Ergod. Th. & Dynam. Sys. (2005), **25**, 1091–1113
doi:10.1017/S0143385704000240

Pre-image entropy

WEN-CHIAO CHENG† and SHELDON E. NEWHOUSE‡

† *Department of Mathematics, National Chung Cheng University, Chia-Yi, Taiwan 621, R.O.C.*
(e-mail: wccheng@math.ccu.edu.tw)

‡ *Department of Mathematics, Michigan State University, E. Lansing, MI 48824, USA*
(e-mail: newhouse@math.msu.edu)

(*Received* 17 *January* 2005)

Abstract. We define and study new invariants called *pre-image entropies* which are similar to the standard notions of topological and measure-theoretic entropies. These new invariants are only non-zero for non-invertible maps, and they give a quantitative measurement of how far a given map is from being invertible. We obtain analogs of many known results for topological and measure-theoretic entropies. In particular, we obtain product rules, power rules, analogs of the Shannon–Breiman–McMillan theorem, and a variational principle.

1. *Introduction*

The notions of entropy are fundamental to our current understanding of dynamical systems. The two main notions are, of course, the topological entropy and the measure-theoretic (or metric) entropy. The former measures the maximal exponential growth rate of orbits for an arbitrary topological dynamical system, and the latter measures the maximal loss of information of the iteration of finite partitions in a measure-preserving transformation. It is well-known that these two invariants are related by the so-called variational principle which states that the topological entropy is the supremum of the metric entropies for all invariant probability measures of a given topological dynamical system. There are many useful properties shared by these two invariants. Using the word 'entropy' to denote either of these concepts, we have:

1. entropy is an invariant in the appropriate category: topologically conjugate systems have the same topological entropy and measure-theoretically conjugate systems have the same metric entropy;
2. the entropy of a direct product is the sum of the entropies of the factors;
3. the entropy of f^n equals n times the entropy of f;
4. if g is a factor of f, then the entropy of g is no larger than the entropy of f;
5. entropy is preserved under the passage to natural extensions.

The last property above shows a weakness of both of these invariants. Namely, they provide no information for non-invertible systems which cannot be gleaned from their associated (invertible) natural extensions.

In several recent papers [**3, 4, 6–8**] some new topological invariants of dynamical systems have been defined and studied which only give non-trivial information when the associated systems are non-invertible. In a certain sense, these new invariants give a quantitative estimate of how 'non-invertible' a system is. For simplicity, if we have a numerical quantity $h(f)$ defined for dynamical systems which is preserved under topological conjugacy and which is zero on invertible systems, we call the quantity a *non-invertible invariant*.

Let us be more precise and recall some non-invertible invariants defined by Hurley.

Let $f : X \to X$ be a continuous self-map of the compact metric space (X, d). Let $n > 0$ be a positive integer. Define the $d_{f,n}$ metric on X by

$$d_{f,n}(x, y) = \max_{0 \le j < n} d(f^j x, f^j y).$$

A set E is an (n, ϵ)-separated set if, for any $x \neq y$ in E, one has $d_{f,n}(x, y) > \epsilon$.

Given a subset $K \subseteq X$, we define the quantity $r(n, \epsilon, K)$ to be the maximal cardinality of (n, ϵ)-separated subsets of K. Uniform continuity of f^j for $0 \le j < n$ guarantees that $r(n, \epsilon, K)$ is finite for each $n, \epsilon > 0$. Following Hurley [**4**], we define the quantities $h_p(f), h_m(f)$ by

$$h_p(f) = \sup_{x \in X} \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log r(n, \epsilon, f^{-n} x),$$

$$h_m(f) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \sup_{x \in X} r(n, \epsilon, f^{-n} x).$$

Of course, the topological entropy, $h_{\text{top}}(f)$ is given by

$$h_{\text{top}}(f) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log r(n, \epsilon, X).$$

It is evident that, in general, one has the inequalities

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

In [**8**], it is proved that for $\alpha = p$ or $\alpha = m$, and any positive integer k we have

$$h_\alpha(f^k) = k h_\alpha(f), \tag{1}$$

$$h_\alpha(f \times g) \le h_\alpha(f) + h_\alpha(g) \tag{2}$$

and

$$\textit{the quantity } h_\alpha(f) \textit{ is preserved under topological conjugacy.} \tag{3}$$

Note that if f is invertible, then $r(n, \epsilon, f^{-n} x) = 1$ for all x, n, ϵ so that $h_p(f) = h_m(f) = 0$. Thus, h_p and h_m are non-invertible invariants.

It is not known if the inequality in (2) can be strict. In a recent paper of Fiebig *et al* [**3**], it is proved that $h_p = h_m = h_{\text{top}}$ for forward expansive maps (in particular for subshifts on finitely many symbols), and examples are constructed in which $h_p(f) < h_m(f)$.

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

It follows from work of Langevin and Przytycki [**6**] and Nitecki and Przytycki [**8**] that $h_m = h_{\text{top}}$ for holomorphic endomorphisms of the Riemann sphere and for continuous endomorphisms of a real interval. Using the sub-additivity of the product above (2), one sees that if $T = f \times g$ with f an arbitrary endomorphism and g an automorphism with positive topological entropy, then, $h_m(T) < h_{\text{top}}(T)$.

We started this research by considering whether there is a variational principle for $h_p(f)$ or $h_m(f)$. This is still unknown. However, we discovered that one can define a new non-invertible invariant $h_{\text{pre}}(f)$ which is between $h_m(f)$ and $h_{\text{top}}(f)$ for which a variational principle does indeed hold. This required a new kind of metric entropy to be defined which takes into account the past behavior of f. The appropriate quantity is simply the metric entropy of f conditioned on the infinite past σ-algebra $\bigcap_n f^{-n}\mathcal{B}$ where $\mathcal{B}$ is the Borel σ-algebra. We call this the *metric pre-image entropy* and denote it by $h_{\text{pre},\mu}(f)$ where μ is a Borel invariant probability measure for f. If one maximizes this quantity over all invariant measures, it is clear that one gets a topological invariant, but it was surprising to us that this can be defined in terms of our quantity $h_{\text{pre}}(f)$. In addition to this variational principle, we will show that the quantities $h_{\text{pre}}(f)$ and $h_{\text{pre},\mu}(f)$ also satisfy power and product rules analogous to the standard topological and metric entropy quantities, that the map $\mu \to h_{\text{pre},\mu}(f)$ is affine, and that there is an analog of the Shannon–Breiman–McMillan theorem for the metric pre-image entropy.

2. *Statement of results*

Given $f : X \to X$ as above, $\epsilon > 0$ and $n \in \mathbb{N}$, let us define

$$h_{\text{pre}}(f) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \sup_{x \in X, k \geq n} r(n, \epsilon, f^{-k}x). \tag{4}$$

It is clear that $h_m(f) \leq h_{\text{pre}}(f) \leq h_{\text{top}}(f)$, and that $h_{\text{pre}}(f) = 0$ if f is a homeomorphism.

Our first results are the following.

