ISRAEL JOURNAL OF MATHEMATICS 91 (1995), 253–271

# METRICS AND ENTROPY FOR NON-COMPACT SPACES

BY

MICHAEL HANDEL

*Department of Mathematics and Computer Science*
*Lehman College, Bronx, NY 10468, USA*
*e-mail: chflc@cunyvm.cuny.edu*

AND

BRUCE KITCHENS

*Mathematical Sciences Department*
*IBM Watson Research Center, Yorktown Heights, NY 10598, USA*
*e-mail: kitch@watson.ibm.com*

AND

# APPENDIX

BY

DANIEL J. RUDOLPH

*Department of Mathematics*
*University of Maryland, College Park, MD 20742, USA*
*e-mail: djr@math.umd.edu*

ABSTRACT

We investigate Bowen's metric definition of topological entropy for homeomorphisms of non-compact spaces. Different equivalent metrics may assign to the homeomorphism different entropies. We show that the infimum of the metric entropies is greater than or equal to the supremum of the measure theoretic entropies. An example shows that it may be strictly greater. If the entropy of the homeomorphism can vary as the metrics vary we see that the supremum is infinity.

Received August 18, 1993 and in revised form February 15, 1994

## Introduction

Topological entropy is an important tool in the study of dynamics on compact spaces. One reason for this is that topological entropy can be studied from more than one point of view. The original definition of Adler, Konheim, and McAndrew [AKM] used open covers. Bowen gave two definitions; the first [Bo1] used metrics and the second [Bo2] used information theory. A second reason for the usefulness of topological entropy on compact spaces is the variational principal ([Gw],[Di],[Gm]), which states that the topological entropy is equal to the supremum of the measure theoretic entropies over all invariant Borel probability measures. The connection between the measure theoretic and the topological entropy is also illustrated by Katok's ([Kk],[BK]) definition of measure theoretic entropy using the metric methods of Bowen.

For non-compact spaces, there appears to be no such unity. Bowen generalized his metric definition to the non-compact setting in [Bo1]. His definition from the information theory point of view [Bo2] was originally stated in terms of non-compact spaces and has since been greatly expanded by Pesin [Pe] and by Pesin and B.S. Pitskel' [PP]. The open cover definition of Adler, Konheim, and McAndrew can also be generalized. Unfortunately, there are no known relations between these three. One can still define measure theoretic entropy in the same way, although it is now possible that the space $\mathcal{M}(T)$ of all $T$ invariant Borel probability measures on $X$ will be empty. See [O] for a complete discussion of the existence of invariant measures.

Our focus in this paper is on the Bowen metric definition. Let $X$ be a (not necessarily compact) metric space with metric $d$ and let $T$ be a homeomorphism of $X$ to itself. We denote the metric entropy, defined below, by $h_d(T)$. One immediate problem with this invariant is that it depends on the choice of $d$. For example, if $X = \mathbb{R}$ and $T(x) = 2x$, then one would expect (and it is easy to check) that $h_d(T) = \log 2$ with respect to the usual metric on $\mathbb{R}$. But $T$ is conjugate to $T'(x) = cx$ for any positive constant $c$. Using this conjugacy to change the metric, we see that $h_d(T)$ takes on all values in $(0, \infty)$ as $d$ varies among topologically equivalent metrics. To create an intrinsic definition that does not depend on the choice of metric, we consider the extreme values for $h_d(T)$, namely, $h^D(T) = \sup\{h_d(T)\}$ and $h_D(T) = \inf\{h_d(T)\}$. We show in section 2 that $h^D$ is not a useful invariant; if $h_d(T)$ varies as a function of $d$, then $h^D(T) = \infty$.

We denote $\sup\{h_\mu(T): \mu \in \mathcal{M}(T)\}$ by $h_V(T)$. It is not hard to show (Proposition 1.4) that $h_V$ gives a lower bound for $h_D$. One main result (Theorem 1.8) of this paper is that a full variational principal does not hold. Namely, there exists a $\sigma_4$ invariant Borel subset $X \subset \Sigma_4$ (the full 4-shift) such that $h_V(\sigma_4|X) < h_D(\sigma_4|X)$. To prove this we rely on an example of an invariant Borel set $X \subset \Sigma_4$ that has measure zero with respect to any invariant Borel probability measure but for any open set containing $X$ there is an invariant Borel probability measure that gives the open set measure one. The example is due to Dan Rudolph and is explained in the appendix. This result is similar to one of Pesin and Pitskel' [PP], in which they show that the variational principal does not hold for the information theory type entropy and its generalizations. The spaces and the techniques involved in these two examples are quite different.

We conclude this section with Bowen's definition of $h_d(T)$. Let $K \subseteq X$ be a compact subset. A set $E \subseteq K$ is an $(n, \epsilon, K)$ separated set if for every $x, y \in E$ there is an $i$, $0 \le i < n$, so that $d(T^i x, T^i y) > \epsilon$. Let $r_d(n, \epsilon, K)$ be the maximal cardinality of an $(n, \epsilon, K)$ separated set and let

$$h_d(T, K) = \lim_{\epsilon \to 0} \overline{\lim_{n \to \infty}} \frac{1}{n} \log r_d(n, \epsilon, K).$$

Then define the entropy of $T$ with respect to $d$ by $h_d(T) = \sup h_d(T, K)$, where the supremum is over all compact $K \subseteq X$. For a thorough discussion of these ideas see [W].

## 1. The relationship between $h_V$ and $h_D$

The main results of this section are (Proposition 1.4) that $h_V \le h_D$ and that (Theorem 1.8) strict inequality $h_V < h_D$ sometimes occurs. We also present some special cases (Lemmas 1.5, 1.6 and 1.7) in which $h_D = h_V$.

We begin with some standard counting arguments.

Let $\mathcal{C} = \{C_1, \dots, C_k\}$ be any finite collection of disjoint subsets of $X$; $\mathcal{C}$ need not be a cover of $X$. We say that a finite set $E \subset X$ is $(n, \mathcal{C})$- **separated** if for any pair of distinct points $x, y \in K$, there exists $0 \le i < n$ and $1 \le r \ne s \le k$ such that $T^i x \in C_r$ and $T^i y \in C_s$. Define $r(n, \mathcal{C}, K)$ to be the maximal cardinality of a $(n, \mathcal{C})$-separated subset of $K$.

Our first observation will allow us to relate $r(n, \mathcal{C}, K)$ to $h_D$. Note that the hypothesis on $\gamma$ is automatically satisfied if the $C_r$'s are compact.

