*Ergod. Th. & Dynam. Sys.* (0), **0**, 1–29
*Printed in the United Kingdom* 

# On the Topological Entropy of Saturated Sets

C.-E. Pfister† and W.G. Sullivan‡

† *EPF-L, Institut d'analyse et calcul scientifique CH-1015 Lausanne, Switzerland*
*(e-mail: charles.pfister@epfl.ch)*
‡ *School of Mathematical Sciences, UCD Dublin, Belfield, Dublin 4, Ireland*
*(e-mail: Wayne.Sullivan@ucd.ie)*

(*Received 18 November 2005*)

*Abstract.* Let $(X, d, T)$ be a dynamical system, where $(X, d)$ is a compact metric space and $T : X \to X$ a continuous map. We introduce two conditions for the set of orbits, called respectively g-*almost product property* and *uniform separation property*. The g-almost product property holds for dynamical systems with the specification property, but also for many others. For example all $\beta$-shifts have the g-almost product property. The uniform separation property is true for expansive and more generally asymptotically $h$-expansive maps. Under these two conditions we compute the topological entropy of saturated sets. If the uniform separation condition does not hold, then we can compute the topological entropy of the set of generic points, and show that for *any* invariant probability measure $\mu$, the (metric) entropy of $\mu$ is equal to the topological entropy of generic points of $\mu$. We give an application of these results to multi-fractal analysis and compare our results with those of Takens and Verbitskiy (Ergod. Th. & Dynam. Sys. **23** (2003)).

## 1. *Introduction*

In this paper a dynamical system $(X, d, T)$ means always that $(X, d)$ is a compact metric space and $T : X \to X$ is a continuous map. The set $M(X)$ of all (Borel) probability measures is a compact space for the weak*-topology of measures, and $M(X, T)$ is the subset of $T$-invariant probability measures with the induced topology. We are interested in comparing the (metric) entropy of $\mu \in M(X, T)$ with the topological entropy, which is a measure of the complexity of the dynamical system. In order to do that we study time-averages along orbits and introduce the empirical measure (of order $n$) of $x \in X$, which is the probability measure

$$\mathcal{E}_n(x) := \frac{1}{n}\sum_{k=0}^{n-1} \delta_{T^k x}\,,$$

where $\delta_y$ is the Dirac mass at $y \in X$. A subset $D \subset X$ is called saturated if $x \in D$ and the sequences $\{\mathcal{E}_n(x)\}$ and $\{\mathcal{E}_n(y)\}$ have the same limit-point set, then $y \in D$. The limit-point set of $\{\mathcal{E}_n(x)\}$ is always a compact connected subset $V(x) \subset M(X,T)$. Of particular interest are the points $x \in X$ such that $V(x) = \{\mu\}$. For such points, the time-average along the orbit of $x$ of any continuous function $f$ is equal to the average of $f$ with respect to $\mu$ (space-average)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j x) = \int_X f \, d\mu \,.$$

These points are called generic points of $\mu$. By $G_\mu$ we denote the saturated set of generic points of $\mu$, and more generally, if $K \subset M(X,T)$ is a compact connected subset, we denote by $G_K$ the saturated set of points $x$ such that $V(x) = K$. In [**Bo2**] Bowen proved the remarkable result that

$$h_{\rm top}(T, G_\mu) = h(T,\mu) \quad \text{if } \mu \text{ is ergodic.} \tag{1}$$

Here $h_{\rm top}(T, G_\mu)$ is the topological entropy of $\mu$ and $h(T,\mu)$ is the metric (or Kolmogorov-Sinai) entropy of $\mu$. These concepts are quite different. The notion of metric entropy is affine while the Bowen entropy is a dimension-like characteristic. Generalization and extension of Bowen's result are given in [**PePi**]. Essential in these papers is the fact that one deals with large sets in the measure theoretic sense. Indeed the Ergodic Theorem implies that $\mu$ is ergodic if and only if $\mu(G_\mu) = 1$, while for non-ergodic measures $\nu$ one always has $\nu(G_\nu) = 0$. For non-ergodic measures the situation is quite different. It is clear that (1) cannot hold for general $\mu \in M(X,T)$ since it is not difficult to provide examples with $G_\mu = \emptyset$ and $h(T,\mu) > 0$. However, the following inequalities hold (Theorem 4.1) for $\mu \in M(X,T)$,

$$h_{\rm top}(T, G_\mu) \leq h(T,\mu) \,, \tag{2}$$

and for $K \subset M(X,T)$, a compact connected set,

$$h_{\rm top}(T, G_K) \leq \inf\{h(T,\mu) \colon \mu \in K\} \,. \tag{3}$$

Formulas (2) and (3) are subcases of Bowen's variational principle.

To treat non-ergodic measures we introduce two conditions. The first one is a weaker form of the notion of specification, which is verified for a large class of interesting and important systems. The new condition (see Section 2), called g-*almost product property*, is inspired by our previous works [**PfS1**] and in particular [**PfS2**], where we introduced a similar condition for deriving large deviations estimates. When dealing with the topological entropy we had to strengthen a little bit the property introduced in [**PfS2**], but the principal feature is retained: while for the specification property one requires the existence of an orbit, which $\varepsilon$-shadows all specified orbit-segments, we only require the existence of an orbit, which *partially* $\varepsilon$-shadows the specified orbit-segments (see Section 2). In the case of $\beta$-shifts it is known that the specification property holds for a set of $\beta$ of Lebesgue measure 0 (see [**Sc**]), whereas the g-almost product property always holds (see Section 2). To deal with the g-almost product property we must consider the notion introduced in [**PfS2**] of a $(\delta, n, \varepsilon)$-separated subset. For $\delta > 0$ and $\varepsilon > 0$, two points $x$ and $y$ are $(\delta, n, \varepsilon)$-separated if

$$\mathrm{card}\{j : d(T^j x, T^j y) > \varepsilon\,,\ 0 \leq j \leq n-1\} \geq \delta n\,.$$

A subset $E$ is $(\delta, n, \varepsilon)$-separated if any pair of different points of $E$ are $(\delta, n, \varepsilon)$-separated. The reason for introducing this stronger separation property comes from the basic principle for error-correcting codes: in order to correct errors one introduces more information then a priori necessary. Here the stronger form of separation compensates the fact that we require only partial shadowing. We proved in [**PfS2**] the following proposition. Let $F \subset M(X)$ be a neighborhood; we set

$$X_{n,F} := \{x \in X : \mathcal{E}_n(x) \in F\}\,.$$

PROPOSITION 1.1. *Let $\nu \in M(X,T)$ be ergodic and $h^* < h(T,\nu)$. Then there exist $\delta^* > 0$ and $\varepsilon* > 0$ so that for each neighborhood $F$ of $\nu$ in $M(X)$, there exists $n^*_{F,\nu} \in \mathbf{N}$ such that for any $n \geq n^*_{F,\nu}$, there exists a $(\delta^*,\, n,\, \varepsilon^*)$-separated set $\Gamma_n$, such that*

$$\Gamma_n \subset X_{n,F} \quad \textit{and} \quad |\Gamma_n| \geq 2^{nh^*}\,.$$

The second condition is introduced in Section 3; it is called *uniform separation property.* It states that the affirmation of Proposition 1.1 holds uniformly, that is, $\delta^*$ and $\varepsilon^*$ can be chosen independently of the ergodic measure $\nu$. Expansive dynamical systems, and more generally asymptotic $h$-expansive dynamical systems ([**Bo1**] and [**Mi1**]), have the uniform separation property (Theorem 3.1). Let $F \subset M(X)$ be a neighborhood of $\nu$, and $\varepsilon > 0$. We define

$$N(F; n, \varepsilon) := \text{maximal cardinality of a } (n,\varepsilon)\text{-separated subset of } X_{n,F}\,.$$

Following Kolmogorov, the quantity $N(F; n, \varepsilon)$ is the $(n, \varepsilon)$-capacity of the set $X_{n,F}$ (see [**W**] p.170). Let

$$\underline{s}(\nu;\varepsilon) := \inf_{F \ni \nu} \liminf_{n\to\infty} \frac{1}{n} \log_2 N(F;n,\varepsilon)\ ,\ \overline{s}(\nu;\varepsilon) := \inf_{F \ni \nu} \limsup_{n\to\infty} \frac{1}{n} \log_2 N(F;n,\varepsilon)\,.$$

In these definitions we can take the infimum over any base of neighborhoods of $\nu$. For any ergodic measure† $\mu$, and general dynamical system (Corollary 3.2)

$$\lim_{\varepsilon\to 0} \underline{s}(\mu;\varepsilon) = \lim_{\varepsilon\to 0} \overline{s}(\mu;\varepsilon) = h(T,\mu)\,.$$

If the uniform separation condition is true and the ergodic measures are entropy-dense, that is, for each $\nu \in M(X,T)$, each neighborhood $F \ni \nu$, and $h^* < h(T,\nu)$, there exists an ergodic measure $\rho \in F$ such that $h^* < h(T,\rho)$, then (Proposition 3.2)

$$\lim_{\varepsilon\to 0} \underline{s}(\mu;\varepsilon) = \lim_{\varepsilon\to 0} \overline{s}(\mu;\varepsilon) = h(T,\mu) \quad \forall\ \mu \in M(X,T)\,. \tag{4}$$

Under the same two conditions, we also prove that the entropy map on $M(X,T)$, $\mu \mapsto h(T,\mu)$, is upper semi-continuous (Proposition 3.3). The g-almost product property implies entropy-density (Theorem 2.1 in [**PfS2**]).

In Section 5 we prove

† See [**K**] and [**BK**] for related results.

THEOREM 1.1. *If the* $\mathsf{g}$*-almost product property and the uniform separation property hold, then for any compact connected non-empty set* $K \subset M(X,T)$

$$\inf\{h(T,\mu) : \mu \in K\} = h_{\mathrm{top}}(T, G_K)\,.$$

The difficult part of the proof is to get a lower bound for $h_{\mathrm{top}}(T, G_K)$. The method consists in constructing a subset of $G_K$, which is simple enough, so that we can estimate its topological entropy. The construction, which generalizes the construction of normal numbers by Champernowne [**Ch**], is inspired by a similar construction in [**PfS1**]. In Section 6 we drop the condition of uniform separation and assume only the $\mathsf{g}$-almost product property. We prove (Proposition 6.1) that the $\mathsf{g}$-almost product property alone implies (4). We are able to treat the particular case (but the most important one) of saturated sets of generic points for any invariant probability measure. Adapting the method of proof of Section 5, the main result of Section 6 is

THEOREM 1.2. *If the* $\mathsf{g}$*-almost product property holds, then*

$$h_{\mathrm{top}}(T, G_\mu) = h(T,\mu) \quad \forall\, \mu \in M(X,T)\,.$$

In the last section we give an application of our results to multi-fractal analysis; we compute the entropy spectrum of ergodic averages. Let $\varphi$ be a continuous function with values in a topological vector space $Y$, such that the map

$$\Phi : M(X,T) \to Y\ ,\ \Phi(\mu) := \int_X \varphi\, d\mu$$

is continuous. Let $a \in Y$ and consider the level-set

$$K_a := \Big\{x \in X : \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \varphi(T^j x) = a\Big\}\,.$$

If the $\mathsf{g}$-almost property holds, then

$$h_{\mathrm{top}}(T, K_a) = \sup\Big\{h(T,\mu) : \mu \in M(X,T)\,,\ \int_X \varphi\, d\mu = a\Big\}\,. \tag{5}$$

Finally we compare (5) with the corresponding (weaker) result in [**TV**], and comment about their proof.

We write $\mathbf{Z}_+$ for the set of nonnegative integers $\mathbf{N} \cup \{0\}$. If $r, s \in \mathbf{Z}_+$, $r \le s$, we set $[r,s] := \{k \in \mathbf{Z}_+ : r \le k \le s\}$, and $\Lambda_n := [0, n-1]$. The cardinality of a finite set $\Lambda$ is denoted by $|\Lambda|$. We set

$$\langle\, f, \mu\,\rangle := \int_X f\, d\mu\,.$$

There exists a countable and separating set of continuous functions $\{f_1, f_2, \ldots\}$ with $0 \le f_k(x) \le 1$, and such that

$$d(\mu,\nu) \equiv \|\mu - \nu\| := \sum_{k\ge 1} 2^{-k} |\langle\, f_k, \mu - \nu\,\rangle| \tag{6}$$

defines a metric for the weak*-topology on $M(X,T)$. Notice that $\|\mu-\nu\| \le 1$. In the rest of the paper we find convenient to use an equivalent metric on $X$, still denoted by $d$,

$$d(x,y) := d(\delta_x, \delta_y)\,. \tag{7}$$

We refer to [**W**] for Ergodic Theory, and to [**Pe**] for Dimension Theory. The problem of the equality of the topological entropy of generic points of an invariant measure and its metric entropy has a long history. Let us mention the classical earlier results by Eggleston [**Eg**], Wegmann [**We**] and Colebrook [**Co**]. Two of the main impulses to this field are Cajar's monograph [**C**] and Young's paper [**Y**]. More recent references include [**O2**] and [**Te**]. Another important related work is Barreira and Schmeling's paper [**BaSc**] about Hausdorff dimension and topological entropy of sets of "non-typical" points.