THEOREM 2.1. *Let $f : X \to X$ and $g : Y \to Y$ be continuous self-maps of the compact metric spaces X, Y, respectively.*

Then:

1. *the pre-image entropy $h_{\text{pre}}(f)$ is independent of the choice of metric on X;*
2. (power rule) *for any positive integer τ, we have $h_{\text{pre}}(f^\tau) = \tau \cdot h_{\text{pre}}(f)$;*
3. (product rule) $h_{\text{pre}}(f \times g) = h_{\text{pre}}(f) + h_{\text{pre}}(g)$;
4. (topological invariance) *if f is topologically conjugate to g, then $h_{\text{pre}}(f) = h_{\text{pre}}(g)$.*

Remark. We have already mentioned that, for forward expansive maps, Fiebig *et al* showed that $h_p(f) = h_m(f) = h_{\text{top}}(f)$. Hence, also for these maps we have $h_{\text{pre}}(f) = h_{\text{top}}(f)$ and our pre-image entropy is nothing new. On the other hand, analogous to the case of h_m above, our product rule implies that if $T = f \times g$ where f is any endomorphism and g is an automorphism with $h_{\text{top}}(g) > 0$, then $h_{\text{pre}}(T) = h_{\text{pre}}(f) + h_{\text{pre}}(g) = h_{\text{pre}}(f) \leq h_{\text{top}}(f)$ since $h_{\text{pre}}(g) = 0$. But, $h_{\text{top}}(T) = h_{\text{top}}(f) + h_{\text{top}}(g) > h_{\text{top}}(f)$. Thus there are many natural examples of endomorphisms T with $h_{\text{pre}}(T) < h_{\text{top}}(T)$.

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

Next, we consider the analogous notion of metric pre-image entropy. Let $(f, X, \mathcal{B}, \mu)$ denote a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. That is, $(X, \mathcal{B}, \mu)$ is a measure space with $\mu(X) = 1$, and, $X = X_1 \sqcup X_2$ is a disjoint union of measurable sets in which X_2 is a possibly empty, at most countable, set of atoms and $(X_1, \mathcal{B} \mid X_1, \mu \mid X_1)$ is isomorphic (mod 0) to a subinterval of the real closed unit interval with Lebesgue measure and the σ-algebra of Lebesgue measurable sets. The map $f : X \to X$ is such that $f^{-1}(E) \in \mathcal{B}$ for each $E \in \mathcal{B}$, and $\mu(f^{-1}(E)) = \mu(E)$ for each $E \in \mathcal{B}$. Following standard terminology, we call $(f, X, \mathcal{B}, \mu)$ a *measure-preserving transformation*. Set $\mathcal{B}^- = \bigcap_{n\geq 0} f^{-n}\mathcal{B}$. We call $\mathcal{B}^-$ the *infinite past σ-algebra* related to $\mathcal{B}$.

As usual (for example, see Petersen [**9**]), given a subset $A \in \mathcal{B}$, we set $\mu(A \mid \mathcal{B}^-) = E(\chi_A \mid \mathcal{B}^-)$ where χ_A denotes the characteristic function of A, and $E(\psi \mid \mathcal{A})$ denotes the conditional expectation of the function ψ given the sub-σ-algebra $\mathcal{A}$.

For finite partitions α, β, we set $\alpha \vee \beta = \{A \cap B : A \in \alpha, B \in \beta\}$. If $0 \leq j \leq n$, are positive integers, we let $\alpha_j^n = \bigvee_{i=j}^{i=n} f^{-i}\alpha$, and $\alpha^n = \alpha_0^{n-1}$.

Define the information function $I_{\alpha \mid \mathcal{B}^-}$ of α given the infinite past σ-algebra $\mathcal{B}^-$ to be

$$I_{\alpha \mid \mathcal{B}^-} = -\sum_{A\in\alpha} \log \mu(A \mid \mathcal{B}^-)\chi_A.$$

Set

$$H_\mu(\alpha \mid \mathcal{B}^-) = \int I_{\alpha \mid \mathcal{B}^-}\, d\mu = -\sum_{A\in\alpha} \log \mu(A \mid \mathcal{B}^-)\mu(A).$$

It is standard (see, for example, [**9**]) that the quantity $H_\mu(\alpha \mid \mathcal{B}^-)$ is increasing in the first variable and decreasing in the second variable. That is, if β is a partition which refines α and $\mathcal{A}$ is a sub σ-algebra of $\mathcal{B}^-$, then,

$$H_\mu(\alpha \mid \mathcal{B}^-) \leq H_\mu(\beta \mid \mathcal{A}).$$

Also, $H_\mu(f^{-1}\alpha \mid f^{-1}\mathcal{B}^-) = H_\mu(\alpha \mid \mathcal{B}^-)$.

It follows that the numbers $a_n = H_\mu(\alpha^n \mid \mathcal{B}^-)$ form a subadditive sequence. Indeed, for positive integers n, m, using $f^{-n}\mathcal{B}^- = \mathcal{B}^-$, we have

$$\begin{aligned} a_{n+m} &= H_\mu(\alpha^{n+m} \mid \mathcal{B}^-) \\ &= H_\mu((\alpha^n \vee f^{-n}\alpha^m) \mid \mathcal{B}^-) \\ &\leq H_\mu(\alpha^n \mid \mathcal{B}^-) + H_\mu(f^{-n}\alpha^m \mid \mathcal{B}^-) \\ &= H_\mu(\alpha^n \mid \mathcal{B}^-) + H_\mu(f^{-n}\alpha^m \mid f^{-n}\mathcal{B}^-) \\ &= H_\mu(\alpha^n \mid \mathcal{B}^-) + H_\mu(\alpha^m \mid \mathcal{B}^-) \\ &= a_n + a_m. \end{aligned}$$

Consequently, there is a well-defined number $h_\mu(\alpha \mid \mathcal{B}^-)$ given by

$$h_\mu(\alpha \mid \mathcal{B}^-) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\alpha^n \mid \mathcal{B}^-) = \inf_{n\to\infty} \frac{1}{n} H_\mu(\alpha^n \mid \mathcal{B}^-).$$

We define the metric (or measure-theoretic) pre-image entropy of f with respect to μ and $\mathcal{B}$ to be

$$h_{\mathrm{pre},\mu}(f) = \sup_\alpha h_\mu(\alpha \mid \mathcal{B}^-).$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

It is easy to see that $h_{\mathrm{pre},\mu}(f)$ is an invariant of measure-theoretic conjugacy. That is, if $(f, X, \mathcal{B}, \mu)$ and $(g, Y, \mathcal{B}', \nu)$ are measure-preserving transformations, and $\pi : X \to Y$ is a bimeasurable bijection (mod 0) such that $g\pi = \pi f$, then $h_{\mathrm{pre},\mu}(f) = h_{\mathrm{pre},\nu}(g)$. This is the measure-theoretic analog of statement 4 in Theorem 2.1. Next, we have the measure-theoretic analog of statements 2 and 3.