OBSERVATION 1.1: *Choose a metric $d$ on $X$. If*

$$\gamma = \min\{d(C_r, C_s): 1 \leq r \neq s \leq k\}$$

*is greater than zero, then $r(n, \mathcal{C}, K) \leq r_d(n, \gamma', K)$ for all $\gamma' \leq \gamma$.*

The following two lemmas will allow us to relate $r(n, \mathcal{C}, K)$ to $h_V$.

LEMMA 1.2: *For any invariant measure $\mu$ and compact set $C$ with $\mu(C) > 1-\delta/2$, there is a compact set $K \subseteq C$ with $\mu(K) > 1-\delta$ and an $N$ so that for every $n \geq N$ and $x \in K$*

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_C(T^i x) \geq 1-\delta.$$

*Proof of Lemma 1.2:* For each $N > 0$, let

$$G_N(C) = \left\{x \in X: \frac{1}{n}\sum_{i=0}^{n-1} \chi_C(T^i x) \geq 1-\delta, \quad \forall n \geq N\right\}.$$

It suffices to show that some $K = C \cap G_N(C)$ is compact and satisfies $\mu(K) > 1-\delta$. The first property follows from the fact that each $G_N(C)$ is closed in $X$. For the second property, note that every regular point of $\mu$ is contained in some $G_N(C)$, so that $\mu(\cup G_N(C)) = 1$. Since the $G_N(C)$'s are an increasing family, $\mu(G_N(C)) \geq 1-\delta/2$, and hence $\mu(C \cap G_N(C)) \geq 1-\delta$, for all sufficiently large $N$. ∎

LEMMA 1.3: *Let $\mathcal{P} = \{P_1, \ldots, P_k\}$ be a Borel partition of $X$, let $C_r \subset P_r$ be compact subsets and let $\mathcal{C} = \{C_1, \ldots, C_k\}$. Suppose that $K \subseteq C = \bigcup C_r$ is compact and that there exists $\delta > 0$ so that for all $n \geq N$ and all $x \in K$*

$$\frac{1}{n}\sum_{i=0}^{n-1} \chi_C(T^i x) \geq 1-\delta.$$

*Then*

$$\overline{\lim}\, \frac{1}{n} \log r(n, \mathcal{C}, K) \geq \overline{\lim}\, \frac{1}{n} \log r(n, \mathcal{P}, K) - \epsilon(\delta)$$

*where $\epsilon(\delta) \to 0$ as $\delta \to 0$.*

*Proof of Lemma 1.3:* If $x, y \in K$ and $\{x, y\}$ is not a $(n, \mathcal{C})$-separated set, then we say that $x$ and $y$ are $(n, \mathcal{C})$-**indistinguishable**. In other words, $x$ and $y$ are $(n, \mathcal{C})$-indistinguishable if for each $0 \leq i < n$, either both $T^i(x)$ and $T^i(y)$

are contained in the same element of $\mathcal{C}$ or at least one of the points $T^i(x)$ or $T^i(y)$ is contained in $X \setminus C$. Assuming without loss that $\delta$ is rational, choose $n \geq N$ so that $\delta n$ is an integer. The first step in the proof is to show that the cardinality of a $(n, \mathcal{P})$-separated subset $S \subset K$, all of whose elements are $(n, \mathcal{C})$-indistinguishable, is at most $k^{2\delta n}\binom{n}{2\delta n}$.

For $x \in S$, let $J(x) = \{j\colon 0 \leq j < n; f^j(x) \in X \setminus C\}$. By hypothesis, $J(x)$ has cardinality at most $\delta n$. Choose a collection $J$ of $2\delta n$ integers between 0 and $n-1$. Suppose that $x, y \in S$ and that $J(x) \cup J(y) \subset J$. If $0 \leq i \leq n-1$ is not an element of $J$, then both $T^i(x)$ and $T^i(y)$ are contained in the same element of $\mathcal{C}$ and hence also in the same element of $\mathcal{P}$. Since $\{x, y\}$ is $(n, \mathcal{P})$-separated, there exists $i \in J$ such that $T^i(x)$ and $T^i(y)$ lie in different elements of $\mathcal{P}$. We conclude that there are at most $k^{2\delta n}$ points $x \in S$ with $J(x) \subset J$. Our claim now follows from the fact that there are at most $\binom{n}{2\delta n}$ ways to choose $J$.

We now know that

$$r(n, \mathcal{C}, K) \geq \frac{r(n, \mathcal{P}, K)}{k^{2\delta n}\binom{n}{2\delta n}}.$$

Use Stirling's formula

$$\binom{n}{k} \sim \frac{n^{\frac{3n}{2}}}{2\pi(n-k)^{\frac{3(n-k)}{2}} k^{\frac{3k}{2}}}$$

to asymptotically estimate $\binom{n}{2\delta n}$.

$$\begin{aligned} \overline{\lim} \frac{1}{n} \log r(n, \mathcal{C}, K) &\geq \overline{\lim} \frac{1}{n}\Big( \log r(n, \mathcal{P}, K) - \log k^{2\delta n} - \log \binom{n}{2\delta n}\Big) \\ &= \overline{\lim} \frac{1}{n} \log r(n, \mathcal{P}, K) - 2\delta \log k \\ &\quad + \frac{3}{2}[(1-2\delta)\log(1-2\delta) + 2\delta \log(2\delta)] \end{aligned}$$

All terms involving $\delta$ go to zero as $\delta$ goes to zero. ∎

We can now apply our counting arguments to obtain an inequality relating $h_D$ and $h_V$.

PROPOSITION 1.4: $h_V \leq h_D$.

*Proof of Proposition 1.4:* Fix a metric $d$, an invariant Borel probability measure $\mu$, a finite Borel partition $\mathcal{P} = \{P_1, \ldots, P_k\}$ and $\delta > 0$. Choose compact sets $C_r \subseteq P_r$ so that $C = \bigcup_{r=1}^{k} C_r$ satisfies $\mu(C) > 1 - \delta/2$. Choose $K$ and $N$ as in Lemma 1.2. By the Shannon–McMillan–Breiman theorem

$$r(n, \mathcal{P}, K) \geq (1-2\delta)2^{(h_\mu(T,\mathcal{P})-\delta)n}$$

for all sufficiently large n. This means $\overline{\lim}\,\frac{1}{n}\log r(n,\mathcal{P},K) \geq h_\mu(T,\mathcal{P}) - \delta$. Observation 1.1 and Lemma 1.3 imply that for all sufficiently small $\gamma > 0$

$$\begin{aligned} h_d(T) &\geq \overline{\lim}\,\frac{1}{n}\log r_d(n,\gamma,K) \\ &\geq \overline{\lim}\,\frac{1}{n}\log r(n,\mathcal{C},K) \\ &\geq \overline{\lim}\,\frac{1}{n}\log r(n,\mathcal{P},K) - \epsilon(\delta) \geq h_\mu(T,\mathcal{P}) - (\epsilon(\delta)+\delta). \end{aligned}$$

Since $\epsilon(\delta)+\delta$ can be made arbitrarily small, $h_d(T) \geq h_\mu(T)$ for all metrics $d$ and all invariant measures $\mu$. ∎

We now present some cases in which $h_V = h_D$.