## 2. $\mathsf{g}$-*almost product property*

Let $x \in X$. The dynamical ball $B_n(x,\varepsilon)$ is the set

$$B_n(x,\varepsilon) := \big\{y \in X:\ \max\{d(T^jx, T^jy) \le \varepsilon : j \in \Lambda_n\}\big\}\,.$$

A point $x \in X$ $\varepsilon$-shadows a sequence $\{x_0, x_1, \ldots, x_k\}$ if

$$d(T^jx, x_j) \le \varepsilon \quad \forall\, j = 0, \ldots, k\,.$$

We first recall the definition of specification.

Definition 2.1. *The dynamical system $(X,d,T)$ has the* specification property *if the following is true. For any $\Delta > 0$ there exists $\mathtt{k}(\Delta)$ such that for any finite collection of $p$ intervals $I_j = [a_j, b_j] \subset \mathbf{Z}^+$, $j = 0, \ldots, p-1$, such that $a_j - b_{j-1} \ge \mathtt{k}(\Delta)$, $j = 1, \ldots, p-1$, and any collection of $p$ points $\{x_0, \ldots, x_{p-1}\}$, there exists $x \in X$ such that*

$$d(T^{a_j+m}x, T^m x_j) \le \Delta \quad \forall\, m = 0, \ldots, b_j - a_j \quad \textit{and} \quad \forall\, j = 0, \ldots, p-1\,.$$

Our main definition is Definition 2.3. The terminology of Definition 2.2 is taken from [**S**].

Definition 2.2. *Let $\mathsf{g} : \mathbf{N} \to \mathbf{N}$ be a given nondecreasing unbounded map with the properties*

$$\mathsf{g}(n) < n \quad \textit{and} \quad \lim_{n\to\infty} \frac{\mathsf{g}(n)}{n} = 0\,.$$

*The function* $\mathsf{g}$ *is called* blowup function. *Let $x \in X$ and $\varepsilon > 0$. The* $\mathsf{g}$-blowup *of $B_n(x,\varepsilon)$ is the closed set*

$$B_n(\mathsf{g}; x, \varepsilon) := \big\{y \in X : \exists \Lambda \subset \Lambda_n\,,\ |\Lambda_n \backslash \Lambda| \le \mathsf{g}(n) \ \textit{and}\ \max\{d(T^jx, T^jy) : j \in \Lambda\} \le \varepsilon\big\}\,.$$

Definition 2.3. *The dynamical system $(X,d,T)$ has the* $\mathsf{g}$-almost product property *with blowup function* $\mathsf{g}$, *if there exists a nonincreasing function $\mathtt{m} : \mathbf{R}^+ \to \mathbf{N}$, such* *that for any* $k \in \mathbf{N}$*, any* $x_1 \in X, \ldots, x_k \in X$*, any positive* $\varepsilon_1, \ldots, \varepsilon_k$*, and any integers* $n_1 \geq \mathtt{m}(\varepsilon_1), \ldots, n_k \geq \mathtt{m}(\varepsilon_k)$*,*

$$\bigcap_{j=1}^{k} T^{-M_{j-1}} B_{n_j}(\mathtt{g}; x_j, \varepsilon_j) \neq \emptyset\,,$$

*where* $M_0 := 0$*,* $M_i := n_1 + \cdots + n_i$*,* $i = 1, \ldots, k-1$*.*

The main idea expressed in this definition is that one requires only partial shadowing of the specified orbit segments, contrary to Definition 2.1. If $(Y, d', S)$ is a factor of $(X, d, T)$, and $(X, d, T)$ has the $g$-almost product property, then $(Y, d', S)$ has the $g$-almost product property. Indeed, let $\psi : X \to Y$ be a continuous surjection such that $\psi \circ T = S \circ \psi$. Given $\varepsilon' > 0$, there exists $\varepsilon > 0$ such that if $d(x, y) \leq \varepsilon$, then $d'(\psi(x), \psi(y)) \leq \varepsilon'$. Then define $\mathtt{m}'(\varepsilon') := \mathtt{m}(\varepsilon)$. In particular, the $g$-almost product property does not depend on the choice of the metric, as long as the topology on $X$ remains the same.

**Example.** All $\beta$-shifts have the $\mathtt{g}$-almost product property. We use the notations of Section 5 in [**PfS2**]. Let $\beta > 1$ be a real number, and $\mathtt{A} := \{0, 1, \ldots, \lfloor \beta \rfloor\}$. The $\beta$-shift is a subshift $X^\beta \subset \mathtt{A}^{\mathbf{N}}$. The shift map is denoted by $T$, and the elements of $X^\beta$ are denoted by $\omega = \{\omega_i\}$. The language $\mathcal{L}$ of the $\beta$-shift is the set of all words $\omega_1^k \equiv (\omega_1, \ldots, \omega_k)$ with $\omega \in X^\beta$ and $k \in N$. The set of all words of $\mathcal{L}$ of length $k$ is denoted by $\mathcal{L}_k$. For the separating continuous functions $\{f_1, f_2, \ldots\}$ we choose the functions

$$f_{\mathrm{w}}(\omega) := \begin{cases} 1 & \text{if w is a prefix of } \omega \\ 0 & \text{otherwise} \end{cases} \quad \text{with } \mathrm{w} \in \mathcal{L}\,.$$

We enumerate these functions by enumerating first the functions with $\mathrm{w} \in \mathcal{L}_1$, then those with $\mathrm{w} \in \mathcal{L}_2$ and so on. If $\omega$ and $\omega'$ are two points of $X^\beta$ having a common largest prefix of length $k$, then by our definition of the distance (see (6))

$$d(\omega, \omega') \leq 2^{-\ell_k} \quad \text{with } \ell_k := \sum_{j=1}^{k} |\mathcal{L}_j|\,.$$

The main observation is that, given any $\mathrm{w} \in \mathcal{L}$ we can find a word $\widehat{\mathrm{w}} \in \mathcal{L}$, which differs from w by at most one character, and such that for any $\mathrm{w}' \in \mathcal{L}$, the concatenated word $\widehat{\mathrm{w}}\mathrm{w}' \in \mathcal{L}$ (see Proposition 5.1, in particular formula (5.7) in [**PfS2**]). Let $\omega \in X^\beta$ with $\mathrm{w} = \omega_1^n$, and $\widehat{\omega} \in X^\beta$ with the prefix $\widehat{\mathrm{w}} = \widehat{\omega}_1^n$. Suppose that w and $\widehat{\mathrm{w}}$ differ only at the character $j$, $1 \leq j \leq n$. Then

$$d(T^m\omega, T^m\widehat{\omega}) \leq \begin{cases} 2^{-\ell_{(j-1)-m}} & \text{if } m = 0, \ldots, j-1 \\ 2^{-\ell_{n-m}} & \text{if } m = j, \ldots, n-1. \end{cases} \tag{8}$$

From (8) it follows that we can take any blowup function $\mathtt{g}$ and determine the corresponding function $\mathtt{m}$.

Proposition 2.1. *Let* $\mathtt{g}$ *be any blowup function and* $(X, d, T)$ *have the specification property. Then it has the* $\mathtt{g}$*-almost product property.*

**Proof**: We may assume that the function $\mathtt{k}$ in Definition 3.1 is nonincreasing. Let $\{x_1, \ldots, x_p\}$ and $\{\varepsilon_1, \ldots, \varepsilon_p\}$ be given. Let

$$\Delta'_j := 2^{-j} \quad j \in \mathbf{N}\,.$$

Notice that $\Delta'_j = \sum_{k>j} \Delta'_k$. Define

$$\mathtt{m}(\varepsilon) := \begin{cases} \min\{m : \mathtt{g}(m) \geq 2\mathtt{k}(\Delta'_j)\} & \text{if } \varepsilon = 2\Delta'_j \\ \mathtt{m}(2\Delta'_j) & \text{if } j \text{ is the maximum of } i \text{ with } 2\Delta'_i \leq \varepsilon. \end{cases}$$

It is sufficient to prove the statement for $\varepsilon_i$ of the form $2\Delta'_{j_i}$, $i = 1, \ldots, p$, where, as above $j_i$ is the maximum of $k$ such that $2\Delta'_k \leq \varepsilon_i$. Precisely, if $\varepsilon_i$ is not of that form, we change it into $2\Delta'_{j_i}$. From now on we assume that, for all $i$, $\varepsilon_i$ is of the form $2\Delta'_{j_i}$ and $n_i > \mathtt{m}(\varepsilon_i)$.

We prove the proposition by an iterative construction. Let $\Delta(x_i) := \varepsilon_i/2 = \Delta'_{j_i}$ and $n(x_i) := n_i$. The sequence $\{x_1, \ldots, x_p\}$ is considered as an ordered sequence; its elements are called *original points*. The possible values of $\Delta(x_i)$ are written $\Delta_1 > \Delta_2 > \ldots > \Delta_q$. A *level-k point* is by definition an original point $x_j$ such that $\Delta(x_j) = \Delta_k$. We set

$$k_i := \mathtt{k}(\Delta_i) \quad i = 1, \ldots q\,.$$

There are $q$ steps; at each step we replace some subsets of consecutive points by single points, called *concatenated points*, and obtain in this way a new ordered sequence, consisting of original and concatenated points. After the $q$ steps have been completed, the sequence is reduced to a single element $y$ such that

$$y \in \bigcap_{j=1}^{p} T^{-M_{j-1}} B_{n_j}(\mathtt{g}; x_j, \varepsilon_j)\,.$$

At step 1 we consider the level-1 points labeled by

$$S_1 := \{k \in [1, p] : \Delta(x_k) = \Delta_1\}\,.$$

If $S_1 = [1, p]$, then by the specification property there exists $y$ such that

$$\begin{aligned} d\big(T^m x_1, T^m y\big) &\leq \Delta_1 \quad \forall\, m = k_1, \ldots, n(x_1) - k_1 - 1 \\ d\big(T^m x_2, T^{n_2+m} y\big) &\leq \Delta_1 \quad \forall\, m = k_1, \ldots, n(x_2) - k_1 - 1 \\ \cdots &\leq \cdots \\ d\big(T^m x_p, T^{n_1+\cdots+n_{p-1}+m} y\big) &\leq \Delta_1 \quad \forall\, m = k_1, \ldots, n(x_p) - k_1 - 1\,, \end{aligned}$$

which proves this case. If $S_1 \neq [1, p]$, then we decompose it into maximal subsets of consecutive points, called *components*. (The components are defined with respect to the whole sequence.) Let $J$ be a component, say $[r, s]$ with $r < s$. By the specification property there exists $y$ such that

$$\begin{aligned} d\big(T^m x_r, T^m y\big) &\leq \Delta_1 \quad \forall\, m = k_1, \ldots, n(x_r) - k_1 - 1 \\ \cdots &\leq \cdots \\ d\big(T^m x_s, T^{n_r+\cdots+n_{s-1}+m} y\big) &\leq \Delta_1 \quad \forall\, m = k_1, \ldots, n(x_s) - k_1 - 1\,. \end{aligned}$$

Here $n_r+\cdots+n_{s-1}$ stands for $n(x_r)+\cdots+n(x_{s-1})$. This implies that

$$y\in B_{n_r}(\mathsf{g};x_r,\varepsilon_r)\cap\cdots\cap T^{-(n_r+\cdots+n_{s-1})}B_{n_s}(\mathsf{g};x_s,\varepsilon_s)\,.$$

We replace the sequence $\{x_1,\ldots,x_p\}$ by the (ordered) sequence

$$\{x_1,\ldots,x_{r-1},y,x_{s+1}\ldots,x_p\}\,,$$

and set, for the concatenated point $y$,

$$\Delta(y):=\Delta_1\quad n(y):=n(x_r)+\cdots+n(x_s)\,.$$

We do this operation for all components which are not singletons. After these operations we have a new (ordered) sequence $\{z_1,\ldots,z_{p_1}\}$, $p_1\leq p$, where the point $z_i$ is either a point of the original sequence, or a concatenated point. This ends the construction at step 1. Let

$$S_2:=\{k\in[1,p_1]:\Delta(z_k)\geq\Delta_2\}\,.$$

We decompose this set into components. Let $[r,s]$ be a component which is not a singleton ($r<s$). We replace that component by a single concatenated point $y$ such that

$$d\big(T^m z_r,T^m y\big)\leq\begin{cases}\Delta_2 & \forall\, m=k_i,\ldots,n(z_r)-k_i-1 \quad \text{if } z_r \text{ level-}i\text{ point}\\ \Delta_2 & \forall\, m=0,\ldots,n(z_r)-1 \quad \text{if } z_r \text{ concatenated point}\end{cases}$$

$$\cdots\leq\cdots$$

$$d\big(T^m z_s,T^{n_r+\cdots+n_{s-1}+m}y\big)\leq\begin{cases}\Delta_2 & \forall\, m=k_i,\ldots,n(z_s)-k_i-1 \quad \text{if } z_r \text{ level-}i\text{ point}\\ \Delta_2 & \forall\, m=0,\ldots,n(z_s)-1 \quad \text{if } z_r \text{ concatenated point.}\end{cases}$$