THEOREM 2.2. *Let $(f, X, \mathcal{B}, \mu)$ and $(g, Y, \mathcal{B}', \nu)$ be measure-preserving transformations. Then:*

1. (power rule) *for any positive integer τ, we have*
$$h_{\mathrm{pre},\mu}(f^\tau) = \tau \cdot h_{\mathrm{pre},\mu}(f);$$
2. (product rule)
$$h_{\mathrm{pre},\mu\times\nu}(f \times g) = h_{\mathrm{pre},\mu}(f) + h_{\mathrm{pre},\nu}(g).$$

Our next three results are analogs of well-known theorems concerning metric entropy adapted to the setting of metric pre-image entropy.

THEOREM 2.3. (Affinity of metric pre-image entropy) *Let $(X, \mathcal{B})$ be a measurable space, $f : X \to X$ be a measurable transformation, and let μ and ν be two f-invariant probability measures so that both $(X, \mathcal{B}, \mu)$ and $(X, \mathcal{B}, \nu)$ are Lebesgue spaces together with the possible exception of countable sets of atoms. Let $q \in [0, 1]$.*

Then,
$$h_{\mathrm{pre},q\mu+(1-q)\nu}(\alpha, f) = q h_{\mathrm{pre},\mu}(\alpha, f) + (1-q) h_{\mathrm{pre},\nu}(\alpha, f), \tag{5}$$
and
$$h_{\mathrm{pre},q\mu+(1-q)\nu}(f) = q h_{\mathrm{pre},\mu}(f) + (1-q) h_{\mathrm{pre},\nu}(f). \tag{6}$$

THEOREM 2.4. (Shannon–Breiman–McMillan theorem for metric pre-image entropy) *Let $(f, X, \mathcal{B}, \mu)$ be an ergodic measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$, and let α be a finite measurable partition.*

Then,
$$\lim_{n\to\infty} \frac{1}{n} I_{\alpha^n \mid \mathcal{B}^-} = h_\mu(\alpha \mid \mathcal{B}^-), \tag{7}$$
where the convergence is μ-almost everywhere and in $L^1(\mu)$.

Remark. The previous theorem actually works if $\mathcal{B}^-$ is replaced by any invariant sub-σ-algebra $\mathcal{A}$, i.e. $f^{-1}\mathcal{A} = \mathcal{A}$.

Next, we present the main result of this paper: that the topological and metric pre-image entropies are related by a variational principle.

We consider measure-preserving transformations $(f, X, \mathcal{B}, \mu)$ in which X is a compact metric space, $f : X \to X$ is continuous, and $\mathcal{B}$ is the σ-algebra of Borel subsets of X. As usual in this case we call μ an f-invariant Borel probability measure.

THEOREM 2.5. (Variational principle for pre-image entropy) *Let $f : X \to X$ be a continuous self-map of the compact metric space X, and let $\mathcal{M}(f)$ denote the set of f-invariant Borel probability measures. Then,*
$$h_{\mathrm{pre}}(f) = \sup_{\mu\in\mathcal{M}(f)} h_{\mathrm{pre},\mu}(f). \tag{8}$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

Remarks.

1. Another type of pre-image topological entropy could be defined as

$$h'_{\text{pre}}(f) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \sup_{x \in X, k \geq 1} r(n, \epsilon, f^{-k}x).$$

   It is clear that $h_{\text{pre}}(f) \leq h'_{\text{pre}}(f)$, but it is not obvious whether one might have strict inequality. It turns out that one can follow through the proofs of Theorem 2.5 with this new quantity and show that it also equals $\sup_\mu h_{\text{pre},\mu}(f)$. Hence, it makes no difference whether one uses $k \geq n$ or $k \geq 1$ in the definition of $h_{\text{pre}}(f)$.
2. It is obvious from the variational principle that one can compute $h_{\text{pre}}(f)$ by restricting to any set which contains the supports of the f-invariant probability measures. We do not know if this is true for the quantities $h_p(f)$, $h_m(f)$ of Hurley.
3. It will be clear to experts that many of our methods of proof here are obtained by making appropriate modifications of many known techniques in ergodic theory and topological dynamics. However, the proof of the part of the variational principle which asserts that the pre-image entropy is no larger than the supremum of the metric pre-image entropies of invariant measures is considerably more delicate than the standard result for topological entropy.

3. *Proof of Theorem 2.1*

Before proceeding to the proof, we recall some concepts from entropy theory.

Let f, (X, d), $d_{f,n}$ be as above.

Let $\epsilon > 0$ and $n > 0$. A subset $F \subset X$ is an (n, ϵ, f)-separated subset if whenever $x, y \in F$ and $x \neq y$, we have $d_{f,n}(x, y) > \epsilon$. Given a subset $K \subset X$, we let $r(n, \epsilon, K, f)$ denote the maximal cardinality of an (n, ϵ, f)-separated subset of K. A subset $E \subset K$ is an (n, ϵ, K, f)-spanning set if, for every $x \in K$, there is a $y \in E$ such that $d_{f,n}(x, y) \leq \epsilon$. Let $s(n, \epsilon, K, f)$ be the minimal cardinality of any (n, ϵ, K, f)-spanning set. It is standard that for any subset $K \subset X$,

$$s(n, \epsilon, K, f) \leq r(n, \epsilon, K, f) \leq s\left(n, \frac{\epsilon}{2}, K, f\right). \tag{9}$$

Next, using techniques as in Bowen [**1**], we have the following.

If n_1, n_2, ℓ are positive integers such that $\ell \geq n_1$, then

$$\begin{aligned} r(n_1 + n_2, \epsilon, f^{-\ell}K, f) &\leq s\left(n_1, \frac{\epsilon}{2}, f^{-\ell}K, f\right) s\left(n_2, \frac{\epsilon}{2}, f^{-\ell+n_1}K, f\right) \\ &\leq r\left(n_1, \frac{\epsilon}{2}, f^{-\ell}K, f\right) r\left(n_2, \frac{\epsilon}{2}, f^{-\ell+n_1}K, f\right). \end{aligned} \tag{10}$$

Also, observe that if $m > 0$, $n > 0$, E is an (n, ϵ, f)-separated subset of $f^{-k+m}x$, and $F \subset f^{-m-n}x$ is such that f^m maps F bijectively onto E, then, F is an $(m + n, \epsilon, f)$-separated subset of $f^{-k}x$. Hence,

$$r(n, \epsilon, f^{-k+m}x, f) \leq r(m + n, \epsilon, f^{-k}x, f). \tag{11}$$

It will be convenient to use the notation

$$h_{\text{pre}}(f, \epsilon) = \limsup_{n \to \infty} \frac{1}{n} \log \sup_{k \geq n, x \in X} r(n, \epsilon, f^{-k}x, f)$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

so that

$$h_{\mathrm{pre}}(f) = \lim_{\epsilon \to 0} h_{\mathrm{pre}}(f, \epsilon) = \sup_{\epsilon > 0} h_{\mathrm{pre}}(f, \epsilon).$$

When the metric d needs to be explicitly mentioned, we write

$$h_{\mathrm{pre}}(f, d), h_{\mathrm{pre}}(f, \epsilon, d), r(n, \epsilon, f^{-k}x, f, d),$$

and so on.

