §1 we obtain Cheng-Newhouse's "preimage entropy of f with respect to μ and $\mathcal{B}^-$"

$$h_{pre,\mu}(f) := \sup\{\lim_{n\to\infty} \frac{1}{n} H(\mathcal{P}_n, \mathcal{B}^-) \,|\, \mathcal{P} \text{ a finite measurable partition of } X\}.$$

Cheng-Newhouse obtain a number of basic properties of this invariant, such as affineness with respect to μ and a Shannon-Breiman-McMillan theorem, which they use to prove the following preimage analogue of Theorem 1:

Theorem 10 (Restricted Variational Principle for Preimage Entropy, [CN]) *For $f : X \to X$ any continuous map on a compact metric space,*

$$h_{pre}(f) = h'_{pre}(f) =$$
$$\sup\{h_{pre,\mu}(f) \,|\, \mu \textit{ is an } f\textit{-invariant Borel probability measure on } X\}.$$

Semigroup Actions: The dynamics of a single map $f : X \to X$ can be viewed as an action of the semigroup $\mathbb{N}$ on X. Andrzej Biś [Biś02] has formulated analogues of the various preimage entropies in the context of an action of any finitely-generated semigroup of continuous maps on a compact metric space. One might speculate that a combination of these ideas with those of Mihailescu and Urbański might yield more general results on the dimension of fractals defined by iterated function systems.

References

[AKM65] Roy L. Adler, A. G. Konheim, and M. H. McAndrew, *Topological entropy*, Transactions, American Mathematical Society **114** (1965), 309–319.

[Biś02] Andrzej Biś, *Entropies of a semigroup of maps*, preprint, Univ. Łódź, Poland, 2002.

[Bow71] Rufus Bowen, *Entropy for group endomorphisms and homogeneous spaces*, Transactions, American Mathematical Society **153** (1971), 401–414, erratum, **181**(1973) 509-510.

[Bow78] ______, *On axiom A diffeomorphisms*, CBMS Regional Conference Series in Mathematics, no. 35, American Mathematical Society, 1978.

[Bow79] ______, *Hausdorff dimension of quasi-circles*, Publ. Math. IHES **50** (1979), 11–26.

[Bri63] Jan Brinkhuis, *Non-repetitive sequences on three symbols*, Quarterly Journal of Math., Oxford **34** (1963), 145–149.

[Buz97] Jérôme Buzzi, *Intrinsic ergodicity of smooth interval maps*, Israel Journal of Mathematics **100** (1997), 125–161.

[CN] Wen-Chiao Cheng and Sheldon Newhouse, *Pre-image entropy*, preprint, 2003.

[Din70] E. I. Dinaburg, *The relation between topological entropy and metric entropy*, Soviet Math. Dokl. **11** (1970), 13–16.

[FFN03] Doris Fiebig, Ulf Fiebig, and Zbigniew Nitecki, *Entropy and preimage sets*, Ergodic Theory and Dynamical Systems (2003), to appear.

[Goo69] L. Wayne Goodwyn, *Topological entropy bounds measure-theoretic entropy*, Proceedings, American Mathematical Society **23** (1969), 679–688.

[Goo71] T. N. T. Goodman, *Relating topological entropy and and measure entropy*, Bulletin, London Mathematical Society **3** (1971), 176–180.

[Gri01] Uwe Grimm, *Improved bounds on the number of ternary square-free words*, J. Integer Sequences (electronic) **4** (2001), article no. 01.2.7.

[HR69] E. Hemmingsen and William L. Reddy, *Expansive homeomorphisms on homogeneous spaces*, Fundamenta Mathematica **64** (1969), 203–207.

[Hur95] Mike Hurley, *On topological entropy of maps*, Ergodic Theory and Dynamical Systems **15** (1995), 557–568.

[KH95] Anatole Katok and Boris Hasselblatt, *Introduction to the modern theory of dynamical systems*, Encyclopedia of Mathematics and its Applications, vol. 54, Cambridge Univ. Press, London and New York, 1995.

[Khi57] Alexander I. Khinchin, *The entropy concept in probability theory*, Mathematical Foundations of Information Theory, Dover, New York, 1957, (transl. R. A. Silverman).

[Kol58] Andrei N. Kolmogorov, *A new metric invariant of transitive dynamical systems and automorphisms of lebesgue spaces*, Dokl. Akad. Nauk SSSR **119** (1958), 861–864, English translation: Proc. Steklov Inst. **169**(1986) 97-102.

[Kri72] Wolfgang Krieger, *On the uniqueness of the equilibrium state*, Mathematical Systems Theory **8** (1972), 97–104.