In that formula $n_r+\cdots+n_{s-1}$ stands for $n(z_r)+\cdots n(z_{s-1})$. Existence of such a $y$ is a consequence of the specification property. We set

$$\Delta(y):=\Delta_2\quad n(y):=n(z_r)+\cdots+n(z_s)\,.$$

The construction of $y$ involves consecutive points of the original sequence (via the concatenated points), say points $x_j$, $j\in[u,t]$. Since $\Delta'_j=\sum_{k>j}\Delta'_k$,

$$y\in B_{n_u}(\mathsf{g};x_u,\varepsilon_u)\cap\cdots\cap T^{-(n_u+\cdots+n_{t-1})}B_{n_t}(\mathsf{g};x_t,\varepsilon_t)\,.$$

We do these operations for all components of $S_2$, which are not singletons. We get a new ordered sequence, still denoted by $\{z_1,\ldots,z_{p_2}\}$. This ends the construction at level 2. The construction at level 3 is similar to the construction at level 2, using

$$S_3:=\{k\in[1,p_2]:\Delta(z_k)\geq\Delta_3\}\,.$$

Once step $q$ is completed, since $\Delta'_j=\sum_{k>j}\Delta'_k$, we have a single concatenated point $y$ such that for all $i=1,\ldots,p$

$$d\big(T^m x_i,T^{n_1+\cdots+n_{i-1}+m}y\big)\leq\varepsilon_i\quad\forall\, m=k_i,\ldots,n_i-k_i-1\,.$$

□

COROLLARY 2.1. *Assume that $(X, d, T)$ has the specification property. There exists a nonincreasing function $\mathtt{k} : \mathbf{R}^+ \to \mathbf{N}$ with the following property. For $y_1 \in X, \ldots, y_p \in X$, $\varepsilon_1 > 0, \ldots, \varepsilon_p > 0$, and $n_1' \in \mathbf{N}, \ldots, n_p' \in \mathbf{N}$, there exists $z \in X$ such that*

$$d(T^{M_k+m}z, T^m y_k) \leq \varepsilon_k \quad \forall\, m = 0, \ldots n_k' - 1 \text{ and } k = 1, \ldots, p\,,$$

*with*

$$M_1 := \mathtt{k}(\varepsilon_1) \quad \text{and} \quad M_k := \sum_{i=1}^{k-1} 2\mathtt{k}(\varepsilon_i) + n_i' + \mathtt{k}(\varepsilon_k) \quad k = 2, \ldots, p\,.$$

**Proof**: Specification implies surjectivity, so we may choose $x_i$ so that $T^{\mathtt{k}(\varepsilon_i)}x_i = y_i$. With $n_i := n_i' + 2\mathtt{k}(\varepsilon_i)$ we then proceed as in the proof of Proposition 2.1. □

We denote a ball in $M(X)$ by

$$\mathcal{B}(\nu, \zeta) := \{\mu \in M(X) : d(\nu, \mu) \leq \zeta\}\,.$$

LEMMA 2.1. *Assume that $(X, d, T)$ has the $\mathtt{g}$-almost product property. Let $x_1 \in X, \ldots, x_k \in X$, $\varepsilon_1 > 0, \ldots, \varepsilon_k > 0$, and $n_1 \geq \mathtt{m}(\varepsilon_1), \ldots, n_k \geq \mathtt{m}(\varepsilon_k)$ be given. Assume that*

$$\mathcal{E}_{n_j}(x_j) \in \mathcal{B}(\nu_j, \zeta_j) \quad j = 1, \ldots, k\,.$$

*Then for any $y \in \bigcap_{j=1}^k T^{-M_{j-1}} B_{n_j}(\mathtt{g}; x_j, \varepsilon_j)$ and any probability measure $\alpha$*

$$d\big(\mathcal{E}_{M_k}(y), \alpha\big) \leq \sum_{j=1}^k \frac{n_j}{M_k}\big(\zeta_j' + d(\nu_j, \alpha)\big)\,,$$

*where $M_i = n_1 + \cdots + n_i$ and*

$$\zeta_i' = \zeta_i + \varepsilon_i + \frac{\mathtt{g}(n_i)}{n_i} \quad i = 1, \ldots, k\,.$$

**Proof**: We have

$$\mathcal{E}_{M_k}(y) = \sum_{j=1}^k \frac{n_j}{M_k}\mathcal{E}_{n_j}(T^{M_{j-1}}y)\,,$$

and, because of our choice (6) of the distance on $X$,

$$d\big(\mathcal{E}_{n_j}(x_j), \mathcal{E}_{n_j}(T^{M_{j-1}}y)\big) \leq \frac{1}{n_j}\sum_{m=0}^{n_j-1} d(T^m x_j, T^{M_{j-1}+m}y) \leq \frac{\mathtt{g}(n_j)}{n_j} + \varepsilon_j \frac{n_j - \mathtt{g}(n_j)}{n_j}\,.$$

The result follows from the triangle inequality and the definition of the distance (6):

$$d\big(\mathcal{E}_{M_k}(y), \alpha\big) \leq \sum_{j=1}^k \frac{n_j}{M_k} d\big(\mathcal{E}_{n_j}(T^{M_{j-1}}y), \alpha\big)\,.$$

□

3. *Uniform separation*

Let $F \subset M(X)$ be a neighborhood of $\nu \in M(X,T)$. Recall that

$$X_{n,F} := \{x \in X : \mathcal{E}_n(x) \in F\}\,,$$

$$N(F;n,\varepsilon) := \text{maximal cardinality of a } (n,\varepsilon)\text{-separated subset of } X_{n,F}\,, \tag{9}$$

and define

$$N(F;\delta,n,\varepsilon) := \text{maximal cardinality of a } (\delta,n,\varepsilon)\text{-separated subset of } X_{n,F}\,.$$

Let $\xi = \{V_i\colon\, i = 1,\ldots,k\}$, be a finite partition of measurable sets of $X$. The entropy of $\nu \in M(X)$ with respect to $\xi$ is

$$H(\nu,\xi) := -\sum_{V_i \in \xi} \nu(V_i)\log_2 \nu(V_i).$$

We write $T^{\vee n}\xi := \bigvee_{k\in\Lambda_n} T^{-k}\xi$. The entropy of $\nu \in M(X,T)$ with respect to $\xi$ is

$$h(T,\nu,\xi) := \lim_n \frac{1}{n} H(\nu, T^{\vee n}\xi)\,,$$

and the metric entropy of $\nu$ is $h(T,\nu) := \sup_\xi h(T,\nu,\xi)$.

3.1. *Uniform separation property*

Definition 3.1. *The dynamical system $(X,d,T)$ has* uniform separation property *if the following holds. For any $\eta$, there exist $\delta^* > 0$ and $\varepsilon^* > 0$ so that for $\mu$ ergodic and any neighborhood $F \subset M(X)$ of $\mu$, there exists $n^*_{F,\mu,\eta}$, such that for $n \geq n^*_{F,\mu,\eta}$,*

$$N(F;\delta^*,n,\varepsilon^*) \geq 2^{n(h(T,\mu)-\eta)}\,.$$

**Remark.** Notice that uniform separation implies that $h_{\rm top}(T,X)$ is finite. Indeed,

$$2^{n(h(T,\mu)-\eta)} \leq N(F;\delta^*,n,\varepsilon^*) \leq N(X;n,\varepsilon^*)\,,$$

so that, for all $\mu$ ergodic

$$h(T,\mu) \leq \limsup_{n\to\infty} \frac{1}{n}\log_2 N(X;n,\varepsilon^*) + \eta < \infty\,.$$

The result follows from Corollary 8.6.1 in [**W**].

Theorem 3.1. *Suppose that $T$ is expansive, or that $T$ is asymptotically $h$-expansive. Then $(X,d,T)$ has the uniform separation property.*

**Proof**: If $T$ is expansive, and $\varepsilon > 0$ is an expansive constant for $T$, then for a partition $\alpha$ with diameter smaller than $\varepsilon$ (see [**W**])

$$h(T,\rho) = h(T,\rho,\alpha) \quad \text{if } \rho \in M(X,T)\,.$$

If $T$ is asymptotically $h$-expansive, then there exists $h(\varepsilon)$ such that

$$\lim_{\varepsilon\to 0} h(\varepsilon) = 0\,,$$

and if $\alpha$ is a finite partition with diameter smaller than $\varepsilon$,

$$h(T,\rho) \leq h(T,\rho,\alpha) + h(\varepsilon) \quad \text{if } \rho \in M(X,T)\,. \tag{10}$$

(See [**N**] Theorem 1 and [**Mi1**].) Let $\eta > \eta' > 0$. The value of $\eta'$ is fixed in (12). We choose $\varepsilon$ so that either $\varepsilon$ is an expansive constant, or

$$h(\varepsilon) \leq \frac{\eta'}{4}\,. \tag{11}$$

Let $\nu \in M(X,T)$. We first construct a neighborhood of $\nu$, $W_\nu \subset M(X)$. It is sufficient to prove the result for an ergodic $\mu \in W_\nu \cap M(X,T)$, since $M(X,T)$ can be covered by a finite number of neighborhoods $W_{\nu_i} \subset M(X)$, $i = 1,\ldots,n$.

Let $\xi = \{A_1,\ldots,A_k\}$ be a fixed finite partition of $X$ with $2\,\mathrm{diam}\xi \leq \varepsilon$. We choose $\eta'$ and a constant $\delta^* > 0$ so that $2\eta' + \delta^* < \frac{1}{2}$ and

$$\phi(\delta^* + 2\eta') + (\delta^* + 2\eta')\log_2(2k-1) < \eta - \eta'\,, \tag{12}$$

where

$$\phi(\delta) := -\delta\log_2\delta - (1-\delta)\log_2(1-\delta)\,. \tag{13}$$

Since $\nu$ is regular, let $V_j \subset A_j$, $j = 1,\ldots,k$, be compact subsets, with

$$\nu(A_j \backslash V_j) < \frac{\eta'}{4k\log_2(2k)}\,.$$

Define $\varepsilon^* > 0$ so that $d(x,y) > 2\varepsilon^*$ whenever $x \in V_i$, $y \in V_j$, $i \neq j$. There exists an integer $n^*$, such that for $n \geq n^*$,

$$\frac{\eta'}{4\log_2 k} \geq \frac{\log_2 2}{n\log_2(2k)}\,.$$

Let $U_i \supset V_i$, $i = 1,\ldots,k$, be $k$ open neighborhoods with $\mathrm{diam}U_i < \varepsilon$, and such that $d(x,y) > \varepsilon^*$ whenever $x \in U_i$, $y \in U_j$, $i \neq j$. Let $\xi'$ be a partition of $X$ which contains $U_j$, $j = 1,\ldots,k$, and the sets $A_i \setminus \cup_j U_j$ so that $\mathrm{diam}\xi' < \varepsilon$. Let $K := X \setminus \cup_j U_j$; since $K$ is closed, $I_K$, the indicator function of $K$, is upper semi-continuous. We define the neighborhood $W_\nu$ by

$$W_\nu := \Big\{\rho \in M(X) \colon \langle\, I_K, \rho\,\rangle \leq \langle\, I_K, \nu\,\rangle + \frac{\eta'}{4\log_2(2k)}\Big\}\,.$$

It is convenient to label the atoms of $T^{\vee n}\xi'$ by words w of length $n$ over an alphabet $\mathtt{A}$ of at most $2k$ letters. The letters $1,\ldots,k$ label the atoms $U_1,\ldots,U_k$ of $\xi'$, and the other letters the non-empty atoms among $A_1 \setminus \cup_j U_j, \ldots, A_k \setminus \cup_j U_j$. We define an application $\mathrm{w} : X \to \mathtt{A}^{\Lambda_n}$,

$$\mathrm{w}(x)_j := i \quad \text{if } T^j x \text{ is in the atom of } \xi' \text{ labeled by } i\,.$$

We prove the result for an ergodic $\mu \in W_\nu \cap M(X,T)$. Because of our choice of $\varepsilon$ (see (10) and (11)), there exists $n^*$ so that

$$H(\mu, T^{\vee n}\xi') > n\Big(h(T,\mu) - \frac{\eta'}{2}\Big) \quad \text{if } n \geq n^*\,.$$

By definition of $W_\nu$

$$\mu(X\setminus\cup_j U_j) \leq \nu(X\setminus\cup_j U_j)+\frac{\eta'}{4\log_2(2k)} \leq \sum_{j=1}^k \nu(A_j\setminus V_j)+\frac{\eta'}{4\log_2(2k)} < \frac{\eta'}{2\log_2(2k)}\,.$$