Let us prove that $h_{\mathrm{pre}}(f)$ is independent of the metric on X.

Let d_1, d_2 be two metrics on X. Then, by compactness of X, for every $\epsilon > 0$ there is a $\delta > 0$ such that, for all $x, y \in X$, if $d_1(x, y) < \delta$, then $d_2(x, y) < \epsilon$. It follows that $r(n, \epsilon, f^{-k}x, f, d_2) \leq r(n, \delta, f^{-k}x, f, d_1)$ for all n, ϵ, k, x. This gives that $h_{\mathrm{pre}}(f, \epsilon, d_2) \leq h_{\mathrm{pre}}(f, \delta, d_1)$. Letting $\delta \to 0$ gives $h_{\mathrm{pre}}(f, \epsilon, d_2) \leq h_{\mathrm{pre}}(f, d_1)$. Now, letting $\epsilon \to 0$ gives $h_{\mathrm{pre}}(f, d_2) \leq h_{\mathrm{pre}}(f, d_1)$. Interchanging d_1 and d_2 then gives the opposite inequality, proving that $h_{\mathrm{pre}}(f, d_1) = h_{\mathrm{pre}}(f, d_2)$.

Now, let us proceed to prove the power rule. Write $g = f^{\tau}$.

Assertion 1. For any $\epsilon > 0$ we have $h_{\mathrm{pre}}(g, \epsilon) \leq \tau \cdot h_{\mathrm{pre}}(f, \epsilon)$.

Let $k \geq n$ and $x \in X$.

It is clear that

$$r(n, \epsilon, g^{-k}x, g) \leq r(\tau n, \epsilon, f^{-\tau k}x, f).$$

Hence, we have

$$\begin{aligned}
h_{\mathrm{pre}}(g, \epsilon) &= \limsup_{n \to \infty} \frac{1}{n} \log \sup_{k \geq n, x \in X} r(n, \epsilon, g^{-k}x, g) \\
&\leq \limsup_{n \to \infty} \frac{1}{n} \log \sup_{k \geq n, x \in X} r(\tau n, \epsilon, f^{-\tau k}x, f) \\
&= \limsup_{n \to \infty} \frac{\tau}{n\tau} \log \sup_{\tau k \geq \tau n, x \in X} r(\tau n, \epsilon, f^{-\tau k}x, f) \\
&\leq \limsup_{n \to \infty} \frac{\tau}{n\tau} \log \sup_{k \geq \tau n, x \in X} r(\tau n, \epsilon, f^{-k}x, f) \\
&= \tau \limsup_{n \to \infty} \frac{1}{n\tau} \log \sup_{k \geq \tau n, x \in X} r(\tau n, \epsilon, f^{-k}x, f) \\
&\leq \tau \limsup_{n \to \infty} \frac{1}{n} \log \sup_{k \geq n, x \in X} r(n, \epsilon, f^{-k}x, f) \\
&= \tau \cdot h_{\mathrm{pre}}(f, \epsilon),
\end{aligned}$$

proving Assertion 1.

Assertion 2. Given $\epsilon > 0$, let $\delta > 0$ be such that if $d(x, y) < \delta$, then $d(f^j x, f^j y) < \epsilon/4$ for $j \in [0, \tau)$. Then,

$$h_{\mathrm{pre}}(g, \delta) \geq \tau \cdot h_{\mathrm{pre}}(f, \epsilon).$$

Proof. Let $n > 0, k \geq n$. From the definition of δ, we have that if ℓ, s are positive integers such that $\ell \geq \tau s$, then

$$r\left(\tau s, \frac{\epsilon}{4}, f^{-\ell}x, f\right) \leq r(s, \delta, f^{-\ell}x, g). \tag{12}$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

Write $k = \tau n_2 + \ell_2$ with $0 \le \ell_2 < \tau$ and $n - \ell_2 = \tau n_1 + \ell_1$ with $0 \le \ell_1 < \tau$. Let $C(j)$ denote a constant depending on the positive integer j. From (9), (10), (11) and (12), we have

$$\begin{aligned} r(n, \epsilon, f^{-k}x, f) &\le s\left(n - \ell_2, \frac{\epsilon}{2}, f^{-k}x, f\right) s\left(\ell_2, \frac{\epsilon}{2}, f^{-k+n-\ell_2}x, f\right) \\ &\le C(\ell_2) s\left(n - \ell_2, \frac{\epsilon}{2}, f^{-k}x, f\right) \\ &= C(\ell_2) s\left(\tau n_1 + \ell_1, \frac{\epsilon}{2}, f^{-k}x, f\right) \\ &\le C(\ell_2) s\left(\tau n_1, \frac{\epsilon}{4}, f^{-k}x, f\right) s\left(\ell_1, \frac{\epsilon}{4}, f^{-k+\tau n_1}x, f\right) \\ &\le C(\ell_2)C(\ell_1) r\left(\tau n_1, \frac{\epsilon}{4}, f^{-k}x, f\right) \\ &= C(\ell_2)C(\ell_1) r\left(\tau n_1, \frac{\epsilon}{4}, f^{-\tau n_2 - \ell_2}x, f\right) \\ &= C(\ell_2)C(\ell_1) r\left(\tau n_1, \frac{\epsilon}{4}, f^{-\tau(n_2+1)+\tau-\ell_2}x, f\right) \\ &\le C(\ell_2)C(\ell_1) r(n_1, \delta, f^{-\tau(n_2+1)}(f^{\tau-\ell_2}x), g) \\ &= C(\ell_2)C(\ell_1)\, r(n_1, \delta, g^{-(n_2+1)}(f^{\tau-\ell_2}x), g). \end{aligned}$$

Note that $\tau(n_2 + 1) > k \ge n = \tau n_1 + \ell_1 + \ell_2 \ge \tau n_1$, so that $n_2 + 1 > n_1$.

As $n \to \infty$, so does n_1. Hence, using that $(1/n)\log(C(\ell_2)C(\ell_1)) \to 0$ as $n \to \infty$, we get

$$\begin{aligned} h_{\text{pre}}(f, \epsilon) &= \limsup_{n\to\infty} \frac{1}{n} \log \sup_{k\ge n, x\in X} r(n, \epsilon, f^{-k}x, f) \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log \sup_{k\ge n_1, x\in X} C(\ell_2)C(\ell_1) r(n_1, \delta, g^{-k}x, g) \\ &= \limsup_{n\to\infty} \frac{1}{\tau n_1 + \ell_1 + \ell_2} \log \sup_{k\ge n_1, x\in X} r(n_1, \delta, g^{-k}x, g) \\ &= \frac{1}{\tau} \limsup_{n_1\to\infty} \frac{1}{n_1} \log \sup_{k\ge n_1, x\in X} r(n_1, \delta, g^{-k}x, g) \\ &= \frac{1}{\tau} h_{\text{pre}}(g, \delta). \end{aligned}$$

Letting $\epsilon \to 0$ in Assertion 1, we get $h_{\text{pre}}(g) \le \tau \cdot h_{\text{pre}}(f)$. Also, letting $\delta \to 0$ and then taking $\epsilon \to 0$ in Assertion 2, we get $h_{\text{pre}}(g) \ge \tau \cdot h_{\text{pre}}(f)$, completing the proof of the power rule in Theorem 2.1.