LEMMA 1.5: *If $X$ is a locally compact metric space, then $h_V(T) = h_D(T)$. In fact, there is a metric compactification $X^*$ of $X$ and an extension $T^*$ of $T$ such that $h(T^*) = h_V(T)$.*

*Proof of Lemma 1.5:* Since $X$ is a locally compact metric space, its one point compactification $X^*$ is metric and $T$ extends to a homeomorphism $T^*\colon X^* \to X^*$ that fixes the point at infinity. The only ergodic invariant Borel probability measure on $X^*$ that is not a measure on $X$ is the point mass measure that gives the point at infinity mass one. Since this measure has entropy equal to zero, $h_V(T^*) = h_V(T)$. The variational principal for compact spaces implies that $h_V(T^*) = h_d(T^*)$ for every metric $d$ on $X^*$. Lemma 1.5 now follows from Proposition 1.4 and the fact that $h_{d|X}(T) \leq h_d(T^*)$. ∎

LEMMA 1.6: *If $X$ is a countable state Markov shift and $T$ is the shift transformation, then $h_V(T) = h_D(T)$. In fact, there is a metric compactification $X^*$ of $X$ and an extension $T^*$ of $T$ such that*

$$\sup\{h(\Lambda,T)\colon \textit{compact invariant } \Lambda\} = h_V(T) = h_D(T) = h(T^*).$$

*Remark:* The variational principal for compact spaces implies that $\sup\{h(\Lambda,T)\colon$ compact invariant $\Lambda\} \leq h_V(T)$. In general, equality does not hold. For example, if $T\colon X \to X$ is a positive entropy minimal homeomorphism of a compact metric space, and if $\mathcal{O}(T,x)$ is the orbit of some $x \in X$, then $h_V(T \setminus \mathcal{O}(T,x)) = h_V(T) > 0$ but $(T \setminus \mathcal{O}(T,x))$ contains no compact invariant sets.

*Proof of Lemma 1.6:* This is a result of D. Vere-Jones [V-J] and B.M. Gurevich [G]. A discussion of the dynamics of countable state Markov shifts can be found in

[K]. A countable state Markov chain is defined by a countably infinite, irreducible, 0-1 matrix. Vere-Jones showed that the supremum over the spectral radius of the finite submatrices, $\lambda$, is in many ways analogous to the Perron eigenvalue of a finite, nonnegative, irreducible square matrix. A finite square 0-1 matrix defines a subshift of finite type. It is compact with topological entropy $\log\lambda$ where $\lambda$ is the Perron value. This means $\sup\{h(\Lambda,T)\colon$ compact invariant $\Lambda\subseteq X\}\geq\log\lambda$ where $\lambda$ is the Perron value defined by Vere-Jones. Gurevich defined a compactification with entropy $\log\lambda$ as follows. Let $T$ be a countably infinite 0-1 matrix. It defines a countable state Markov shift, $\Sigma_T$, that is a shift invariant subset of $\{1,2,3,\dots\}^{\mathbb{Z}}$. We think of the one point compactification of the positive integers as $\{1,1/2,1/3,\dots,0\}$, let $\overline{\Sigma}_T$ be the closure of $\Sigma_T$ in $\{1,1/2,\dots,0\}^{\mathbb{Z}}$ and observe that the shift on $\Sigma_T$ extends to the shift on the compact space $\overline{\Sigma}_T$. Gurevich used a coding argument to show the entropy of the shift on $\overline{\Sigma}_T$ is $\log\lambda$. Salama [S] observed that the metric $d(x,y)=\sum_{i=-\infty}^{\infty}|1/x_i-1/y_i|/2^{|i|}$ on $\overline{\Sigma}_T$ can easily be seen to give $h_d(\overline{\Sigma}_T,\sigma)=\log\lambda$. ∎

LEMMA 1.7: *Let $X\subset\Sigma_k$ be the set of nonrecurrent points for the shift $\sigma_k$. Then for all $\epsilon>0$, there is a metric compactification $X^*$ of $X$ and an extension $T^*$ of $T=\sigma_k|X$ such that $h(T^*)<\epsilon$. In particular, $h_D=0$.*

*Proof of Lemma 1.7:* The only facts we use about $X$ is that it is a $\sigma$-compact subset of $\Sigma_k$ with universal measure zero (i.e., $\mu(X)=0$ for all invariant Borel probability measures $\mu$).

Write $X=\bigcup_{i=1}^{\infty}A_i$ where $A_i=\{x\in\Sigma_k\colon d(T^j(x),x)\geq\frac{1}{i}$ for all $j>0\}$. Suppose that a sequence of positive rational $\epsilon_i$'s have been chosen. Since $X$ has universal measure zero and $A_i$ is compact, there exists [O] $m_i$ such that

$$\frac{1}{m_i}\sum_{j=0}^{m_i-1}\chi_{A_i}(T^jx)\leq\epsilon_i$$

for all $x\in X$. Since $A_i$ is the intersection of a decreasing family of open and closed sets in $\Sigma_k$, there is an open and closed set $U_i\supset A_i$ such that

$$\frac{1}{m_i}\sum_{j=0}^{m_i-1}\chi_{U_i}(T^jx)\leq\epsilon_i$$

for all $x\in X$. Inductively define $B_1=U_1$ and $B_i=U_i\setminus U_{i-1}$. Then

(B.1) The $B_i$'s are disjoint open and closed sets in $\Sigma_k$.

(B.2) $X\subset\bigcup_{i=1}^{\infty}B_i$.

Let $\{P_r\colon 1 \le r \le k\}$ be the usual time zero cylinder set partition of $\Sigma_k$. Replacing each $B_i$ by $B_i \cap P_1, \ldots, B_i \cap P_k$, we may also assume that

(B.3) Each $B_i$ is contained in some $P_r$.