Let

$$Y_n := \big\{x\colon |\{j\in\Lambda_n : \mathrm{w}(x)_j > k\}| \leq n\eta'\big\}\,,$$

so that $\lim_n \mu(X\setminus Y_n)=0$. Therefore, by the Ergodic Theorem and by taking $n^*$ large enough, if $F$ is a neighborhood of $\mu$, then we may suppose that

$$\mu(X\setminus(X_{n,F}\cap Y_n)) < \frac{\eta'}{2\log_2(2k)} - \frac{\log_2 2}{n\log_2(2k)} \quad \text{if } n\geq n^*\,.$$

Set

$$A_0^n := X\setminus(X_{n,F}\cap Y_n)\,.$$

By conditioning with respect to $\{A_0^n, X\backslash A_0^n\}$ we get

$$\begin{aligned} H(\mu, T^{\vee n}\xi') \leq &\log_2 2 + \mu(A_0^n)H(\mu(\,\cdot\,|A_0^n), T^{\vee n}\xi') + \\ &(1-\mu(A_0^n))H(\mu(\,\cdot\,|X\setminus A_0^n), T^{\vee n}\xi')\,. \end{aligned}$$

Since the number of atoms of $\xi'$ is at most $2k$, $H(\mu(\,\cdot\,|A_0^n), T^{\vee n}\xi') \leq n\log_2(2k)$. Thus, for $n\geq n^*$,

$$H(\mu(\,\cdot\,|X\setminus A_0^n), T^{\vee n}\xi') > n(h(T,\mu)-\eta')\,. \tag{14}$$

Let $\Xi_n$ denote the image of $X_{n,F}\cap Y_n$ by the map w. Inequality (14) implies that

$$|\Xi_n| \geq 2^{n(h(T,\mu)-\eta')}\,.$$

The Hamming distance $d_n^H(\mathrm{w},\mathrm{w}')$ of two different elements w and w′ of $\mathtt{A}^n$ is the number of different letters of w and w′. Let $\Xi_n'\subset\Xi_n$ of maximum cardinality such that $d_n^H(\mathrm{w},\mathrm{w}') > n(2\eta'+\delta^*)$ for any pair $(\mathrm{w},\mathrm{w}')$ of different elements of $\Xi_n'$. Let $\Gamma_n$ be defined by selecting exactly one point of $X_{n,F}\cap Y_n$ from each atom of $T^{\vee n}\xi_n'$, which is labeled by a word of $\Xi_n'$. By construction $\Gamma_n\subset X_{n,F}$ is $(\delta^*,n,\varepsilon^*)$-separated. The maximum cardinality of $\Xi_n'$ implies that for each $\mathrm{w}\in\Xi_n$ there exists $\mathrm{w}'\in\Xi_n'$ so that $d_n^H(\mathrm{w}',\mathrm{w})\leq n(2\eta'+\delta^*)$. For a given $\mathrm{w}\in\mathtt{A}^n$ (see e.g. Lemma 2.2 in [**PfS2**])

$$|\{\mathrm{w}'\in\mathtt{A}^n : d_n^H(\mathrm{w}',\mathrm{w})\leq n(2\eta'+\delta^*)\}| \leq 2^{n\phi(2\eta'+\delta^*)}(|\mathtt{A}|-1)^{n(2\eta'+\delta^*)}\,.$$

Therefore, by our choice (12) of $\eta'$ and $\delta^*$

$$|\Gamma_n| = |\Xi_n'| \geq \frac{2^{n(h(T,\mu)-\eta')}}{2^{n\phi(2\eta'+\delta^*)}(2k-1)^{n(2\eta'+\delta^*)}} \geq 2^{n(h(T,\mu)-\eta)}\,.$$

□

Corollary 3.1. *Assume that $(X,d,T)$ has the uniform separation property, and that the ergodic measures are entropy-dense. For any $\eta$, there exist $\delta^*>0$ and*

$\varepsilon^* > 0$ *so that for* $\mu \in M(X,T)$ *and any neighborhood* $F \subset M(X)$ *of* $\mu$*, there exists* $n^*_{F,\mu,\eta}$*, such that*

$$N(F;\delta^*,n,\varepsilon^*) \geq 2^{n(h(T,\mu)-\eta)} \quad \textit{if } n \geq n^*_{F,\mu,\eta}\,.$$

*For any* $\mu \in M(X,T)$*,*

$$h(T,\mu) \leq \lim_{\varepsilon\to 0}\lim_{\delta\to 0}\inf_{F\ni\mu}\liminf_{n\to\infty}\frac{1}{n}\log_2 N(F;\delta,n,\varepsilon)\,.$$

**Proof**: Let $\eta > 0$ and $\mu \in F$. For $\eta/2$ there exist $\delta^* > 0$ and $\varepsilon^* > 0$ such that the statement is true for the ergodic measures with $\eta/2$ instead of $\eta$. If $\mu$ is not ergodic, and $F \ni \mu$, then we choose $\nu \in F$, ergodic, so that $h(T,\mu) - h(T,\nu) \leq \eta/2$. The statement is then true with $n^*_{F,\mu,\eta} = n^*_{F,\nu,\eta/2}$. The second statement is an easy consequence of the first one. □

3.2. *The entropy map* We study the entropy map when the uniform separation property holds. Proposition 3.2 gives another expression for the entropy of $\nu \in M(X,T)$, and Proposition 12 gives a sufficient condition implying that the entropy map is upper semi-continuous. Recall that for $\varepsilon > 0$ and $\nu \in M(X,T)$ (see (9))

$$\underline{s}(\nu;\varepsilon) := \inf_{F\ni\nu}\liminf_{n\to\infty}\frac{1}{n}\log_2 N(F;n,\varepsilon) \quad\text{and}\quad \underline{s}(\nu) := \lim_{\varepsilon\to 0}\underline{s}(\nu;\varepsilon)\,,$$

with similar definitions for $\overline{s}(\nu;\varepsilon)$ and $\overline{s}(\nu)$. If $\underline{s}(\nu) = \overline{s}(\nu)$, then the common value is denoted by $s(\nu)$.

The next three results are valid for any dynamical systems. Lemma 3.1 is essentially one-half of the variational principle for the topological entropy, [**G**], [**Mi2**] and [**W**] section 8.2.

LEMMA 3.1. *Let* $(X,d,T)$ *be a dynamical system. Let* $\{E_n\}$ *be a sequence of* $(n,\varepsilon)$*-separated subsets and define*

$$\nu_n := \frac{1}{n|E_n|}\sum_{x\in E_n}\sum_{k=0}^{n-1}\delta_{T^k x}\,.$$

*Assume that* $\lim_n \nu_n = \mu$*. Then*

$$\limsup_{n\to\infty}\frac{1}{n}\log_2|E_n| \leq h(T,\mu)\,.$$

**Proof**: Up to minor modifications, the proof is identical with the second part of the proof of Theorem 8.6 in [**W**]. □

PROPOSITION 3.1. *Let* $(X,d,T)$ *be a dynamical system and* $\mu \in M(X,T)$*. Then*

$$\overline{s}(\mu) \leq h(T,\mu)\,.$$

**Proof**: If $h(T,\mu) = \infty$, then there is nothing to prove. Let $h(T,\mu) < \infty$. Suppose that

$$\lim_{\varepsilon\to 0}\inf_{F\ni\mu}\limsup_{n\to\infty}\frac{1}{n}\log_2 N(F;n,\varepsilon) > h(T,\mu)\,.$$

There exist $\varepsilon^* > 0$ and $\delta > 0$ so that for $\varepsilon \leq \varepsilon^*$,

$$\inf_{F \ni \mu} \limsup_{n\to\infty} \frac{1}{n} \log_2 N(F; n, \varepsilon) \geq h(T, \mu) + 2\delta\,.$$

Let $0 < \varepsilon \leq \varepsilon^*$. There exists a decreasing sequence of convex closed neighborhoods $\{C_n\}$ so that

$$\bigcap_n C_n = \{\mu\} \quad \text{and} \quad \limsup_{n\to\infty} \frac{1}{n} \log_2 N(C_n; n, \varepsilon) \geq h(T, \mu) + 2\delta\,. \tag{15}$$

Let $E_n \subset X_{n,C_n}$ be $(n, \varepsilon)$-separated with maximal cardinality, and define

$$\nu_n := \frac{1}{n|E_n|} \sum_{x\in E_n} \sum_{k=0}^{n-1} \delta_{T^k x} \in C_n\,.$$

By definition $\lim_n \nu_n = \mu$. By Lemma 3.1

$$\limsup_{n\to\infty} \frac{1}{n} \log_2 |E_n| = \limsup_{n\to\infty} \frac{1}{n} \log_2 N(C_n; n, \varepsilon) \leq h(T, \mu)\,,$$

which contradicts (15). □

Corollary 3.2. *Let $(X, d, T)$ be a dynamical system. Then $s(\mu)$ is well-defined, and $s(\mu) = h(T, \mu)$, for all ergodic measures. Moreover,*

$$s(\mu) = \lim_{\varepsilon\to 0} \lim_{\delta\to 0} \inf_{F\ni\mu} \liminf_{n\to\infty} \frac{1}{n} \log_2 N(F; \delta, n, \varepsilon) = \lim_{\varepsilon\to 0} \lim_{\delta\to 0} \inf_{F\ni\mu} \limsup_{n\to\infty} \frac{1}{n} \log_2 N(F; \delta, n, \varepsilon)\,.$$

**Proof**: From Proposition 3.1 and $N(F; \delta, n, \varepsilon) \leq N(F; n, \varepsilon)$,

$$\lim_{\varepsilon\to 0} \lim_{\delta\to 0} \inf_{F\ni\mu} \limsup_{n\to\infty} \frac{1}{n} \log_2 N(F; \delta, n, \varepsilon) \leq \overline{s}(\mu) \leq h(T, \mu)\,.$$

From Proposition 1.1

$$h(T, \mu) \leq \lim_{\varepsilon\to 0} \lim_{\delta\to 0} \inf_{F\ni\mu} \liminf_{n\to\infty} \frac{1}{n} \log_2 N(F; \delta, n, \varepsilon) \leq \underline{s}(\mu)\,.$$

□

Proposition 3.2. *Let $(X, d, T)$ be a dynamical system. If the uniform separation condition is true and the ergodic measures are entropy-dense, then $s(\mu)$ is well-defined, and $s(\mu) = h(T, \mu)$, for all $\mu \in M(X, T)$.*

**Proof**: Same proof as the proof of Corollary 3.2, using Corollary 3.1 instead of Proposition 1.1. □

Proposition 3.3. *Let $(X, d, T)$ be a dynamical system. If the uniform separation condition is true and the ergodic measures are entropy-dense, then the entropy map is upper semi-continuous.*

**Proof**: Let $F$ be a neighborhood of $\nu$. Given $\eta > 0$, by Corollary 3.1 there exists $\varepsilon^*$ so that

$$\sup_{\mu\in F} h(T, \mu) \leq \limsup_{n\to\infty} \frac{1}{n} \log_2 N(F; n, \varepsilon^*) + \eta\,.$$

Hence

$$\inf_{F\ni\nu}\sup_{\mu\in F} h(T,\mu) \leq \inf_{F\ni\nu}\limsup_{n\to\infty}\frac{1}{n}\log_2 N(F;n,\varepsilon^*)+\eta\,.$$

Since $\eta$ is arbitrary, we get from Proposition 3.1

$$\inf_{F\ni\nu}\sup_{\mu\in F} h(T,\mu) \leq h(T,\nu)\,.$$

□

4. *Upper bound for* $h_{\rm top}(T,G_K)$

Let $E\subset X$, and $\mathcal{G}_n(E,\varepsilon)$ be the collection of all finite or countable covers of $E$ by sets of the form $B_m(x,\varepsilon)$ with $m\geq n$. We set

$$C(E;t,n,\varepsilon,T):=\inf\Big\{\sum_{B_m(x,\varepsilon)\in\mathcal{C}} 2^{-tm} : \mathcal{C}\in\mathcal{G}_n(E,\varepsilon)\Big\}\,,$$

and

$$C(E;t,\varepsilon,T):=\lim_{n\to\infty} C(E;t,n,\varepsilon,T)\,.$$

Then

$$h_{\rm top}(E,\varepsilon,T):=\inf\{t\colon C(E;t,\varepsilon,T)=0\}=\sup\{t\colon C(E;t,\varepsilon,T)=\infty\}\,,$$

and the topological entropy of $E$ is

$$h_{\rm top}(E,T):=\lim_{\varepsilon\to 0} h_{\rm top}(E,\varepsilon,T)\,.$$

Theorem 4.1. *Let $(X,d,T)$ be a dynamical system and $\mu\in M(X,T)$.*
*(1) Let $K\subset M(X,T)$ be a closed subset, and let*

$${}^K G:=\Big\{x\in X: \{\mathcal{E}_n(x)\}_n \text{ has a limit-point in } K\Big\}\,.$$