Next, we go to the product rule.

Let $(X, d_1), (Y, d_2)$ be compact metric spaces with continuous maps $f : X \to X$, $g : Y \to Y$. Give $X \times Y$ the metric

$$d((x_1, y_1), (x_2, y_2)) = \max(d_1(x_1, x_2), d_2(y_1, y_2)).$$

We have two things to prove:

$$h_{\text{pre}}(f \times g) \le h_{\text{pre}}(f) + h_{\text{pre}}(g), \tag{13}$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

and

$$h_{\text{pre}}(f \times g) \geq h_{\text{pre}}(f) + h_{\text{pre}}(g). \tag{14}$$

Given $(x, y) \in X \times Y$, $n > 0, k \geq n$, let $E \subset (f \times g)^{-k}(x, y)$ be a maximal $(n, \epsilon, f \times g)$-separated set. Let E_1 be a minimal $(n, \epsilon/2, f)$ spanning set in $f^{-k}x$, and let E_2 be a minimal $(n, \epsilon/2, g)$-spanning set in $g^{-k}y$. For each $(u, v) \in E$, there is a pair $(x_1(u), y_1(v)) \in E_1 \times E_2$ such that

$$d_1(f^j u, f^j x_1(u)) \leq \frac{\epsilon}{2} \quad \text{and} \quad d_2(f^j v, f^j y_1(v)) \leq \frac{\epsilon}{2}$$

for $0 \leq j < n$. If $(u_1, v_1), (u_2, v_2)$ are in E and are such that $x_1(u_1) = x_1(u_2)$ and $y_1(v_1) = y_1(v_2)$, then $d_1(f^j u_1, f^j u_2) \leq \epsilon$ and $d_2(g^j v_1, g^j v_2) \leq \epsilon$ for all $0 \leq j < n$. It follows that (u_1, v_1) and (u_2, v_2) remain ϵ close for all interates $(f \times g)^j$ with $0 \leq j < n$. Since E was $(n, \epsilon, f \times g)$-separated, we have $(u_1, v_1) = (u_2, v_2)$. Thus, the map $(u, v) \to (x_1(u), y_1(v))$ is injective. This implies that

$$r(n, \epsilon, (f \times g)^{-k}(x, y), f \times g) \leq r\left(n, \frac{\epsilon}{2}, f^{-k}(x), f\right) \cdot r\left(n, \frac{\epsilon}{2}, g^{-k}(y), g\right).$$

Let

$$a_n(\epsilon) = \sup\{r(n, \epsilon, (f \times g)^{-k}(x, y), f \times g) : (x, y) \in X \times Y,\ k \geq n\},$$
$$b_n(\epsilon) = \sup\{r(n, \epsilon, f^{-k}(x), f) : x \in X,\ k \geq n\},$$

and

$$c_n(\epsilon) = \sup\{r(n, \epsilon, g^{-k}(y), g) : y \in Y,\ k \geq n\}.$$

Then, we have

$$a_n(\epsilon) \leq b_n\left(\frac{\epsilon}{2}\right) \cdot c_n\left(\frac{\epsilon}{2}\right) \tag{15}$$

so that

$$\begin{aligned} h_{\text{pre}}(\epsilon, f \times g) &= \limsup_{n\to\infty} \frac{1}{n} \log a_n(\epsilon) \\ &\leq \limsup_{n\to\infty} \frac{1}{n} \log b_n\left(\frac{\epsilon}{2}\right) + \limsup_{n\to\infty} \frac{1}{n} \log c_n\left(\frac{\epsilon}{2}\right) \\ &= h_{\text{pre}}\left(\frac{\epsilon}{2}, f\right) + h_{\text{pre}}\left(\frac{\epsilon}{2}, g\right). \end{aligned}$$

Letting $\epsilon \to 0$ implies (13).

For the converse statement, it is tempting to try the following.

Let $n > 0, k \geq n, \epsilon > 0$. If E_1 is a maximal (n, ϵ, f)-separated set in $f^{-k}x$, and E_2 is a maximal (n, ϵ, g)-separated set in $g^{-k}y$, then $E_1 \times E_2$ is clearly an $(n, \epsilon, f \times g)$-separated set in $(f \times g)^{-k}(x, y)$.

This gives

$$r(n, \epsilon, f^{-k}(x), f) \cdot r(n, \epsilon, g^{-k}(y), g) \leq r(n, \epsilon, (f \times g)^{-k}(x, y), f \times g),$$

and

$$b_n(\epsilon) \cdot c_n(\epsilon) \leq a_n(\epsilon).$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

If

$$\lim \frac{1}{n} \log a_n(\epsilon), \quad \lim \frac{1}{n} \log b_n(\epsilon) \quad \text{and} \quad \lim \frac{1}{n} \log c_n(\epsilon)$$

all existed, then this would give that $h_{\text{pre}}(\epsilon, f) + h_{\text{pre}}(\epsilon, g) \leq h_{\text{pre}}(\epsilon, f \times g)$, and (14) would follow. In general, these limits do not exist, and this argument breaks down.

Instead, we bring the variational principle and the product rule for metric pre-image entropy to the rescue. Let $\epsilon > 0$, and let μ, ν be invariant measures so that

$$h_{\text{pre},\mu}(f) > h_{\text{pre}}(f) - \epsilon \quad \text{and} \quad h_{\text{pre},\nu}(g) > h_{\text{pre}}(g) - \epsilon.$$

Then, we have

$$\begin{aligned} h_{\text{pre}}(f \times g) &\geq h_{\text{pre},\mu\times\nu}(f \times g) \\ &= h_{\text{pre},\mu}(f) + h_{\text{pre},\nu}(g) \\ &\geq h_{\text{pre}}(f) - \epsilon + h_{\text{pre}}(g) - \epsilon. \end{aligned}$$

Since ϵ is arbitrary, (14) follows.

Finally, we prove the topological invariance statement.

Let $\phi : X \to Y$ be a topological conjugacy from f to g. That is, ϕ is a homeomorphism and $g\phi = \phi f$.

Let d_1 be the metric on X and d_2 be the metric on Y. It is obvious that the induced metric $\phi_\star d_1$ on Y defined by

$$\phi_\star d_1(y_1, y_2) = d_1(\phi^{-1} y_1, \phi^{-1} y_2)$$

gives the same pre-image entropy to g as $h_{\text{pre}}(f)$. Since, the pre-image entropy is independent of the choice of metric on Y, we have $h_{\text{pre}}(f) = h_{\text{pre}}(g)$ as required.

This completes the proof of Theorem 2.1.