Let $Y \subset [0,1]$ be the compact set $\{1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{l} \ldots, 0\}$ and let $Y^{\mathbb{Z}} = \prod_{i=-\infty}^{\infty} Y$. Define $\phi\colon X \to Y^{\mathbb{Z}}$ by $(\phi(x))_i = \frac{1}{l}$ where $T^i(x) \in B_l$. Properties (B1)–(B3) imply that $\phi$ is a well defined embedding. We may therefore view the closure $X^*$ of $\phi(X) \subset Y^{\mathbb{Z}}$ as a compactification of $X$ and the restriction $T^*$ of the shift on $Y^{\mathbb{Z}}$ as an extension of $T$. It suffices to show that for any $\epsilon > 0$, the $\epsilon_i$'s can be chosen so that $h(T^*) < \epsilon$.

Define $\alpha_n\colon Y \to \{0, 1, \ldots, n-1\}$ by

$$\alpha_n\left(\frac{1}{l}\right) = \begin{cases} l & \text{if } 1 \le l \le n-1 \\ 0 & \text{otherwise} \end{cases}$$

and define $\rho_n\colon \{0, 1, \ldots, n\} \to \{0, 1, \ldots, n-1\}$ by

$$\rho_n(l) = \begin{cases} l & \text{if } 1 \le l \le n-1, \\ 0 & \text{if } l = n. \end{cases}$$

Then $\alpha_n$ and $\rho_n$ are continuous and satisfy $\alpha_n = \rho_{n+1}\alpha_{n+1}$. Extend these coordinate wise to continuous maps $A_n\colon Y^{\mathbb{Z}} \to \Sigma_n$ and $P_n\colon \Sigma_{n+1} \to \Sigma_n$. Since $A_n = P_{n+1}A_{n+1}$, the $A_n$'s determine a limit map $F\colon Y^{\mathbb{Z}} \to \varprojlim \Sigma_n$. It is easy to check that $F$ is injective, surjective and continuous and is therefore a homeomorphism.

Define $\phi_n\colon X \to \Sigma_n$ by $\phi_n = A_n\phi$. In other words,

$$(\phi_n(x))_i = \begin{cases} l & \text{if } T^i(x) \in B_l \quad \text{for } l < n, \\ 0 & \text{otherwise.} \end{cases}$$

Denote the closure of $\phi_n(X)$ in $\Sigma_n$ by $X_n^*$ and the restriction of the shift map by $T_n^*\colon X_n^* \to X_n^*$. Then $F$ conjugates $T^*\colon X^* \to X^*$ to $\varprojlim(X_n^*, T_n^*)$. It therefore suffices to show that each $h(T_n^*) < \epsilon$.

We estimate $h(T_n^*|X_n^*)$ as follows. We restrict our attention to those integers $m$ such that $m/m_i$ and $m\epsilon_i$ are integers for $1 \le i \le n$. Since there is a bounded distance between such values of $m$, they are sufficient for computing entropy. The projection of any $\phi_n(x)$ onto $\prod_{i=0}^{m-1}\{0, 1, \ldots, n-1\}$ is a word of length $m$ in the letters $0, 1, \ldots, n-1$ with the property that each letter $i \neq 0$ occurs at most

$m\epsilon_i$ times. The number or such words is bounded by $\prod_{i=1}^{n-1}\binom{m}{m\epsilon_i}$, so Stirlings formula implies that

$$h(T_n^*) \leq \overline{\lim}\,\frac{1}{m}\log\prod_{i=1}^{n-1}\binom{m}{m\epsilon_i}$$
$$\leq \frac{-3}{2}\sum_{i=1}^{n-1}[(1-\epsilon_i)\log(1-\epsilon_i)+\epsilon_i\log\epsilon_i].$$

We need only choose the $\epsilon_i$'s so that $-\frac{3}{2}(1-\epsilon_i)\log(1-\epsilon_i)+\epsilon_i\log\epsilon_i < \epsilon/2^i$. ∎

We now prove that $h_V$ may be strictly less than $h_D$.

Dan Rudolph (see Appendix) constructed a $\sigma_4$-invariant Borel subset $X \subset \Sigma_4$ satisfying:

(X.1) $X$ has universal measure zero (and so $h_V(\sigma_4|X) = 0$).

(X.2) For every neighborhood $U \subseteq \Sigma_4$ of $X$, there is an invariant measure $\mu$ satisfying $\mu(U) = 1$ and $h_\mu(\sigma_4) = \log 2$.

By taking the union of this set with the set of non-recurrent points, we may assume that $X$ also satisfies

(X.3) $X$ contains all non-recurrent points.

THEOREM 1.8: *Let $T = \sigma_4|X\colon X \to X$. For any metric $d$ on $X$, $h_d(T) \geq \log 2$.*

*Proof of Theorem 1.8:* Let $\{P_1, P_2, P_3, P_4\}$ be the usual time zero cylinder set partition of $\Sigma_4$. The $P_i$'s are open and closed and they generate a basis for the topology on $\Sigma_4$. Denote $P_r \cap X$ by $X_r$.

Since each $x \in X_r$ has positive distance from $\bigcup_{s\neq r} X_s$, there are basis elements (i.e. cylinder sets) $\{B_i\}$ and constants $\epsilon_i > 0$ such that

(B.1) $X \subset U = \bigcup_{i=1}^{\infty} B_i$.

(B.2) Each $B_i$ is contained in some $P_r$.

(B.3) $x \in B_i \cap X_r, y \in X_s, r \neq s \Longrightarrow d(x,y) > \epsilon_i$.

Let $\mu$ be an invariant measure associated to $U$ by property $(X.2)$. For any $\delta > 0$, there exists $M$ so that $\mu(\bigcup_{i=1}^{M} B_i) > 1 - \delta/2$. For $1 \leq i \leq 4$, let $C_i = \bigcup B_j$ where the union is taken over $\{B_j\colon 1 \leq j \leq M; B_j \subset P_i\}$. Define $\mathcal{C} = \{C_1, C_2, C_3, C_4\}$, $C = \bigcup_{i=1}^{4} C_i$ and $\gamma = \min\{\epsilon_1, \ldots, \epsilon_M\}$.

By Observation 1.1, it suffices to show that for all $\epsilon > 0$, there is a compact set $K \subset X$ such that $\overline{\lim}\,\frac{1}{n}\log(r(n,\mathcal{C},K)) \geq \log 2 - \epsilon$. As a preliminary step toward the construction of $K$, we construct a compact set $K' \subset C$ with the desired growth rate.