*Then*

$$h_{\rm top}(T,{}^K G)\leq \sup\{h(T,\rho)\colon \rho\in K\}\,.$$

*(2) If $\mu\in M(X,T)$, then*

$$h_{\rm top}(T,G_\mu)\leq h(T,\mu)\,.$$

*(3) Let $K\subset M(X,T)$ non-empty, connected and compact. Then*

$$h_{\rm top}(T,G_K)\leq \inf\{h(T,\mu)\colon \mu\in K\}\,.$$

**Proof**: Let $\mu\in M(X,T)$ and $s:=\sup_{\mu\in K} h(T,\mu)$. If $s=\infty$, there is nothing to prove. Assume that $s<\infty$; let $s'-s=2\delta>0$. Since $N(F;n,\varepsilon)$ is a nonincreasing function of $\varepsilon$, by Proposition 3.1

$$\inf_{F\ni\mu}\limsup_{n\to\infty}\frac{1}{n}\log_2 N(F;n,\varepsilon)\leq h(T,\mu)\quad \forall\,\varepsilon>0\,.$$

For any $\varepsilon>0$ there exist a neighborhood of $\mu$, $F(\mu,\varepsilon)$, and $M(F(\mu,\varepsilon))\in\mathbf{N}$, so that

$$\frac{1}{n}\log_2 N(F(\mu,\varepsilon);n,\varepsilon)\leq h(T,\mu)+\delta\quad \forall\, n\geq M(F(\mu,\varepsilon))\,.$$

Since maximal $(n,\varepsilon)$-separated subsets of a set $A$ are also $(n,\varepsilon)$-spanning subsets of $A$, for any $n \geq M(F(\mu,\varepsilon))$,

$$C(X_{n,F(\mu,\varepsilon)}; s', n, \varepsilon, T) \leq N(F(\mu,\varepsilon); n, \varepsilon) 2^{-s'n} \leq 2^{-\delta n}\,.$$

Since $K$ is compact, given a fixed $\varepsilon > 0$, we can find a finite open covering of $K$ by sets of the form $F(\mu,\varepsilon)$, say $F(\mu_j,\varepsilon)$, $j = 1,\ldots m_\varepsilon$, with $\mu_j \in K$. If $\{\mathcal{E}_n(x)\}$ has a limit-point in $K$, then $x$ is an element of

$$A_M := \bigcup_{n\geq M} \bigcup_{j=1}^{m_\varepsilon} X_{n,F(\mu_j,\varepsilon)}\,,$$

for arbitrarily large $M$. Thus, for $M \geq \max_j M(F(\mu_j,\varepsilon))$,

$$C({}^K G; s', M, \varepsilon, T) \leq m_\varepsilon \sum_{n\geq M} 2^{-\delta n}\,,$$

which implies that

$$h_{\rm top}({}^K G, T; \varepsilon) \leq \sup\{h(T,\mu) \colon \mu \in K\}\,.$$

(2) is a consequence of (1). For the third statement notice that $G_K \subset {}^{\{\mu\}}G$ for all $\mu \in K$, so that

$$h_{\rm top}(T, G_K) \leq \inf\{h(T,\mu) \colon \mu \in K\}\,.$$

□

5. *Lower bound for $h_{\rm top}(T, G_K)$*

THEOREM 5.1. *Let $(X,d,T)$ be a dynamical system with the uniform separation and $\mathsf{g}$-almost product properties. Let $K$ be a connected non-empty compact subset of $M(X,T)$. Then*

$$\inf\{h(T,\mu) \colon \mu \in K\} \leq h_{\rm top}(T, G_K)\,.$$

**Proof**: For each $\varepsilon > 0$ there exists a finite sequence of measures $\alpha_1,\ldots,\alpha_n$ in $K$ such that each point of $K$ is within $\varepsilon$ of some $\alpha_j$. Because $K$ is connected, possibly repeating some $\alpha_j$, we can choose this sequence so that $\alpha_1,\ldots,\alpha_{n'}$ is not more than $\varepsilon$ away from any point of $K$ and $d(\alpha_j,\alpha_{j+1}) < 2\varepsilon$ for each $j$. Extending this argument we deduce that there exists a sequence $\alpha_1,\alpha_2,\ldots$ in $K$ so that the closure of $\{\alpha_j : j \in \mathbf{N},\ j > n\}$ for each $n \in \mathbf{N}$ equals $K$ and

$$\lim_{j\to\infty} d(\alpha_j,\alpha_{j+1}) = 0.$$

Let $\eta > 0$, and

$$h^* := \inf\{h(T,\mu) \colon \mu \in K\} - \eta\,.$$

Given this sequence of measures† $\{\alpha_k\}$, we construct a subset $G$ such that for each $x \in G$, $\{\mathcal{E}_n(x)\}$ has the same limit-point set as the sequence $\{\alpha_k\}$, and

† If $K = \{\alpha\}$, then $\alpha_k = \alpha$ for all $k$.

$h_{\rm top}(T,G) \geq h^*$. The construction of $G$ is the core of the proof. It is also used in the proof of Theorem 6.1.

By Corollary 3.1 we can find $\delta^* > 0$ and $\varepsilon^* > 0$ so that if $F \ni \mu$, then there exists $n^*_{F,\mu,\eta}$ with

$$N(F;\delta^*,n,\varepsilon^*) \geq 2^{n(h(T,\mu)-\eta)} \quad \forall\, n \geq n^*_{F,\mu,\eta}\,. \tag{16}$$

Let $\{\zeta_k\}$ and $\{\varepsilon_k\}$ be two strictly decreasing sequences†, $\lim_k \zeta_k = 0$ and $\lim_k \varepsilon_k = 0$, with $\varepsilon_1 < \varepsilon^*$. From (16) we deduce the existence of $n_k$ and a $(\delta^*, n_k, \varepsilon^*)$-separated subset $\Gamma_k \subset X_{n_k,\mathcal{B}(\alpha_k,\zeta_k)}$ with

$$|\Gamma_k| \geq 2^{n_k h^*}\,. \tag{17}$$

We may assume that $n_k$ satisfies

$$\delta^* n_k > 2\mathsf{g}(n_k) + 1 \quad \text{and} \quad \frac{\mathsf{g}(n_k)}{n_k} \leq \varepsilon_k\,. \tag{18}$$

The orbit-segments $\{x, Tx, \ldots, T^{n_k-1}x\}$, $x \in \Gamma_k$, are the building-blocks for the construction of the points of $G$. By Lemma 2.1 and (18)

$$x \in \Gamma_k \text{ and } y \in B_{n_k}(\mathsf{g}; x, \varepsilon_k) \implies \mathcal{E}_{n_k}(y) \in \mathcal{B}(\alpha_k, \zeta_k + 2\varepsilon_k)\,. \tag{19}$$

We choose a strictly increasing sequence $\{N_k\}$, with $N_k \in \mathbf{N}$,

$$n_{k+1} \leq \zeta_k \sum_{j=1}^{k} n_j N_j\,, \tag{20}$$

and

$$\sum_{j=1}^{k-1} n_j N_j \leq \zeta_k \sum_{j=1}^{k} n_j N_j\,. \tag{21}$$

Finally we define the (stretched) sequences $\{n'_j\}$, $\{\varepsilon'_j\}$ and $\{\Gamma'_j\}$, by setting for

$$j = N_1 + \cdots + N_{k-1} + q \quad \text{with} \quad 1 \leq q \leq N_k\,,$$

$$n'_j := n_k \quad \varepsilon'_j := \varepsilon_k \quad \Gamma'_j := \Gamma_k\,.$$

Let

$$G_k := \bigcap_{j=1}^{k} \Big( \bigcup_{x_j \in \Gamma'_j} T^{-M_{j-1}} B_{n'_j}(\mathsf{g}; x_j, \varepsilon'_j) \Big) \quad \text{with} \quad M_j := \sum_{\ell=1}^{j} n'_\ell\,.$$

$G_k$ is a non-empty closed set. We can label each set obtained by developing this formula by the branches of a labeled tree of height $k$. A branch is labeled by $(x_1, \ldots, x_k)$ with $x_j \in \Gamma'_j$. Theorem 5.1 is proved by proving Lemma 5.1.

LEMMA 5.1. *Let $\varepsilon$ be such that $4\varepsilon = \varepsilon^*$, and let*

$$G := \bigcap_{k \geq 1} G_k\,.$$

† One can take $\zeta_k = \varepsilon_k$; however the roles of $\zeta_k$ and $\varepsilon_k$ are different. See last remark of Section 5.

(1) *Let* $x_j, y_j \in \Gamma'_j$ *with* $x_j \neq y_j$. *If* $x \in B_{n'_j}(\mathsf{g}; x_j, \varepsilon'_j)$ *and* $y \in B_{n'_j}(\mathsf{g}; y_j, \varepsilon'_j)$, *then*

$$\max\{d(T^m x, T^m y) : m = 0, \ldots n_j - 1\} > 2\varepsilon\,.$$

(2) *G is a closed set, which is the disjoint union of non-empty closed sets* $G(x_1, x_2, \ldots)$ *labeled by* $(x_1, x_2, \ldots)$ *with* $x_j \in \Gamma'_j$. *Two different sequences label two different sets.*
(3) $G \subset G_K$.
(4) $h_{\rm top}(T, G) \geq h^*$.

**Proof**: (1) Let $x \in B_{n'_j}(\mathsf{g}; x_j, \varepsilon'_j)$ and $y \in B_{n'_j}(\mathsf{g}; y_j, \varepsilon'_j)$. Since $x_j$ and $y_j$ are $(\delta^*, n'_j, \varepsilon^*)$-separated and (18) holds, there exists $m \in \Lambda_{n'_j}$ such that

$$d(T^m x_j, T^m y_j) > 4\varepsilon\ ,\ d(T^m x_j, T^m x) \leq \varepsilon'_j\ ,\ d(T^m y_j, T^m y) \leq \varepsilon'_j\,.$$

But

$$d(T^m x, T^m y) \geq d(T^m x_j, T^m y_j) - d(T^m x_j, T^m x) - d(T^m y, T^m y_j) > 2\varepsilon\,.$$

(2) $G$ is the intersection of closed sets. Let $(x_1, x_2, \ldots)$ be a sequence with $x_j \in \Gamma'_j$. By the $\mathsf{g}$-almost product property and compactness

$$\bigcap_{j\geq 1} T^{-M_{j-1}} B_{n'_j}(\mathsf{g}; x_j, \varepsilon'_j)$$

is a non-empty closed set. By (1) the sets $B_{n'_j}(\mathsf{g}; x_j, \varepsilon'_j)$ and $B_{n'_j}(\mathsf{g}; y_j, \varepsilon'_j)$ are disjoint when $x_j \neq y_j$. Thus two different sequences label two different sets.
(3) We define the stretched sequence $\{\alpha'_m\}$ by

$$\alpha'_m := \alpha_k \quad \text{if} \quad \sum_{j=1}^{k-1} n_j N_j + 1 \leq m \leq \sum_{j=1}^{k} n_j N_j\,.$$

The sequence $\{\alpha'_m\}$ has the same limit-point set as the sequence $\{\alpha_k\}$. If

$$\lim_{n\to\infty} d(\mathcal{E}_n(y), \alpha'_n) = 0\,,$$

then the two sequences $\{\mathcal{E}_n(y)\}$ and $\{\alpha'_n\}$ have the same limit-point set. Because of (20) and the definition of $\{\alpha'_m\}$, it is sufficient to show that

$$\lim_{k\to\infty} d(\mathcal{E}_{M_k}(y), \alpha'_{M_k}) = 0\,.$$

Suppose that $\sum_{\ell=1}^{j} n_\ell N_\ell < M_k \leq \sum_{\ell=1}^{j+1} n_\ell N_\ell$; hence $\alpha'_{M_k} = \alpha_{j+1}$. By Lemma 2.1, (19) and (21)

$$\begin{aligned} d(\mathcal{E}_{M_k}(y), \alpha'_{M_k}) \leq & \frac{\sum_{\ell=1}^{j-1} n_\ell N_\ell}{M_k} d(\mathcal{E}_{\sum_{\ell=1}^{j-1} n_\ell N_\ell}(y), \alpha'_{M_k}) + \frac{n_j N_j}{M_k}\big(\zeta_j + 2\varepsilon_j + d(\alpha_j, \alpha_{j+1})\big) \\ & + \frac{M_k - \sum_{\ell=1}^{j} n_\ell N_\ell}{M_k}(\zeta_{j+1} + 2\varepsilon_{j+1}) \\ \leq & 2\zeta_j + 2\varepsilon_j + d(\alpha_j, \alpha_{j+1}) + \zeta_{j+1} + 2\varepsilon_{j+1}\,. \end{aligned}$$

Since $\lim_j \zeta_j = 0$, $\lim_j \varepsilon_j = 0$ and $\lim_j d(\alpha_j, \alpha_{j+1}) = 0$ this proves (3).