4. *Proofs of Theorems 2.2 and 2.3*

Here and in the sequel we will assume without further mention that all sets considered are appropriately measurable.

The power rule. Let $g = f^\tau$ with $\tau > 0$, and consider the two σ-algebras $\mathcal{B}^- = \bigcap_n f^{-n}(\mathcal{B})$, $\mathcal{B}_1^- = \bigcap_n g^{-n}\mathcal{B} = \bigcap_n f^{-n\tau}(\mathcal{B})$.

It is obvious that $\mathcal{B}^- \subset \mathcal{B}_1^-$ since the latter is an intersection of fewer sets. On the other hand, if $\ell \geq k$, then $f^{-\ell}(\mathcal{B}) \subset f^{-k}(\mathcal{B})$. So, for each $n > 0$ we have $\mathcal{B}_1^- \subset f^{-n\tau}(\mathcal{B}) \subset f^{-n}(\mathcal{B})$. Hence, $\mathcal{B}_1^- \subset f^{-n}(\mathcal{B})$ for all $n > 0$. This gives

$$\mathcal{B}^- = \mathcal{B}_1^-. \tag{16}$$

Let α be a finite partition, and let $\beta = \bigvee_{i=0}^{\tau-1} f^{-i}\alpha$. Then,

$$\bigvee_{i=0}^{n-1} g^{-i}(\beta) = \bigvee_{i=0}^{n\tau-1} f^{-i}(\alpha).$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

So,

$$\begin{aligned} h_\mu(\alpha \mid \mathcal{B}^-, f) &= \lim_{n\to\infty} \frac{1}{n\tau} H_\mu\Big(\bigvee_{i=0}^{n\tau-1} f^{-i}\alpha \;\Big|\; \mathcal{B}^- \Big) \\ &= \lim_{n\to\infty} \frac{1}{n\tau} H_\mu\Big(\bigvee_{i=0}^{n-1} g^{-i}\beta \;\Big|\; \mathcal{B}^- \Big) \\ &= \frac{1}{\tau} \lim_{n\to\infty} \frac{1}{n} H_\mu\Big(\bigvee_{i=0}^{n-1} g^{-i}\beta \;\Big|\; \mathcal{B}_1^- \Big) \\ &= \frac{1}{\tau} h_\mu(\beta \mid \mathcal{B}_1^-, g) \\ &\le \frac{1}{\tau} h_{\mathrm{pre},\mu}(g). \end{aligned}$$

Taking the supremum over all α then gives

$$h_{\mathrm{pre},\mu}(f) \le \frac{1}{\tau} h_{\mathrm{pre},\mu}(g).$$

On the other hand, returning to $h_\mu(\alpha \mid \mathcal{B}^-, f) = (1/\tau) h_\mu(\beta \mid \mathcal{B}_1^-, g)$ and using $h_\mu(\beta \mid \mathcal{B}_1^-, g) \ge h_\mu(\alpha \mid \mathcal{B}_1^-, g)$, we get

$$h_\mu(\alpha \mid \mathcal{B}^-, f) \ge \frac{1}{\tau} h_\mu(\alpha \mid \mathcal{B}_1^-, g).$$

This time taking the supremum over all α gives

$$h_{\mathrm{pre},\mu}(f) \ge \frac{1}{\tau} h_{\mathrm{pre},\mu}(g).$$

Product rule. Given a collection γ of sets, let $\mathcal{B}(\gamma)$ denote the σ-algebra generated by γ.

Observe that if $\gamma_1, \gamma_2, \ldots$ is an increasing sequence of finite partitions of X such that $\mathcal{B}\big(\bigcup \gamma_i\big) = \mathcal{B}$ (i.e. the sequence γ_i generates $\mathcal{B}$), then for any finite partition α, we have

$$H_\mu(\alpha \mid \mathcal{B}^-) = \lim_{j\to\infty} \lim_{i\to\infty} H_\mu(\alpha \mid f^{-j}\gamma_i).$$

Let $\mathcal{B}_X$ denote the σ-algebra on X and $\mathcal{B}_Y$ denote that on Y. For $U = X, Y$, let $\mathcal{A}_U$ denote the collection of finite partitions of $\mathcal{B}_U$. Let $\gamma_1 \le \gamma_2 \le \cdots$ and $\delta_1 \le \delta_2 \le \cdots$ be increasing sequences of finite partitions such that $\gamma_i \in \mathcal{A}_X$, $\bigcup_i \gamma_i$ generates $\mathcal{B}_X$, $\delta_i \in \mathcal{A}_Y$, and $\bigcup_i \delta_i$ generates $\mathcal{B}_Y$. Then, for any partitions $\alpha \in \mathcal{A}_X$, $\beta \in \mathcal{A}_Y$ we have

$$\begin{aligned} H_\mu(\alpha \mid \mathcal{B}_X^-) &= \lim_{j\to\infty} \lim_{i\to\infty} H_\mu(\alpha \mid f^{-j}(\gamma_i)), \\ H_\nu(\beta \mid \mathcal{B}_Y^-) &= \lim_{j\to\infty} \lim_{i\to\infty} H_\nu(\alpha \mid g^{-j}(\delta_i)), \end{aligned}$$

and

$$H_{\mu\times\nu}(\alpha\times\beta \mid \mathcal{B}_{X\times Y}^-) = \lim_{j\to\infty} \lim_{i\to\infty} H_{\mu\times\nu}(\alpha\times\beta \mid f^{-j}(\gamma_i)\times g^{-j}(\delta_i)).$$

Here, of course, we define $\alpha\times\beta = \{A\times B : A\in\alpha, B\in\beta\}$.

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

Recall the standard formula

$$H_{\mu_1}(\alpha \mid \beta) = H_{\mu_1}(\alpha \vee \beta) - H_{\mu_1}(\beta) \tag{17}$$

for a measure μ_1 and finite partitions α, β.

Let us use this formula with the notation that A, B, C, D run through $\alpha, \beta, \gamma, \delta$, respectively, to get

$$\begin{aligned} H_{\mu\times\nu}(\alpha \times \beta \mid \gamma \times \delta) &= -\sum_{A,B,C,D} \mu(A \cap C)\nu(B \cap D) \log(\mu(A \cap C)\nu(B \cap D)) \\ &\quad + \sum_{C,D} \mu(C)\nu(D) \log(\mu(C)\nu(D)) \\ &= -\sum_{A,C} \mu(A \cap C) \log(\mu(A \cap C)) - \sum_{B,D} \nu(B \cap D) \log(\nu(B \cap D)) \\ &\quad + \sum_{C} \mu(C) \log \mu(C) + \sum_{D} \nu(D) \log \nu(D) \\ &= H_\mu(\alpha \mid \gamma) + H_\nu(\beta \mid \delta). \end{aligned}$$

Putting the above formulas together gives that for any finite partitions α of X, β of Y, we have

$$H_{\mu\times\nu}(\alpha \times \beta \mid \mathcal{B}^-_{X\times Y}) = H_\mu(\alpha \mid \mathcal{B}^-_X) + H_\nu(\beta \mid \mathcal{B}^-_Y).$$