Choose $K'$ and $N$ as in Lemma 1.2 with $\delta < \epsilon/2$ chosen so that the constant $\epsilon(\delta)$ of Lemma 1.3 is less than $\epsilon/2$. Since the fixed point $P = (\dots, 1, 1, 1, \dots)$ is not a regular point for $\mu$, we may assume that $P \notin K'$, and hence that there is a uniform bound to the number of consecutive 1's that occur in an element of $K'$ (this property will be used at the end of the proof). By the Shannon–McMillan–Breiman theorem

$$r(n, \mathcal{P}, K') \geq (1 - 2\delta)2^{(h_\mu(T,\mathcal{P})-\delta)n}$$

for all sufficiently large $n$. This means $\overline{\lim}\, \frac{1}{n} \log r(n, \mathcal{P}, K') \geq h_\mu(T, \mathcal{P}) - \delta$. Lemma 1.3 implies that

$$\begin{aligned} \overline{\lim}\, \frac{1}{n} \log r(n, \mathcal{C}, K) &\geq \overline{\lim}\, \frac{1}{n} \log r(n, \mathcal{P}, K') - \epsilon(\delta) \\ &\geq h_\mu(T, \mathcal{P}) - (\epsilon(\delta) + \delta) \geq \log 2 - \epsilon. \end{aligned}$$

It now suffices to find a compact set $K \subset X$ satisfying $r(n, \mathcal{C}, K) \geq r(n, \mathcal{C}, K')$. Define $r = r_m \colon C \to \Sigma_4$ by

$$r(x) = \begin{cases} 1 & \text{if } i < -m, \\ x_i & \text{otherwise.} \end{cases}$$

Since $C$ is a finite union of cylinder sets, $r(C) \subset C$ for all sufficiently large $m$. We assume that $m$ is chosen so that this property holds. Denote the compact set $r(K') \subset C$ by $K$. Since there is a uniform bound to the number of consecutive 1's that appear in an element of $K'$, each element of $K$ is non-recurrent; by property (X.3), $K \subset X$. Finally note that if $x$ and $y$ are $(n, \mathcal{C})$-separated, then $r(x)$ and $r(y)$ are also $(n, \mathcal{C})$-separated. This implies that $r(n, \mathcal{C}, K) \geq r(n, \mathcal{C}, K')$ as desired. ∎

*Question 1.9:* Perhaps the most natural large set of universal measure zero is the set $NR$ of points that are not regular with respect to any invariant measure. What is $h_D(NR)$? Our proof of Lemma 1.7 does not apply to $NR$ because $NR$ is an $F_{\sigma\delta}$ instead of an $F_\sigma$ (i.e. $NR$ is the intersection of $\sigma$-compact sets but is not itself $\sigma$-compact). The techniques of Theorem 1.8 do not apply to $NR$ because $NR$ has neighborhoods of universal small measure. More precisely, for all $\epsilon > 0$, there is a neighborhood $U \supset NR$ such that $\mu(U) < \epsilon$ for all invariant measures $\mu$. (We thank Dan Rudolph and Mate Wierdl for showing us how this last fact follows from a result of Bourgain [Bg].) We also thank the referee for pointing out to us that this also follows from E. Bishop's constructive ergodic theorem [Bi].

Pesin and Pitskel' [PP] have shown that $h_H(NR) = \log 2$ where $h_H$ is the dimension theory type entropy of [Bo2] and [PP].

## 2. $h^D$

The following proposition is the main result of this section. It shows that $h^D$ is not, in general, an interesting invariant.

PROPOSITION 2.1: *Suppose that $X$ is a separable metric space with no isolated points.*

(1) *If $\overline{\bigcup_{i\geq 0} T^i(K)}$ is compact for every compact $K$, then*

$$\begin{aligned} h^D(T) &= h_V(T) \\ &= \sup\{h(\Lambda, T)\colon \text{ compact invariant } \Lambda \subseteq X\} \\ &= h_D(T). \end{aligned}$$

(2) *If there exists a compact $K$ with $\overline{\bigcup_{i\geq 0} T^i(K)}$ noncompact, then $h^D(T) = +\infty$.*

*Proof of Proposition 2.1:* First suppose that every compact $K$ has a compact forward orbit closure $K^*$. For any metric $d$ on $X$ we have the following relations:

$$h_d(K,T) \leq h_d(K^*,T) = h(T|K^*).$$

Combining this with the variational principal for compact spaces, we see that

$$h^D(T) = \sup\{h(\Lambda,T)\colon \text{ compact invariant } \Lambda \subseteq X\} \leq h_V(T).$$

Part (1) now follows from Proposition 1.4 and the fact that $h_D(T) \leq h^D(T)$.

Next, suppose there is a compact $K$ with a noncompact forward orbit closure. Fix a metric $d$ on $X$. Choose a sequence of points $\{x_i\} \subseteq K$ and an increasing sequence of iterates $n_i \to \infty$ so that the sequence of points $\{T^{n_i}(x_i)\}$ has no convergent subsequences. By adding more points to $K$ if necessary, we may assume that the $x_i$'s are not isolated in $K$. Choose disjoint closed neighborhoods

$U_i = B_d(T^{n_i}(x_i), \delta_i)$ of $T^{n_i}(x_i)$ so that $\bigcup_{i=1}^{\infty} U_i$ is closed. Note that each $x \in X$ has a neighborhood that intersects at most one $U_i$. Define $V_i \subset U_i$ by $V_i = B_d(T^{n_i}(x_i), \delta_i/2)$.

LEMMA 2.2: *For any sequence $\{m_i > 1\}$, there is a metric $d_1$ on $X$ such that*

(1) $d_1 \geq d$,

(2) $d_1 | V_i \times V_i \geq m_i \cdot d | V_i \times V_i$.

*Proof of Lemma 2.2:* We assume without loss that each $\delta_i < 1$. Define

$$\tau(x) = \begin{cases} 2m_i & \text{if } x \in U_i \\ 1 & \text{if } x \notin \bigcup_{i=1}^{\infty} U_i \end{cases}$$

and

$$\rho_1(x, y) = \max\{\tau(x), \tau(y)\} \cdot d(x, y).$$

Finally, define the new metric by

$$d_1(x, y) =$$

$$\inf \left\{ \sum_{i=0}^{k-1} \rho_1(x_i, x_{i+1}) \colon x = x_0, \ldots, x_k = y \text{ any finite collection of points in } X \right\}.$$

It follows immediately from the construction that $d_1$ is a metric and that $d_1 \geq d$. Each $x \in X$ has a neighborhood $W$ that intersects at most one $U_i$. Thus $d|W \times W \leq d_1|W \times W \leq M \cdot d|W \times W$ for some constant $M$ (that is either some $m_i$ or 1). It follows that $d$ and $d_1$ induce the same topology.