(4) For this part of the proof the details of the construction are unimportant. What is required is that $\lim_n M_{n+1}/M_n = 1$ and $|\Gamma_{n+1}| \geq 2^{h^*(M_{n+1}-M_n)}$. Let $s < h^*$. We prove that $C(G; s, \varepsilon, T) \geq 1$. Since $G$ is compact we can consider finite covers $\mathcal{C}$ of $G$ with the property that if $B_m(x,\varepsilon) \in \mathcal{C}$, then $B_m(x,\varepsilon) \cap G \neq \emptyset$. By definition

$$C(G; s, n, \varepsilon, T) = \inf_{\mathcal{C} \in \mathcal{G}_n(G,\varepsilon)} \sum_{B_m(x,\varepsilon) \in \mathcal{C}} 2^{-sm}\,.$$

For each $\mathcal{C} \in \mathcal{G}_n(G,\varepsilon)$ we define the cover $\mathcal{C}'$ in which each ball $B_m(z,\varepsilon)$ is replaced by $B_{M_p}(z,\varepsilon)$, $M_p \leq m < M_{p+1}$. Then

$$C(G; s, n, \varepsilon, T) = \inf_{\mathcal{C} \in \mathcal{G}_n(G,\varepsilon)} \sum_{B_m(x,\varepsilon) \in \mathcal{C}} 2^{-sm} \geq \inf_{\mathcal{C} \in \mathcal{G}_n(G,\varepsilon)} \sum_{B_{M_p}(x,\varepsilon) \in \mathcal{C}'} 2^{-sM_{p+1}}.$$

Consider a specific $\mathcal{C}'$ and let $m$ be the largest value of $p$ for which there exists $B_{M_p}(z,\varepsilon) \in \mathcal{C}'$. Define

$$\mathcal{W}_k := \prod_{i=1}^{k} \Gamma_i,\ \overline{\mathcal{W}}_m := \bigcup_{k=1}^{m} \mathcal{W}_k.$$

Each $x \in B_{M_p}(z,\varepsilon) \cap G$ corresponds to a point in $\mathcal{W}_p$. Lemma 5.1 (1) implies that this point is uniquely defined. For $1 \leq j \leq k$, the word $\mathrm{v} \in \mathcal{W}_j$ is a prefix of $\mathrm{w} \in \mathcal{W}_k$ if the first $j$ entries of w coincide with v. Note that each $\mathrm{w} \in \mathcal{W}_k$ is the prefix of exactly $|\mathcal{W}_m|/|\mathcal{W}_k|$ words of $\mathcal{W}_m$. If $\mathcal{W} \subset \overline{\mathcal{W}}_m$ contains a prefix of each word of $\mathcal{W}_m$, then

$$\sum_{k=1}^{m} |\mathcal{W} \cap \mathcal{W}_k|\, |\mathcal{W}_m|/|\mathcal{W}_k| \geq |\mathcal{W}_m|.$$

Thus if $\mathcal{W}$ contains a prefix of each word of $\mathcal{W}_m$,

$$\sum_{k=1}^{m} |\mathcal{W} \cap \mathcal{W}_k|/|\mathcal{W}_k| \geq 1. \tag{22}$$

Note that since $\mathcal{C}'$ is a cover, each point of $\mathcal{W}_m$ has a prefix associated with some $B_{M_p}(z,\varepsilon) \in \mathcal{C}'$. Also $|\mathcal{W}_k| \geq 2^{h^* M_k}$. Hence

$$\sum_{B_{M_p}(x,\varepsilon) \in \mathcal{C}'} 2^{-h^* M_p} \geq 1.$$

Thus when $p$ is large enough so that $k \geq p$ implies $sM_{k+1} \leq h^* M_k$, $n \geq M_p$ and $\mathcal{C} \in \mathcal{G}_n(G,\varepsilon)$, we have

$$\sum_{B_m(x,\varepsilon) \in \mathcal{C}} 2^{-sm} \geq 1 \Rightarrow C(G; s, n, \varepsilon, T) \geq 1.$$

□

**Remark.** Inequality (22) is a complement to Kraft's inequality of coding theory [**S**], which has the reversed inequality. For Kraft's inequality one requires a mapping into $\overline{\mathcal{W}}_m$ for which no word is a prefix of another. In the current context we have a mapping in which a prefix for each word of $\mathcal{W}_m$ appears.

**Remark.** When the dynamical system is expansive with expansive constant $\varepsilon^*$,

$$\lim_{n\to\infty} \sup\{\mathrm{diam} B_n(x,\varepsilon^*) : x \in X\} = 0\,.$$

Hence, in the proof of Theorem 5.1, we can work with a fixed $\varepsilon = \varepsilon^*$ and use

LEMMA 5.2. *Let $(X,d,T)$ be an expansive dynamical system with expansive constant $\varepsilon^*$. Let $f : X \to \mathbf{R}$ be a continuous function, $\|f\| \leq 1$. Then, for any $\delta > 0$ there exists $N(\delta)$ so that for any $n \geq N(\delta)$, any $x \in X$, and any $y \in B_n(\mathsf{g}; x, \varepsilon^*)$,*

$$\big|\langle\, f, \mathcal{E}_n(x)\,\rangle - \langle\, f, \mathcal{E}_n(y)\,\rangle\big| \leq \delta\,.$$

**Proof**: Let $\delta > 0$. Since $f$ is uniformly continuous, there exists $\eta$ so that $d(x,y) \leq \eta$ implies $|f(x) - f(y)| \leq \delta/2$. We choose $M \in \mathbf{N}$ so that for any $m \geq M$,

$$\sup\{\mathrm{diam} B_m(\mathsf{g}; x, \varepsilon^*) :\; x \in X\} \leq \eta\,.$$

Then we choose $N > M$ so that for any $n \geq N$

$$\frac{M(\mathsf{g}(n)+1)}{n} \leq \frac{\delta}{2}\,.$$

We set $N(\delta) := N$. Let $n \geq N(\delta)$ and $y \in B_n(\mathsf{g}; x, \varepsilon^*)$. By definition there exists $\Lambda \subset \Lambda_n$ so that $|\Lambda_n \backslash \Lambda| \leq \mathsf{g}(n)$ and $d_\Lambda(x,y) \leq \varepsilon^*$. We say that $k \in \Lambda_n$ is *bad* if there exists $j$, $k \leq j \leq k + M - 1$ so that $j \notin \Lambda$. The number of bad $k$ is at most $M(\mathsf{g}(n)+1)$. If $k$ is not bad, then $T^k y \in B_M(T^k x, \varepsilon^*)$, so that $d(T^k y, T^k x) \leq \eta$ and consequently $|f(T^k y) - f(T^k x)| \leq \delta/2$. Hence

$$\big|\langle\, f, \mathcal{E}_n(x)\,\rangle - \langle\, f, \mathcal{E}_n(y)\,\rangle\big| \leq \frac{n - M(\mathsf{g}(n)+1)}{n}\,\frac{\delta}{2} + \frac{M(\mathsf{g}(n)+1)}{n} \leq \delta\,.$$

□

6. *Non-uniform case*

In this section we drop the condition of uniform separation. Let $\varepsilon > 0$, $\delta > 0$ and $\nu \in M(X,T)$. We set

$$\underline{s}(\nu;\delta,\varepsilon) := \inf_{F \ni \nu} \liminf_{n\to\infty} \frac{1}{n} \log_2 N(F;\delta,n,\varepsilon)\,.$$

Similarly we define $\overline{s}(\nu;\delta,\varepsilon)$. In this definition we can compute the infimum using the base of neighborhoods of $\nu$ given by the balls $\mathcal{B}(\nu,\zeta)$. Since $N(F;\delta,n,\varepsilon) \leq N(X;n,\varepsilon)$, Theorem 7.7 and remark (8), p.166, in **[W]** imply that $\underline{s}(\nu;\delta,\varepsilon) \leq \underline{s}(\nu;\varepsilon) < \infty$ and $\overline{s}(\nu;\delta,\varepsilon) \leq \overline{s}(\nu;\varepsilon) < \infty$.

LEMMA 6.1. *The functions $\underline{s}(\,\cdot\,;\delta,\varepsilon)$ and $\overline{s}(\,\cdot\,;\delta,\varepsilon)$ are upper semi-continuous functions.*

**Proof**: Let $\lim_k \mu_k = \nu$. Let $a > \underline{s}(\nu;\delta,\varepsilon)$. Then there exists an open set $F \ni \nu$ such that

$$a > \liminf_{n\to\infty} \frac{1}{n}\log_2 N(F;\delta,n,\varepsilon)\,.$$

If $k$ is large enough, then $\mu_k \in F$ and therefore

$$a > \liminf_{n\to\infty} \frac{1}{n}\log_2 N(F;\delta,n,\varepsilon) \geq \underline{s}(\mu_k;\delta,\varepsilon)\,,$$

which implies that

$$\underline{s}(\nu;\delta,\varepsilon) \geq \limsup_{k\to\infty} \underline{s}(\mu_k;\delta,\varepsilon)\,.$$

Similar proof for $\overline{s}(\,\cdot\,;\delta,\varepsilon)$. □

Lemma 6.2. *Let $\varepsilon^* > 0$, $\delta^* > 0$. Let $\nu \in M(X,T)$ and*

$$\nu = \int \mu_t\,\rho(dt)$$

*be its ergodic decomposition. Then, for any $\Delta > 0$ we can find a finite convex combination $\sum_{i=1}^p a_i\mu_i$ of ergodic measures such that*

$$d\big(\nu, \sum_{i=1}^p a_i\mu_i\big) \leq \Delta\,,$$

*and*

$$\int \underline{s}(\mu_t,\delta^*,\varepsilon^*)\,\rho(dt) \leq \sum_{i=1}^p a_i\underline{s}(\mu_i,\delta^*,\varepsilon^*)\,.$$

*The $\{a_i\}$ can be chosen to be rational numbers. Moreover, if $h^* < h(T,\nu)$, then by choosing $\varepsilon^*$ and $\delta^*$ sufficiently small*

$$\int \underline{s}(\mu_t;\delta^*,\varepsilon^*)\,\rho(dt) > h^*\,.$$

**Proof**: Let $\{A_1,\ldots,A_p\}$ be a partition of $M(X,T)$ with diameter smaller than $\Delta$. For any $A_i$ there exists an ergodic $\mu_i \in A_i$ such that

$$\frac{1}{\rho(A_i)}\int_{A_i} \underline{s}(\mu_t;\delta^*\varepsilon^*)\,\rho(dt) \leq \underline{s}(\mu_i;\delta^*\varepsilon^*)\,,$$

and

$$d\Big(\frac{1}{\rho(A_i)}\int_{A_i}\mu_t\,\rho(dt),\mu_i\Big) \leq \Delta\,.$$

Setting $a_i := \rho(A_i)$, we get the first result. To get rational coefficients, do the above for $\Delta/2$ and note that if $2pn < \Delta$, each $a_j$ may be made rational by adjusting by not more than $1/n$ while maintaining the required properties. By Corollary 3.2

$$\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\underline{s}(\nu;\delta,\varepsilon) = h(T,\nu) \quad \text{if } \nu \text{ ergodic}\,.$$

From the Monotone-Convergence Theorem and the affine character of the entropy we get

$$\lim_{\varepsilon\to 0}\lim_{\delta\to 0}\int \underline{s}(\mu_t;\delta,\varepsilon)\,\rho(dt) = \int \lim_{\varepsilon\to 0}\lim_{\delta\to 0}\underline{s}(\mu_t;\delta,\varepsilon)\,\rho(dt) = \int h(T,\mu_t)\,\rho(dt) = h(T,\nu)\,.$$

Since $\underline{s}(\mu_t;\delta,\varepsilon)$ is nonincreasing in $\delta$ and $\varepsilon$, this proves also the second result. □

THEOREM 6.1. *Let $(X, d, T)$ be a dynamical system verifying the* $\mathsf{g}$*-almost product property. Then*

$$h_{\rm top}(T, G_\nu) = h(T, \nu) \quad \textit{for any } \nu \in M(X, T)\,.$$

**Proof**: The proof follows the same pattern as the proof of Theorem 5.1. Let

$$h(T, \nu) > h' > h^*\,.$$

We prove

$$h_{\rm top}(T, G_\nu) \geq h^*\,.$$

By Lemma 6.2 we can find $\varepsilon^* > 0$, $\delta^* > 0$, and for any $k$, a finite convex combination of ergodic measures with rational coefficients

$$\alpha_k := \sum_{i=1}^{p_k} a_{i,k}\mu_{i,k}\,,$$

so that $\{\alpha_k\}$ converges to $\nu$, and such that

$$h' < \sum_{i=1}^{p_k} a_{i,k}\underline{s}(\mu_{i,k}, \delta^*, \varepsilon^*)\,. \tag{23}$$