This implies that, for any positive integer n, we have

$$\begin{aligned} &H_{\mu\times\nu}\Big(\bigvee_{k=0}^{n-1} (f \times g)^{-k}(\alpha \times \beta) \;\Big|\; \mathcal{B}^-_{X\times Y}\Big) \\ &\quad = H_{\mu\times\nu}\Big(\bigvee_{k=0}^{n-1} f^{-k}\alpha \times \bigvee_{k=0}^{n-1} g^{-k}\beta \;\Big|\; \mathcal{B}^-_{X\times Y}\Big) \\ &\quad = H_\mu\Big(\bigvee_{k=0}^{n-1} f^{-k}\alpha \;\Big|\; \mathcal{B}^-_X\Big) + H_\nu\Big(\bigvee_{k=0}^{n-1} g^{-k}\beta \;\Big|\; \mathcal{B}^-_Y\Big). \end{aligned}$$

Multiplying both sides by $1/n$ and taking the limit as $n \to \infty$ gives

$$h_{\mathrm{pre},\mu\times\nu}(\alpha \times \beta, f \times g) = h_{\mathrm{pre},\mu}(\alpha, f) + h_{\mathrm{pre},\nu}(\beta, g)$$

for any $\alpha \in \mathcal{A}_X$, $\beta \in \mathcal{B}_Y$. Since $h_{\mathrm{pre},\mu\times\nu}(f \times g)$ can be computed as the supremum over product partitions $\alpha \times \beta$, this proves the product rule in Theorem 2.2.

Proof of Theorem 2.3. It clearly suffices to prove formula (5).

Let $m = q\mu + (1-q)\nu$.

Recall from [**2**, p. 61] that, for any finite partition α, we have

$$0 \le H_m(\alpha) - qH_\mu(\alpha) - (1-q)H_\nu(\alpha) \le \log 2. \tag{18}$$

Let $\gamma_1 \le \gamma_2 \le \cdots$ be an increasing sequence of finite partitions such that $\bigcup_i \gamma_i$ generates the σ-algebra $\mathcal{B}$.

Then, for any positive integer n and $\rho = m, \nu, \mu$, we have

$$H_\rho(\alpha^n \mid \mathcal{B}^-) = \lim_{j\to\infty} \lim_{i\to\infty} H_\rho(\alpha^n \mid f^{-j}\gamma_i).$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

By (18), we get

$$0 \le H_m(\alpha^n \vee f^{-j}\gamma_i) - qH_\mu(\alpha^n \vee f^{-j}\gamma_i) - (1-q)H_\nu(\alpha^n \vee f^{-j}\gamma_i) \le \log 2, \quad (19)$$

and

$$0 \ge -[H_m(f^{-j}\gamma_i) - qH_\mu(f^{-j}\gamma_i) - (1-q)H_\nu(f^{-j}\gamma_i)] \ge -\log 2. \quad (20)$$

The second term of (20) is non-positive, so adding it to the second term of (19) does not increase the latter's value, so

$$H_m(\alpha^n \mid f^{-j}\gamma_i) - qH_\mu(\alpha^n \mid f^{-j}\gamma_i) - (1-q)H_\nu(\alpha^n \mid f^{-j}\gamma_i) \le \log 2.$$

Similarly, adding the second term of (19) to that of (20) does not decrease the latter's value, so

$$-\log 2 \le H_m(\alpha^n \mid f^{-j}\gamma_i) - qH_\mu(\alpha^n \mid f^{-j}\gamma_i) - (1-q)H_\nu(\alpha^n \mid f^{-j}\gamma_i).$$

Putting these two inequalities together gives

$$-\log 2 \le H_m(\alpha^n \mid f^{-j}\gamma_i) - qH_\mu(\alpha^n \mid f^{-j}\gamma_i) - (1-q)H_\nu(\alpha^n \mid f^{-j}\gamma_i) \le \log 2.$$

Letting $i \to \infty$ and then $j \to \infty$ gives

$$-\log 2 \le H_m(\alpha^n \mid \mathcal{B}^-) - qH_\mu(\alpha^n \mid \mathcal{B}^-) - (1-q)H_\nu(\alpha^n \mid \mathcal{B}^-) \le \log 2.$$

Now, dividing by n and letting $n \to \infty$ gives that

$$h_m(\alpha \mid \mathcal{B}^-) = qh_\mu(\alpha \mid \mathcal{B}^-) + (1-q)h_\nu(\alpha \mid \mathcal{B}^-),$$

as required.

5. *Proof of Theorem 2.4*

The proof of Theorem 2.4 is a more or less straightforward adaptation of the proof of the standard Shannon–Breiman–McMillan theorem in [**9**]. We sketch the ideas.

For a finite partition α, let $\mathcal{B}(\alpha)$ denote the σ-algebra generated by α.

For notational convenience, denote $I_{\alpha\,|\,\mathcal{A}}$ also by $I(\alpha \mid \mathcal{A})$ for a partition α and sub-σ-algebra $\mathcal{A}$.

Let $f_n = I(\alpha \mid \mathcal{B}(\alpha_1^n) \vee \mathcal{B}^-)$ for $n > 0$, and $f_0 = I(\alpha \mid \mathcal{B}^-)$.

First, observe that

$$\begin{aligned} I(\alpha_0^n \mid \mathcal{B}^-) &= I(\alpha_1^n \mid \mathcal{B}^-) + I(\alpha \mid \mathcal{B}(\alpha_1^n) \cup \mathcal{B}^-) \\ &= I(\alpha_0^{n-1} \mid \mathcal{B}^-) \circ f + I(\alpha \mid \mathcal{B}(\alpha_1^n) \cup \mathcal{B}^-) \\ &= \sum_{k=0}^{n} f_{n-k} f^k. \end{aligned} \quad (21)$$

Integrating both sides, and using f^k-invariance of μ gives

$$H_\mu(\alpha_0^n \mid \mathcal{B}^-) = \sum_{k=0}^{n} \int f_{n-k} f^k = \sum_{k=0}^{n} \int f_{n-k} = \sum_{k=0}^{n} \int f_k.$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

This gives the formula

$$H_\mu(\alpha_0^n \mid \mathcal{B}^-) = H_\mu(\alpha \mid \mathcal{B}^-) + \sum_{k=1}^{n} H_\mu(\alpha \mid \mathcal{B}(\alpha_1^k) \cup \mathcal{B}^-). \tag{22}$$

Since the sequence $H_\mu(\alpha \mid \mathcal{B}(\alpha_1^k) \cup \mathcal{B}^-)$ converges non-increasingly to some number h_0, the associated Cesaro sequence

$$\frac{1}{n} \sum_{k=1}^{n} H_\mu(\alpha \mid \mathcal{B}(\alpha_1^k) \cup \mathcal{B}^-)$$

also converges to h_0 as $n \to \infty$. However, multiplying both sides of (22) by $1/(n+1)$ and taking the limit, we then get

$$h_0 = \lim_{n\to\infty} \frac{1}{n+1} H_\mu(\alpha_0^n \mid \mathcal{B}^-) = h_\mu(\alpha \mid \mathcal{B}^-).$$

Now, the proof of the Shannon–Breiman–McMillan theorem in [**9**] gives that the sequence $\{I(\alpha \mid \alpha_1^k) : k \geq 1\}$ is an L^1-bounded martingale. Since

$$\begin{aligned} \int I(\alpha \mid \mathcal{B}(\alpha_1^k) \cup \mathcal{B}^-)\, d\mu &= H_\mu(\alpha \mid \mathcal{B}(\alpha_1^k) \cup \mathcal{B}^-) \\ &\leq H_\mu(\alpha \mid \mathcal{B}(\alpha_1^k)) \\ &= \int I(\alpha \mid \mathcal{B}(\alpha_1^k))\, d\mu, \end{aligned}$$

we have that $\{I(\alpha \mid \mathcal{B}(\alpha_1^k) \cup \mathcal{B}^-) : k \geq 1\}$ is also an L^1-bounded martingale.