To verify (2), suppose that $x_0, \ldots, x_k$ is any collection of points in $X$ with $x_0, x_k \in V_i$. If each $x_i \in U_i$, then $\sum_{i=0}^{k-1} \rho_1(x_i, x_{i+1}) = 2m_i \sum_{i=0}^{k-1} d(x_i, x_{i+1}) \geq 2m_i d(x_0, x_k)$ as desired. Now suppose that $x_l \notin U_i$ for some smallest $l > 0$. Then $\sum_{i=0}^{k-1} \rho_1(x_i, x_{i+1}) > \sum_{i=0}^{l-1} \rho_1(x_i, x_{i+1}) \geq 2m_i \sum_{i=0}^{l-1} d(x_i, x_{i+1}) \geq 2m_i\delta_i/2 = m_i\delta_i \geq m_i d(x_0, x_k)$. ∎

Using Lemma 2.2, we now complete the proof of Proposition 2.1. Let $H$ be a positive integer. Fix $\epsilon > 0$. For each $i \geq 1$, choose $H^{n_i}$ points $\{y_j^i\}$ in $K$ whose $T^{n_i}$ images lie in $V_i$. Let $s_i = \min\{d(T^{n_i}(y_j^i), T^{n_i}(y_k^i)); j \neq k\}$ and choose integers $m_i > \epsilon/s_i$. Apply Lemma 2.2 to produce $d_1$ and note that the points $\{y_j^i\}$ are $(n_i, \epsilon)$-separated with respect to the metric $d_1$. Thus $r_{d_1}(n_i, \epsilon, K) \geq H^{n_i}$ and so $h_{d_1}(T) \geq \log H$. Since $H$ was arbitrary, $h^D = \infty$. ∎

*Remark 2.3:* Suppose that $T^*\colon X^* \to X^*$ is a homeomorphism of a compact metric space and that $T\colon X \to X$ is the restriction of $T^*$ to an invariant subset. If there exists $x \in X$ such that $\omega(T^*, x) \not\subset X$, then $T\colon X \to X$ falls into the second case of Proposition 2.1. For a more general example of case 2, let $X \subseteq S^1$ be the periodic points of the doubling map $T^*\colon S^1 \to S^1$, $T^*(x) = 2x(\bmod 1)$. Choose periodic points $x_i \to 0$ such that the union of the orbits of the $x_i$'s are dense in $S^1$. Then $K = \{x_i, 0\}$ is a compact set whose orbit closure is not compact.

## References


[AKM] R. Adler, A. Konheim and H. MacAndrew, *Topological entropy*, Transactions of the American Mathematical Society **114** (1965), 309–319.

[Bg] J. Bourgain, *Pointwise ergodic theorems for arithmetic sets*, Publications de Mathématiques de l'IHES **69** (1989), 5–45.

[Bi] E. Bishop, *Foundations of Constructive Analysis*, McGraw-Hill, New York, 1989, p. 232.

[Bo1] R. Bowen, *Entropy for group endomorphisms and homogeneous spaces*, Transactions of the American Mathematical Society **153** (1971), 401–414.

[Bo2] R. Bowen, *Topological entropy for noncompact sets*, Transactions of the American Mathematical Society **184** (1973), 125–136.

[BK] M. Brin and A. Katok, Lecture Notes in Mathematics, Vol. 1007, Springer-Verlag, Berlin, 1983, pp. 30–38.

[Di] E. I. Dinaburg, *The relation between topological entropy and metric entropy*, Doklady Akademii Nauk SSSR **190** (1970), 12–22; English transl. in Soviet Mathematics Doklady **11** (1970), 13–16.

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

[Gw] L. W. Goodwyn, *Topological entropy bounds measure theoretic entropy*, Proceedings of the American Mathematical Society **23** (1969), 679–688.

[Gv1] B. M. Gurevič, *Topological entropy of enumerable Markov chains*, Doklady Akademii Nauk SSSR **187** (1969), 715–718; English transl. in Soviet Mathematics Doklady **10** (1969), 911–915.

[Kk] A. Katok, *Lyaponov exponents, entropy and periodic orbits for diffeomorphisms*, Publications de Mathématiques de l'IHES **51** (1980), 137–173.

[Kn] B. Kitchens, *Countable Markov shifts*, preprint.

[O] J. C. Oxtoby, *Ergodic sets,* Bulletin of the American Mathematical Society **58** (1952), 116–136.

[Pe] Ya.B. Pesin, *Dimension type characteristics for invariant sets of dynamical systems,* Uspekhi Matematicheskikh Nauk **43** (1988), 95–128; English transl. in Russian Mathematical Surveys **42** (1988), 111–151.

[PP] Ya. B. Pesin and B. S. Pitskel', *Topological pressure and the variational principle for noncompact sets,* Funktsional'nyi Analiz i ego Prilozheniya **18** (1984), 50–63; English transl. in Functional Analysis and its Applications **18** (1984), 307–318.

[S] I. A. Salama, *Topological entropy and recurrence of countable chains,* Pacific Journal of Mathematics **134** (1988), 325–397.

[V-J] D. Vere-Jones, *Geometric ergodicity in denumerable Markov chains,* The Quarterly Journal of Mathematics. Oxford. **13** (1962), 7–28.

[W] P. Walters, *An Introduction to Ergodic Theory,* GTM, Springer-Verlag, Berlin, 1981.

## Appendix (by Daniel J. Rudolph)

In this appendix we will describe the construction of the example used in Theorem 1.8. In particular we will construct a Borel set $S_0$, in fact a $G_\delta$ in the two shift $\Sigma_2 = \{0,1\}^{\mathbb{Z}}$ with the property that for any open set $O \supseteq S_0$, there is a shift invariant Borel probability measure $\mu$ with $\mu(O) = 1$, but for all shift invariant measures $\mu$, $\mu(S_0) = 0$.

Handel and Kitchens would also like to know that the set is shift invariant and that the entropy $h_\mu(\sigma)$ of the measure with $\mu(O) = 1$ is bounded away from zero, uniformly over the collection of open sets. Our basic example will not satisfy this, as it is much easier to build zero entropy subshifts. We make the set shift invariant by taking the original set's orbit and eliminate the second defect by lifting the example as the first coordinate of a direct product.