[LM95] Douglas Lind and Brian Marcus, *An introduction to symbolic dynamics and coding*, Cambridge Univ. Press, London and New York, 1995.

[LP92] Rémi Langevin and Feliks Przytycki, *Entropie de l'image inverse d'une application*, Bulletin de la Société Mathématique de France **120** (1992), 237–250.

[LW91] Rémi Langevin and Paweł Walczak, *Entropie d'une dynamique*, Comptes Rendus, Acad. Sci. Paris **312** (1991), 141–144.

[Mañ79] Ricardo Mañé, *Expansive homeomorphisms and topological dimension*, Transactions, American Mathematical Society **252** (1979), 313–319.

[Man81] Anthony Manning, *A relation between Lyapunov exponents, Hausdorff dimension and entropy*, Ergodic Theory and Dynamical Systems **1** (1981), 451–459.

[Mih01] Eugen Mihailescu, *Applications of thermodynamic formalism in complex dynamics on* $\mathbb{P}^2$, Discrete and Continuous Dyn. Syst. **7** (2001), 821–836.

[Mis76] Michał Misiurewicz, *A short proof of the variational principle for a* $\mathbb{Z}_+^{\mathbb{N}}$ *action on a compact space*, astérisque **40** (1976), 147–187.

[MM83] H. McCluskey and Anthony Manning, *Hausdorff dimension for horseshoes*, Ergodic Theory and Dynamical Systems **3** (1983), 251–260.

[MU01] Eugen Mihailescu and Mariusz Urbański, *Estimates for the stable dimension for holomorphic maps*, preprint, 2001.

[MU02] ———, *Inverse topological pressure with applications to holomorphic dynamics of several complex variables*, preprint, 2002.

[NP99] Zbigniew Nitecki and Feliks Przytycki, *Preimage entropy for mappings*, International Journal of Bifurcation and Chaos **9** (1999), 1815–1843.

[Orn70] Donald S. Ornstein, *Bernoulli shifts with the same entropy are isomorphic*, Advances in Mathematics **4** (1970), 337–352.

[Orn74] ______, *Ergodic theory, randomness, and dynamical systems*, Yale Mathematical Monographs, vol. 5, Yale Univ. Press, New Haven and London, 1974.

[Pet83] Karl Petersen, *Ergodic theory*, Cambridge Studies in Advanced Mathematics, vol. 2, Cambridge Univ. Press, 1983.

[Rue73] David Ruelle, *Statistical mechanics on a compact set with $\mathbb{Z}^\nu$ action satisfying expansiveness and specification*, Transactions, American Mathematical Society **185** (1973), 237–251.

[Rue78] ______, *Thermodynamic formalism: The mathematical structures of classical equilibrium statistical mechanics*, Encyclopedia of Mathematics and its Applications, vol. 5, Addison-Wesley, Reading, MA, 1978.

[Rue82] ______, *Repellers for real analytic maps*, Ergodic Theory and Dynamical Systems **2** (1982), 99–107.

[Sha63] Claude E. Shannon, *The mathematical theory of communication*, The Mathematical Theory of Communication, Univ. Illinois Press, 1963, pp. 3–91.

[She81a] Robert Shelton, *Aperiodic words on 3 symbols, I*, J. Reine und Angewandte Mathematik **321** (1981), 195–201.

[She81b] ______, *Aperiodic words on 3 symbols, II*, J. Reine und Angewandte Mathematik **327** (1981), 1–11.

[Shu86] Michael Shub, *Global stability of dynamical systems*, Springer, New York and Berlin, 1986.

[Sin59] Yakov G. Sinai, *On the notion of entropy of dynamical systems*, Dokl. Akad. Nauk SSSR **124** (1959), 768–771.

[SS82] Robert Shelton and Raj P. Soni, *Aperiodic words on 3 symbols, III*, J. Reine und Angewandte Mathematik **330** (1982), 44–52.

[VW96] Alberto Verjovsky and H. Wu, *Hausdorff dimension of Julia sets of complex Hénon maps*, Ergodic Theory and Dynamical Systems **16** (1996), 849–861.

[Wal76] Peter Walters, *A variational principle for the pressure of continuous transformations*, American Journal of Mathematics **17** (1976), 937–971.

[Wal82] ______, *An introduction to ergodic theory*, Graduate Texts in Mathematics, no. 79, Springer, New York and Berlin, 1982.

[Wei73] B. Weiss, *Subshifts of finite type and sofic systems*, Monatshefte für Mathematik **77** (1973), 462–474.