Let $\{\zeta_k\}$ and $\{\varepsilon_k\}$ be two strictly decreasing sequences, $\lim_k \zeta_k = 0$ and $\lim_k \varepsilon_k = 0$, with $\varepsilon_1 < \varepsilon^*$. We assume that

$$d(\nu, \alpha_k) \leq \zeta_k\,.$$

For each $\alpha_k$ there exists an integer $n_k$ so each element of $\{n_k a_{i,k}\}$ is an integer,

$$N(\mathcal{B}(\mu_{i,k}, \zeta_k); \delta^*, a_{i,k}n_k, \varepsilon^*) \geq 2^{a_{i,k}n_k(\underline{s}(\mu_{i,k},\delta^*,\varepsilon^*)-\eta)} \quad i = 1, \ldots, p_k\,, \tag{24}$$

where $\eta := h' - h^*$, and

$$\delta^* a_{i,k}n_k > 2\mathsf{g}(a_{i,k}n_k) + 1 \quad \text{and} \quad \frac{\mathsf{g}(a_{i,k}n_k)}{a_{i,k}n_k} \leq \varepsilon_k \quad \forall\ i = 1, \ldots, p_k. \tag{25}$$

Let $\Gamma_{i,k}$ be a $(\delta^*, a_{i,k}n_k, \varepsilon^*)$-separated subset of $X_{a_{i,k}n_k, \mathcal{B}(\mu_{i,k},\zeta_k)}$. The cardinality of $\Gamma_{i,k}$ is at least (see (24))

$$2^{a_{i,k}n_k(\underline{s}(\mu_{i,k},\delta^*,\varepsilon^*)-\eta)}\,.$$

We define

$$\Gamma_k := \prod_{i=1}^{p_k} \Gamma_{i,k},$$

so that by (23) and (24)

$$|\Gamma_k| \geq 2^{n_k h^*}\,.$$

We denote the elements of $\Gamma_k$ by

$$\mathbf{x}_k := (x_{1,k}, \ldots, x_{p_k,k}) \quad \text{with} \quad \mathcal{E}_{a_{i,k}n_k}(x_{i,k}) \in \mathcal{B}(\mu_{i,k}, \zeta_k)\,.$$

and set

$$B_{n_k}(\mathsf{g}; \mathbf{x}_k, \varepsilon_k) := \bigcap_{i=1}^{p_k} T^{-(a_{1,k}+\cdots+a_{i-1,k})n_k} B_{a_{i,k}n_k}(\mathsf{g}; x_{i,k}, \varepsilon_k) \quad (\text{with } a_{0,k} := 0)\,.$$

We omit the rest of the proof, which from this point follows that of Theorem 5.1. □

LEMMA 6.3. *Assume that $(X, d, T)$ has the $\mathtt{g}$-almost product property, and let $\delta > 0$ and $\varepsilon > 0$. Let $a_1 > 0, \ldots, a_p > 0$, with $a_1 + \cdots + a_p = 1$. Then, for $\nu_i \in M(X,T)$, $\varepsilon' < \varepsilon$ and $\delta' < (\min_i a_i)\,\delta$,*

$$\underline{s}\Big(\sum_{i=1}^{p} a_i\nu_i; \delta', \varepsilon'\Big) \geq \sum_{i=1}^{p} a_i\,\underline{s}(\nu_i; \delta, \varepsilon)\,.$$

**Proof**: Let $\mathtt{m}$ be the function of Definition 2.3. Let $\varepsilon' < \varepsilon$ and $\delta' < (\min_i a_i)\cdot\delta$ be given. We fix $\zeta' > 0$, and choose $\Delta$ and $\zeta$ so small that

$$2\Delta < \varepsilon - \varepsilon' \quad \text{and} \quad \zeta + \Delta < \zeta'\,.$$

We set $F_i = \mathcal{B}(\nu_i, \zeta)$, $F = \mathcal{B}(\nu, \zeta')$ with $\nu = \sum_i a_i\nu_i$. Let $\eta > 0$ and $n$ be a (large) integer. We decompose $n$ into $n_1 + \cdots + n_p = n$, with $n_i \in \mathbf{N}$ and $\lfloor a_i n \rfloor \leq n_i \leq \lceil a_i n \rceil$. There exists $n^*$ such that for $n \geq n^*$, $(\min_i a_i)\,n \geq \mathtt{m}(\Delta)$ and

$$N(F_i; \delta, n_i, \varepsilon) \geq 2^{a_i n(\underline{s}(\nu_i) - \eta)} \quad i = 1, \ldots, p\,.$$

We also assume that for $n \geq n^*$ and all $i$

$$\zeta + \Delta + \frac{\mathtt{g}(n_i)}{n_i} < \zeta' \quad \text{and} \quad n_i\delta > 2\mathtt{g}(n_i) + 1\,.$$

Suppose that $\mathcal{E}_n(x_i) \in F_i$, $i = 1, \ldots, p$. For each choice of $x_1, \ldots, x_p$ we select $y$ in

$$\bigcap_{j=1}^{p} T^{-(n_1+\cdots+n_{j-1})} B_{n_j}(\mathtt{g}; x_j, \Delta) \neq \emptyset\,.$$

In this way we define a subset $S$ of cardinality at least $2^{(\underline{s}(\nu_1;\varepsilon)-\eta)a_1 n} \cdots 2^{(\underline{s}(\nu_p;\varepsilon)-\eta)a_p n}$, such that

$$y \in S \implies \mathcal{E}_n(y) \in \mathcal{B}(\nu, \zeta')\,.$$

The subset $S$ is $(\delta'', n, \varepsilon')$-separated, with

$$\delta'' = \Big(\min_i n_i\Big[\delta - \frac{2\mathtt{g}(n_i)}{n_i}\Big]\Big)\frac{1}{n}\,.$$

Since $a_i$ are fixed, by choosing $n$ large enough,

$$\delta'' \geq \delta' \quad \text{if } \delta' < (\min_i a_i)\,\delta\,.$$

Therefore

$$\liminf_{n\to\infty} \frac{1}{n}\log_2 N(\mathcal{B}(\nu, \zeta'); \delta', n, \varepsilon') \geq \sum_{j=1}^{p} \underline{s}(\nu_j; \delta, \varepsilon) - \eta\,.$$

We may take $\Delta$ arbitrarily small, as well as $\zeta$, so that we can take the limit $\zeta' \to 0$. Thus

$$\underline{s}(\nu; \delta', \varepsilon') \geq \sum_{j=1}^{p} \underline{s}(\nu_j; \delta, \varepsilon) - \eta\,.$$

Since $\eta$ is also arbitrary, this proves the lemma. □

Proposition 6.1. *Assume that $(X,d,T)$ has the* g*-almost product property. Then*

$$s(\nu)=h(T,\nu)\quad \textit{if}\ \nu\in M(X,T)\,.$$

**Proof**: It is sufficient to prove $\lim_\varepsilon \underline{s}(\nu;\varepsilon)=\underline{s}(\nu)\geq h^*$, for any $h^*<h(T,\nu)$. By Lemma 6.2 we can find $\varepsilon^*>0$, $\delta^*>0$, and a sequence $\{\nu_k\}$ such that $\lim_k\nu_k=\nu$, each $\nu_k$ is a finite convex combinations of ergodic measures,

$$\nu_k=\sum_{i=1}^{n_k}a_{i,k}\mu_{i,k}\quad\text{and}\quad\sum_{i=1}^{n_k}a_{i,k}\underline{s}(\mu_{i,k};\delta^*,\varepsilon^*)\geq h^*\,.$$

By Lemma 6.3, if $\varepsilon<\varepsilon^*$, then

$$h^*\leq\sum_{i=1}^{n_k}a_{i,k}\underline{s}(\mu_{i,k};\delta^*,\varepsilon^*)\leq\lim_{\delta\to0}\underline{s}\big(\sum_{i=1}^{n_k}a_{i,k}\mu_{i,k};\delta,\varepsilon\big)\leq\underline{s}\big(\sum_{i=1}^{n_k}a_{i,k}\mu_{i,k};\varepsilon\big)\,.$$

By Lemma 6.1 we get

$$\underline{s}(\nu)=\lim_{\varepsilon'\to0}\underline{s}(\nu;\varepsilon')\geq\underline{s}(\nu;\varepsilon)\geq\limsup_{k\to\infty}\underline{s}(\nu_k;\varepsilon)\geq h^*\,.$$

□

## 7. *Multi-fractal analysis of ergodic averages*

We present one application of the above results. For variants see Section 5 in [**PfS2**].

Proposition 7.1. *Let $(X,d,T)$ be a dynamical system such that for all $\mu\in M(X,T)$, $h_{\rm top}(T,G_\mu)=h(T,\mu)$. Let $Y$ be a topological vector space and $\varphi\colon X\to Y$. Suppose that the map $\phi\colon M(X)\to Y$,*

$$\mu\mapsto\phi(\mu):=\langle\,\varphi,\mu\,\rangle$$

*is continuous. For any $a\in Y$ we set*

$$K_a:=\Big\{x\in X\colon\lim_n\frac{1}{n}\sum_{k=0}^{n-1}\varphi(T^kx)=a\Big\}\,.$$

*Then*

$$h_{\rm top}(T,K_a)=\sup\{h(T,\rho)\colon\rho\in M(X,T)\ \textit{and}\ \langle\,\varphi,\rho\,\rangle=a\}\,.$$

**Proof**: Let $F(a):=\{\rho\in M(X,T)\colon\langle\,\varphi,\rho\,\rangle=a\}$. This is a closed set. Because $\mu\mapsto\phi(\mu):=\langle\,\varphi,\mu\,\rangle$ is continuous, the statement

$$\lim_n\frac{1}{n}\sum_{k=0}^{n-1}\varphi(T^kx)=a$$

is equivalent to the statement

$$\{\mathcal{E}_n(x)\}\ \text{has all its limit-points in}\ F(a)\,.$$

Let

$$G^{F(a)} := \{x \in X : \{\mathcal{E}_n(x)\} \text{ has all its limit points in } F(a)\}\,.$$

For any $\rho \in F(a)$ one has $G_\rho \subset G^{F(a)}$, hence

$$\sup\{h(T,\rho) : \rho \in F(a)\} \leq h_{\rm top}(T, G^{F(a)})\,.$$

For the other inequality notice that $G^{F(a)} \subset {}^{F(a)}G$, which implies by Theorem 4.1 (1)

$$\sup\{h(T,\rho) : \rho \in F(a)\} \geq h_{\rm top}(T, G^{F(a)})\,.$$

□

Proposition 7.1 is the main result in [**TV**]. For earlier results in the context of subshifts of finite type, see [**O1**]. See Theorem 6.2 in [**TV**] for the relationship between $h_{\rm top}(T, K_a)$ and the topological pressure

$$P(T,\varphi) = \sup\{h(T,\rho) + \langle\, \varphi, \rho\,\rangle : \rho \in M(X,T)\}\,,$$

when the map $\varphi$ is real-valued. Takens and Verbitskiy use, independently, a method similar to the method of proof of Section 5, under the assumptions of the specification property and $h_{\rm top}(T,X) < \infty$. Our proof is different and we prove that the hypothesis of Proposition 7.1 are verified for a dynamical system with the $g$-almost product property only. We emphasize the role of the empirical measures $\mathcal{E}_n$. There is an analogy with Large Deviations Theory (see [**LePf**]). In the language of Large Deviations Theory, one could say that Takens and Verbitskiy proved a level-1 result and that we prove a level-3 result, and that Proposition 7.1 is a contraction principle.

The two main properties which are used in this paper are the entropy density of ergodic measures and the uniform separation property. This later property implies that the topological entropy is finite and that the entropy-map is upper semi-continuous. We give an example of a dynamical system with *finite* topological entropy, for which entropy-density of ergodic measures is true (specification property is true), but the uniform separation property and the upper semi-continuity of the entropy map fail.

**Example.**

Consider the shift space $Y := [-1,1]^{\mathbf{Z}_+}$, with the metric

$$d(x,y) := \sum_{k=0}^{\infty} 2^{-k-2}|x_k - y_k|\,.$$

Let $\mathtt{A}$ be the alphabet containing the letters $a_0 = 0$, $a_m = 1/m$, and $a_{-m} = -1/m$, $m \in \mathbf{N}$. We consider $\mathtt{A}$ as a subset of $[-1,1]$ with the induced metric, so that it is a compact set. Let $X \subset \mathtt{A}^{\mathbf{Z}_+}$ be the subshift defined‡ by

$$x_{k+1} = \begin{cases} a_j, \text{ or } a_{-j}, \text{ or } a_{j+1}, \text{ or } a_{j-1} & \text{if } x_k = a_j,\ j \geq 1 \\ a_j, \text{ or } a_{-j}, \text{ or } a_{-j+1}, \text{ or } a_{-j-1} & \text{if } x_k = a_{-j},\ j \geq 1 \\ a_0, \text{ or } a_1, \text{ or } a_{-1} & \text{if } x_k = a_0 \end{cases} \quad \forall\, k \in \mathbf{Z}_+\,.$$

‡ We were inspired by [**Gu**].