We begin with a nonconstructive description of the example. Let $(\Sigma_2, \sigma)$ be the shift map on the full two-shift, and $\mathcal{M}$ be the compact convex set of all $\sigma$ invariant Borel probability measures on $\Sigma$.

The collection of all open sets $\tau$ in $\Sigma_2$ is of cardinality $c$. Let $\mathcal{M}_0 \subseteq \mathcal{M}$ be some collection of nonatomic ergodic measures of cardinality $c$, and

$$\phi\colon \mathcal{M}_0 \to \tau$$

some onto map.

For any $\mu \in \mathcal{M}_0$, either $\mu(\phi(\mu)) = 1$ or is $< 1$. If $< 1$ then $H(\mu) = \{x \notin \phi(\mu)\colon x \text{ is regular for } \mu\}$ is not empty. Let

$$\mathcal{M}_1 = \{\mu \in \mathcal{M}_0\colon H(\mu) \neq \emptyset\}$$

and let

$$f\colon \mathcal{M}_1 \to \Sigma$$

be such that $f(\mu) \in H(\mu)$.

Set

$$S_0 = \text{Range}(f).$$

LEMMA 1: *For $S_0$ as above, for any $\mu \in \mathcal{M}$, $\mu(S_0) = 0$, but for any open set $O$ containing $S_0$, for any $\mu$ with $\phi(\mu) = O$, $\mu(O) = 1$.*

*Proof:* It is enough to show that $\mu(S_0) = 0$ for $\mu$ ergodic. All points in $S_0$ are regular for some ergodic but nonatomic measure $\mu \in \mathcal{M}_1$. Hence for any ergodic $\mu \in \mathcal{M}$, $S_0$ contains at most one regular point for $\mu$, and if one, then $\mu$ is nonatomic. Hence the first part is true.

The second part is obvious as if $S_0 \subseteq O$, then no pullback $\mu$ in $\phi^{-1}(O)$ can be in $\mathcal{M}_1$. ∎

We wish to make the construction outlined above for $S_0$ explicit, i.e. pick $\mathcal{M}_0$ constructively, define $\phi$ constructively and obtain from it a constructive picture of $\mathcal{M}_1$ and then define $f$ constructively so as to show $S_0$ to be a $G_\delta$.

The measures $\mathcal{M}_0$ will all come from nonatomic, minimal and uniquely ergodic subshifts, hence it will only be necessary to describe their supports. This will be done inductively on the levels of a binary tree. At level $k$ we will define $2^k$ disjoint subshifts of finite type $\Sigma_{k,1}, \Sigma_{k,2}, \ldots, \Sigma_{k,2^k} \subset \Sigma_2$. To move to level $k+1$ each $\Sigma_{k,j}$ will be split by taking inside it two smaller disjoint subshifts of finite type, $\Sigma_{k+1,2j-1}$ and $\Sigma_{k+1,2j}$. This construction will be postponed until the end.

The assignment of open sets to elements of $\mathcal{M}_0$ will also take place inductively on the levels of the tree. Let $\{C_1, C_2, \ldots\}$ be some fixed listing of the cylinder sets of $\Sigma_2$. To each support $\Sigma_{k,j}$ in the tree assign a finite union $O_{k,j}$ of cylinders from among $C_1, \ldots, C_k$. In particular set $O_{0,1} = \emptyset$ and

$$O_{k+1,2j-1} = O_{k,j} \quad \text{and} \quad O_{k+1,2j} = O_{k,j} \cup C_{k+1}.$$

Thus as we move down a branch of the tree the subshifts $\Sigma_{k,j(k)}$ nest down to a limit subshift $\Sigma_{\vec{j}}$ and the clopen sets $O_{k,j(k)}$ nest up to an open set $O_{\vec{j}}$. All open sets appear, perhaps many times, among the limits $O_{\vec{j}}$.

Assuming the subshifts $\Sigma_{\vec{j}}$ are minimal and uniquely ergodic, there is a unique $\sigma$ invariant measure $\mu_{\vec{j}}$ supported on $\Sigma_{\vec{j}}$ for which all its points are regular, and we define

$$\phi(\mu_{\vec{j}}) = \phi(\vec{j}) = O_{\vec{j}}.$$

By unique ergodicity

$$H(\mu_{\vec{j}}) = \Sigma_{\vec{j}} \backslash O_{\vec{j}}.$$

Thus

$$\mathcal{M}_1 = \{\mu_{\vec{j}}\colon \Sigma_{\vec{j}} \backslash O_{\vec{j}} \neq \emptyset\}.$$

Imbed $\Sigma_2$ as the classical Cantor set in $[0,1]$ and define, for $\mu_{\vec{j}} \in \mathcal{M}_1$,

$$f(\mu_{\vec{j}}) = f(\vec{j}) = \sup(\Sigma_{\vec{j}} \backslash O_{\vec{j}}).$$

LEMMA 2: *Assuming that the tree of subshifts and clopen sets is as described above,* Range$(f)$ *is a* $G_\delta$.

*Proof:* Each $\Sigma_{k,j}$ is defined by a finite list of allowed cylinders $\mathcal{C}_{k,j}$, whose union is a clopen set $C_{k,j}$, where

$$\Sigma_{k,j} = \bigcap_{i \in \mathbb{Z}} \sigma^i(C_{k,j}).$$

We can assume all the cylinders in $\mathcal{C}_{k,j}$ are on the same indices $[-n(k), \ldots, n(k)]$ where $n(k) \to \infty$ and hence each limit subshift

$$\Sigma_{\vec{j}} = \bigcap_k C_{k,j(k)}.$$

Each set $O_{k,j}$ is clopen, and so can also be assumed to be a union of cylinders on the indices $[-n(k), \ldots, n(k)]$. Hence each set

$$C_{k,j} \backslash O_{k,j}$$

is a finite union of cylinders, and

$$\Sigma_{\vec{j}} \backslash O_{\vec{j}} = \bigcap_k C_{k,j(k)} \backslash O_{k,j(k)}.$$

For each $k, j$, if $C_{k,j} \backslash O_{k,j}$ is not empty, let $S_{k,j}$ be the right-most cylinder in $[0,1]$ on the indices $[-n(k), \ldots, n(k)]$ contained in $C_{k,j} \backslash O_{k,j}$. If $C_{k,j} \backslash O_{k,j} = \emptyset$ then $S_{k,j}$ is not defined. We call $S_{k,j}$ the maximal cylinder of $C_{k,j} \backslash O_{k,j}$ if it exists. If $S_{k,j}$ exists, set $x_{k,j} = \sup S_{k,j}$. There are two possibilities along a branch of the tree $k, j(k)$, either all $x_{k,j}$ exist, in which case they form a decreasing sequence and converge to $f(\vec{j})$, or the sequence terminates at some stage, in which case $\mu_{\vec{j}} \notin \mathcal{M}_1$.

For $k' \geq k$, a maximal cylinder $S_{k',j'}$ will be contained in some cylinder $C \in \mathcal{C}_{k,j}$ where $j$ is the smallest integer $\geq j'/2^{k'-k}$. Let $Q(k,k',j')$ be the union of all cylinders in $\mathcal{C}_{k',j'}$ that are contained in $C$, i.e. all cylinders of $\mathcal{C}_{k',j'}$ which lie in the same cylinder from level $k$ as the maximal cylinder $S_{k',j'}$.

We will show

$$S_0 = \bigcap_{k=0}^{\infty} \bigcup_{\substack{k',j' \\ k-\text{fixed}}} Q(k,k',j').$$

Suppose $x \in S_0$, i.e.

$$x = \sup(\Sigma_{\vec{j}} \backslash O_{\vec{j}}).$$

Then as

$$x_{k',j(k')}, \underset{k'}{\rightarrow} x,$$

for all $k$, $x \in \bigcup_{k-\text{fixed}} Q(k,k',j')$.

For the other direction, suppose $x \in \bigcup_{k-\text{fixed}} Q(k,k',j')$ for all $k$. Then $x \in \Sigma_{\vec{j}}$ for some unique $\vec{j}$. As $x \in Q(k,k',j')$, for some $k' \geq k$, $x \in C_{k'} \in \mathcal{C}_{k',j(k')}$ where $S_{k',j(k')}$ and $C_{k'}$ are both contained in the same cylinder $C \in \mathcal{C}_{k,j(k)}$. Writing $k' = k'(k)$,

$$x = \bigcap_k C_{k'(k)}$$

and

$$\text{dia}(C_{k'(k)} \cup S_{k'(k),j(k'(k))}) \underset{k}{\rightarrow} 0.$$

Hence $x_{k'(k),j(k'(k))} \underset{k}{\rightarrow} x$.

Thus the sequence $x_{k,j(k)}$ does not terminate, and so converges to $f(\vec{j})$ which we see must be $x$. ∎

All that remains now is to describe the construction of the subshifts $\Sigma_{k,j}$. Perhaps the easiest choice is the Chacón family of subshifts (see [R], in particular page 124; [dJ,R,S]). We will give for each $\Sigma_{k,j}$ a pair of finite names $N_1(k,j)$ and $N_2(k,j)$. $\Sigma_{k,j}$ will consist of all doubly infinite words for which any finite subword of length at most half that of $N_1(k,j)$ must actually appear in either $N_1(k,j)$ or $N_2(k,j)$.

To begin

$$N_1(0,1) = 0,$$
$$N_2(0,1) = 1.$$

The next step is

$$N_1(1,1) = 00100010,$$
$$N_2(1,1) = 001010010$$
and
$$N_1(1,2) = 01000100,$$
$$N_2(1,2) = 010010100.$$

Inductively

$$N_1(k+1,2j-1) = N_1(k,j)^2 1\, N_1(k,j) N_1(k,j)^2 1\, N_1(k,j),$$
$$N_2(k+1,2j-1) = N_1(k,j)^2 1\, N_1(k,j)\, 1\, N_1(k,j)^2 1\, N_1(k,j)$$
and
$$N_1(k+1,2j) = N_1(k,j)\, 1\, N_1(k,j)^2 N_1(k,j)\, 1\, N_1(k,j)^2,$$
$$N_2(k+1,2j) = N_1(k,j)\, 1\, N_1(k,j)^2 1\, N_1(k,j)\, 1\, N_1(k,j)^2.$$

It is an observation that each $\Sigma_{k,j}$ so defined is a subshift of finite type. One in fact easily sees they are mixing by noting that $\Sigma_{k,j}$ consists of precisely all infinite concatenations of the block

$$N_1(k-1,(j+1)/2)^2 1\, N_1(k-1,(j+1)/2) \text{ and the symbol } 1$$

if $j$ is even, and the block

$$N_1(k-1,j/2)\, 1\, N_1(k-1,j/2)^2 \text{ and the symbol } 1$$

if $j$ is odd. From this one also sees that the subshifts nest down the branches of the tree and that the limit subshifts are precisely the Chacón family of [R] and hence are minimal, uniquely ergodic and weakly mixing.

This construction does not satisfy all the requirements that Handel and Kitchens require. They would also like to have that the set is shift invariant and that for all $\vec{j}$,

$$h_{\mu_{\vec{j}}}(\sigma) \geq \alpha > 0.$$

This is trivially corrected. Let

$$S = \bigcup_{i=-\infty}^{\infty} \sigma^i(S_0).$$

$S$ is a shift invariant $G_{\delta\sigma}$ and has the same measure properties as $S_0$. All of our measures $\mu_{\vec{j}}$ are of entropy zero, so change to working on the 4-shift $\Sigma_2 \times \Sigma_2 = \Sigma_4$. Set $X = S \times \Sigma_2$. It is a shift invariant $G_{\delta\sigma}$ and has measure zero for all shift invariant measures on $\Sigma_4$ as its projection to its first coordinate does. For any open set $U$ containing $X$, there must be an open set $O$ containing $S$ with $O \times \Sigma_2 \subseteq U$. If $O = O_{\vec{j}}$ then $\mu_{\vec{j}} \times \mu(U) = 1$ for any $\mu \in \mathcal{M}$. In particular if $\mu$ is the Bernoulli $1/2, 1/2$ measure then $h_{\mu_{\vec{j}} \times \mu}(\sigma) = \log 2$. Thus the $X$ in the 4-shift satisfies all their requirements. This completes our work.

Notice that $S_0$ consists completely of regular points for some measure. This is in fact necessary as the work of Bourgain [B] shows that the non-regular points in $\Sigma$ have uniform measure zero.

## References


[B] J. Bourgain, *Pointwise ergodic theorems for Arithmetic sets*, Publications de Mathématiques de l'IHES **69** (1989), 5–45.

[dJ,R,S] A. del Junco, M. Rahe and L. Swanson, *Chacon's automorphism has minimal self-joinings*, Journal d'Analyse Mathématique **37** (1980), 276–284.

[R] D. J. Rudolph, *Fundamentals of Measurable Dynamics*, Oxford University Press, 1990.