There is an injective map from $X$ to $\mathtt{A} \times \{-2,-1,1,2\}^{\mathbf{N}}$. Let $x \in X$ and $y = (y_0, y_1, \ldots) \in \mathtt{A} \times \{-2,-1,1,2\}^{\mathbf{N}}$; we set $y_0 := x_0$ and for $k \geq 1$

$$y_k := \begin{cases} 1 & \text{if } x_{k-1} = x_k \\ -1 & \text{if } x_{k-1} = -x_k \\ 2 & \text{if, for } j \in \mathbf{N},\ x_{k-1} = |a_j| \text{ and } x_k = |a_{j-1}| \\ -2 & \text{if, for } j \in \mathbf{N},\ x_{k-1} = |a_j| \text{ and } x_k = |a_{j+1}| \\ 2 & \text{if } x_{k-1} = a_0 \text{ and } x_k = a_1 \\ -2 & \text{if } x_{k-1} = a_0 \text{ and } x_k = a_{-1}. \end{cases}$$

From this it follows that the topological entropy of $X$ is bounded by $\log_2 4$. We show that the subshift $X$ has the specification property. Let $\varepsilon_0 = 1/m_0$, $m_0 \in \mathbf{N}$ be fixed. Consider a word of length $n$, say $\mathrm{w} = (x_0, x_1, \ldots, x_{n-1})$. We define a new word of length $n$, denoted by $\widehat{\mathrm{w}} = (\widehat{x}_0, \widehat{x}_1, \ldots, \widehat{x}_{n-1})$,

$$\widehat{x}_k := \begin{cases} x_k & \text{if } x_k = a_j \text{ or } x_k = a_{-j}, \text{ with } j \leq m_0 \\ a_{m_0} & \text{if } x_k = a_j \text{ with } j > m_0 \\ a_{-m_0} & \text{if } x_k = a_{-j} \text{ with } j > m_0. \end{cases} \tag{26}$$

We can also extend the transformation $\widehat{\cdot}$ to the points of $X$. Notice that the image of $X$ under this transformation, $\widehat{X}$, is a subset of $X$, and

$$d(T^m x, T^m \widehat{x}) \leq \frac{1}{2m_0} \quad \forall\, m \in \mathbf{Z}_+\,.$$

Let $j_0$ be defined by

$$j_0 := \min\Big\{ j \in \mathbf{Z}_+ : 2^{-(j+1)} < \frac{1}{2m_0} \Big\}\,, \tag{27}$$

so that

$$\sum_{k>j_0}^{\infty} 2^{-k-2}|x_k - y_k| < \frac{1}{2m_0}\,.$$

The subshift $X$ has the specification property with $\mathtt{k}(\varepsilon_0) := j_0 + m_0 + 1$. Let $n_1 \in \mathbf{N}, \ldots, n_p \in \mathbf{N}$, $x_1 \in X, \ldots, x_p \in X$ and $k_1 \geq \mathtt{k}(\varepsilon_0), \ldots, k_{p-1} \geq \mathtt{k}(\varepsilon_0)$ be given. Let $\mathrm{w}_i$ be the prefix of $\widehat{x}_i$ of length $n_i + j_0$, $i = 1, \ldots, p$. Then we can find $y \in X$ such that, for $i = 1, \ldots, p$

$$(y_i, y_{i+1} \ldots, y_j) = \mathrm{w}_i \quad i = \sum_{\ell=1}^{i-1}(n_\ell + k_\ell) \text{ and } j = i + n_i + j_0 - 1\,.$$

For $y$ we have

$$\begin{aligned} d(T^{n_1+k_1+\cdots+n_{j-1}+k_{j-1}+m}y, T^m x_j) &\leq d(T^{n_1+k_1+\cdots+n_{j-1}+k_{j-1}+m}y, T^m \widehat{x}_j) \\ &\quad + d(T^m \widehat{x}_j, T^m x_j) \\ &\leq \frac{1}{m_0} \quad m = 0, \ldots, n_j - 1 \quad \text{and} \quad j = 1, \ldots, p\,. \end{aligned}$$

Thus, $(X, d, T)$ is a dynamical system with finite topological entropy and the specification property. It is neither expansive, nor asymptotically expansive; indeed, for $x^0$, with $x^0_k = 0$ for all $k \in \mathbf{Z}_+$,

$$h_{\mathrm{top}}\Big(\bigcap_n B_n(x^0, \varepsilon)\Big) \geq \log_2 2\,.$$

We show that the entropy-map is not upper semi-continuous at $\delta_{x^0}$. Indeed, let $\mu_m$ be the Bernoulli measure concentrated on the configurations with $x_k = a_{\pm m}$ for all $k$. These measures converge weakly to $\delta_{x^0}$, but $h(T, \mu_m) = \log_2 2$ and $h(T, \delta_{x^0}) = 0$.

Choose $\varphi(x) := |x_0|$, $\mu := \delta_{x^0}$ so that $\langle\, \varphi, \mu\,\rangle = 0$. Let $F_\delta$ be the neighborhood

$$F_\delta := \{\nu : |\langle\, \varphi, \nu\,\rangle - \langle\, \varphi, \mu\,\rangle| < \delta\}\,.$$

Then

$$X_{n,F_\delta} = \Big\{x \in X : \frac{1}{n}\sum_0^{n-1} |x_k| < \delta\Big\}\,. \tag{28}$$

We shall prove that for each $\varepsilon > 0$

$$\lim_{\delta\to 0} \limsup_n \frac{1}{n} \log_2 N(F_\delta; n, \varepsilon) = 0\,. \tag{29}$$

Each neighbourhood $F_\delta$, $\delta > 0$, contains ergodic measures of the form $\mu_m$ with $h(T, \mu_m) = \log_2 2$. Hence (29) implies that $X$ does not have the uniform separation property.

It is convenient to work with $(n, \varepsilon)$-spanning sets instead of $(n, \varepsilon)$-separated sets. Let $\varepsilon = 1/m_0$, $m_0 \in \mathbf{N}$. Let $\delta > 0$ so that $m_0\delta < 1/2$, and $j_0$ be defined as in (27). We select an $(n, \varepsilon)$-spanning set $\mathcal{S}(\delta)$ for $X_{n,F_\delta}$ and bound its cardinality. Let

$$\mathcal{S}(\delta) := \Big\{x \in \{a_j : |j| \leq m_0\}^{Z_+} \cap X : x_k = x_{n+j_0-1}, k \geq n + j_0;\ \sum_0^{n-1} |x_k| < n\delta\Big\}\,.$$

Since $|x_k| \geq m_0^{-1}$ when $x_k \neq 0$, $x \in \mathcal{S}(\delta)$ has at most $n\delta m_0$ nonzero terms among its first $n$ coordinates. The number of possible subsets with these nonzero coordinates is at most

$$\sum_{k \leq (\delta m_0)n} \binom{n}{k} \leq 2^{n\phi(m_0\delta)}\,,$$

where $\phi$ is the entropy-function (13). Therefore the cardinality of $\mathcal{S}(\delta)$ is at most

$$2^{n\phi(m_0\delta)}(2m_0)^{m_0\delta n}(2m_0 + 1)^{j_0}\,.$$

It remains to verify that $\mathcal{S}(\delta)$ is $(n, \varepsilon)$-spanning with $\varepsilon = 1/m_0$. Let $y \in X_{n,F_\delta}$ and let $\widetilde{y}$ be the element of $\mathcal{S}(\delta)$ such that

$$\widetilde{y}_k = \widehat{y}_k \text{ for } k < n + j_0,$$

where $\widehat{y}_k$ is defined as in (26). We have $|y_k - \widetilde{y}_k| \leq m_0^{-1}$ for all $k < n + j_0$; for $0 \leq m < n$ it follows from (27) that

$$d(T^m y, T^m \widetilde{y}) \leq \sum_{k=0}^{j_0} 2^{-k-2}|y_{m+k} - \widetilde{y}_{m+k}| + \sum_{k>j_0} 2^{-k-2}|y_{m+k} - \widetilde{y}_{m+k}|$$
$$\leq \frac{1}{m_0}\sum_{k=0}^{j_0} 2^{-k-2} + \frac{1}{2m_0} \leq \frac{1}{m_0}\,.$$

From the above estimates (29) follows.

Notice that we could use here Proposition 3.3 to conclude that the uniform separation property does not hold. Notice also that the proof of Theorem 5.1 in [**TV**] is incomplete. Without additional hypothesis one cannot satisfy points (1), (2) and (3) of p.328. As this example shows, it is not sufficient to assume finiteness of the topological entropy and the specification property.

## References


[**BaSc**] L. Barreira, J. Schmeling, Sets of non-typical points have full topological entropy and full Hausdorff dimension, *J. Isr. Math.* **116**, 29-70 (2000).

[**Bo1**] R. Bowen, Entropy-expansive maps, *Trans. A.M.S.*, **164**, 323-331 (1972).

[**Bo2**] R. Bowen, Topological entropy for noncompact sets, *Trans. A.M.S.*, **184**, 125-136 (1973).

[**BK**] M. Brin, A. Katok, On local entropy, in Geometric Dynamic, 30-38, Ed. J. Palis Jr., LNM 1007, Springer, Berlin (1983).

[**C**] H. Cajar, Billingsley Dimension in Probability Spaces, Springer LMN 892, Berlin (1980).

[**Ch**] D.G. Champernowne, The construction of decimal normal in the scale of ten, *J. London Math. Soc.* **8**, 254-260 (1933).

[**Co**] C.M. Colebrook, The Hausdorff dimension of certain sets of nonnormal numbers, *Michigan Math. J.* **17** 103–116 (1970).

[**Eg**] H.G. Eggleston, The fractional dimension of a set defined by decimal properties, *Quart. J. Math. Oxford Ser.* **20**, 31-36 (1949).

[**G**] L.W. Goodwyn, Topological entropy bounds measure-theoretic entropy, *Proc. A.M.S.*, **23**, 679-688 (1969).

[**Gu**] B.M. Gurevič, Topological entropy of enumerable Markov chains, *Soviet Math. Dokl.* **10**, 911-915 (1969).

[**K**] A. Katok, Lyapunov exponents, entropy and periodic orbits for diffeomorphisms, p.137-145 *IHES Publications mathématiques*, **51**, 137-173 (1980).

[**LePf**] J.T. Lewis, C.-E. Pfister, Thermodynamic probability theory: some aspects of large deviations, *Russian Mathematical Surveys* **50**, 279-317 (1995).

[**Mi1**] M. Misiurewicz, Topological conditional entropy, *Studia Mathematica*, **55**, 175-200 (1976).

[**Mi2**] M. Misiurewicz, A short proof of the variational principle for a $\mathbf{Z}_+^{\mathbf{N}}$-action on a compact metric space, *Astérisque*, **40**, 147-157 (1976).

[**N**] S. Newhouse, Continuity properties of entropy, *Ann. Math* **129**, 215-235 (1989).

[**O1**] E. Olivier, Analyse multifractale de fonctions continues, *C. R. Acad. Sci. Paris* **326**, 1171-1174 (1998).

[**O2**] E. Olivier, Dimension de Billingsley d'ensembles saturés, *C. R. Acad. Sci. Paris* **328**, 13-16 (1999).

**[Pe]** Ya.B. Pesin, Dimension Theory in Dynamical Systems, Contemporary Views and Applications, Chicago Lectures in Mathematics, University of Chicago Press, Chicago (1997).

**[PePi]** Ya.B. Pesin, S.B. Pitskel´, Topological Pressure and the Variational Principle for Noncompact Sets, *Funct. Anal. and Appl.* **18**, 307-318 (1985).

**[PfS1]** C.-E. Pfister, W.G. Sullivan, Billingsley dimension on shift spaces, *Nonlinearity* **16**, 661-682 (2003).

**[PfS2]** C.-E. Pfister, W.G. Sullivan, Large Deviations Estimates for Dynamical Systems without the Specification Property. Application to the $\beta$-shifts, *Nonlinearity* **18**, 237-261 (2005).

**[S]** P.C. Shields, The Ergodic Theory of Discrete Sample Paths, Graduate Studies in Mathematics Volume **13**, American Mathematical Society(1996).

**[Sc]** J. Schmeling, Symbolic dynamics for $\beta$-shifts and self-normal numbers, *Ergod. Th. & Dynam. Sys.* **17**, 675-694 (1997).

**[TV]** F. Takens, E. Verbitskiy, On the variational principle for the topological entropy of certain non-compact sets *Ergod. Th. & Dynam. Sys* **23**, 317-348 (2003).

**[Te]** A.A. Tempelman, Dimension of Random Fractals in Metric Spaces, *Teor. Veroyatnost. i Primenen.* **44**, 589-616 (1999); translation in *Theory Probab. Appl.* **44**, 537-557 (2000).

**[W]** P. Walters, An Introduction to Ergodic Theory, Graduate Texts in Mathematics **79**, Springer, Berlin (2000).

**[We]** H. Wegmann, Über den Dimensionsbegriff in Wahrsheinlichkeitsräumen von P. Billingsley I and II, *Z. Wahrscheinlichkeitstheorie verw. Geb.* **9**, 216-221 and 222-231 (1968).

**[Y]** L.-S. Young, Some Large Deviation Results for Dynamical Systems, *Trans. Amer. Math. Soc.* **318**, 525-543 (1990).