Hence, the last sequence converges almost everywhere and in L^1 to some integrable function h. Its Cesaro sums must also converge in both senses to h as well. Note, also, that $\int h\, d\mu = h_0$.

Again, following the method in [**9**], we have

$$\begin{aligned} \lim_{n\to\infty} \frac{I(\alpha_0^n \mid \mathcal{B}^-)}{n+1} &= \lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^{n} f_{n-k} f^k \\ &= \frac{1}{n+1} \sum_{k=0}^{n} (f_{n-k} - h) \circ f^k + \frac{1}{n+1} \sum_{k=0}^{n} h \circ f^k. \end{aligned}$$

Since $f_n \to h$, the first term in the last equality approaches 0 and the ergodicity of the μ gives that the second term approaches $\int h\, d\mu = h_0$.

Hence,

$$\lim_{n\to\infty} \frac{I(\alpha^n \mid \mathcal{B}^-)}{n} = \lim_{n\to\infty} \frac{I(\alpha_0^n \mid \mathcal{B}^-)}{n+1} = h_0,$$

completing the proof of Theorem 2.4.

6. *Proof of Theorem 2.5*

We adapt the well-known arguments of Misiurewicz (as in [**9**, pp. 269–271]) to pre-image entropy.

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

Step 1. $h_{\mathrm{pre},\mu}(f) \le h_{\mathrm{pre}}(f)$ for all $\mu \in \mathcal{M}(f)$.

Let $\mu \in \mathcal{M}(f)$. If $E_1, E_2, \ldots, E_s$ is a finite disjoint collection of compact subsets of X, we call the partition

$$\alpha = \left\{ E_1, E_2, \ldots, E_s, X \setminus \bigcup_{i=1}^{s} E_i \right\}$$

a *compact* partition. It is easy to verify that $h_{\mathrm{pre},\mu}(f)$ is the supremum of $h_{\mathrm{pre},\mu}(\alpha, f)$ where α varies over the compact partitions.

Thus, using a standard technique, it suffices to prove that, for any compact partition α,

$$h_{\mathrm{pre},\mu}(\alpha, f) \le h_{\mathrm{pre}}(f) + \log 2. \tag{23}$$

Indeed, once this is done, it follows that

$$h_{\mathrm{pre},\mu}(f) \le h_{\mathrm{pre}}(f) + \log 2$$

for every f. Applying this to f^N for large N, gives

$$N h_{\mathrm{pre},\mu}(f) = h_{\mathrm{pre},\mu}(f^N) \le h_{\mathrm{pre}}(f^N) + \log 2,$$

or

$$h_{\mathrm{pre},\mu}(f) \le h_{\mathrm{pre}}(f) + \frac{\log 2}{N}.$$

Letting $N \to \infty$ then, gives $h_{\mathrm{pre},\mu}(f) \le h_{\mathrm{pre}}(f)$.

For (23), it suffices to show that there is an $\epsilon > 0$ such that for any $n > 0$ and any $k > 0$, we have

$$H_\mu(\alpha^n \mid f^{-k}\mathcal{B}) \le n \log 2 + \log \sup_{x \in X} r(n, \epsilon, f^{-k}x). \tag{24}$$

Let ϵ be such that any 4ϵ-ball meets at most two elements of α.

Let $\beta_1 \le \beta_2 \le \cdots$ be a non-decreasing sequence of finite partitions with diameters tending to zero. Thus,

$$\mathcal{B} = \bigvee_{j=1}^{\infty} \beta_j,$$

and

$$H_\mu(\alpha^n \mid f^{-k}\mathcal{B}) = \lim_{j \to \infty} H_\mu(\alpha^n \mid f^{-k}\beta_j).$$

So, it suffices to show that, for sufficiently large j, we have

$$H_\mu(\alpha^n \mid f^{-k}\beta_j) \le n \log 2 + \log \sup_{x \in X} r(n, \epsilon, f^{-k}x). \tag{25}$$

Let $\epsilon_1 = \epsilon_1(n, \epsilon) > 0$ be small enough so that if $d(x, y) < \epsilon_1$, then $d(f^i x, f^i y) < \epsilon$ for $0 \le i < n$.

The collection $\{f^{-k}x : x \in f^k X\}$ is an upper-semicontinuous decomposition of X. Hence, for each $x \in f^k X$ there is an $\epsilon_2(x, k, \epsilon_1) > 0$ such that if $d(y, x) < \epsilon_2(x, k, \epsilon_1)$, $y \in f^k X$ and $y_1 \in f^{-k}y$, then there is an $x_1 \in f^{-k}x$ such that

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

$d(y_1, x_1) < \epsilon_1$. Let $\mathcal{U}$ be the collection of open $\epsilon_2(x, k, \epsilon_1)$ balls in $f^k X$ as x varies in $f^k X$, and let ϵ_3 be a Lebesgue number for $\mathcal{U}$.

Since $\operatorname{diam}(\beta_j) \to 0$ as $j \to \infty$, we may choose j_0 large enough so that if $j \geq j_0$ and B is an element of β_j, then $\operatorname{diam}(B) < \epsilon_3$.

Assume $j \geq j_0$.

For a set $C \in f^{-k}\beta_j$, let μ_C denote the conditional measure of μ restricted to C. Let a_C^n denote the set $\{A \cap C : A \in \alpha^n, A \cap C \neq \emptyset\}$.

We have

$$\begin{aligned} H_\mu(\alpha^n \mid f^{-k}\beta_j) &= \sum_{C \in f^{-k}\beta_j} H_{\mu_C}(\alpha_C^n)\mu(C) \\ &\leq \max_{C \in f^{-k}\beta_j} H_{\mu_C}(\alpha_C^n) \\ &\leq \max_{C \in f^{-k}\beta_j} \log \operatorname{card} \alpha_C^n. \end{aligned}$$

Fix a $B \in \beta_j$ so that $C = f^{-k}B$ is non-empty. For each $A \in \alpha_C^n$, let x_A be an element of A.

Since $f^k x_A \in B$ and $\operatorname{diam}(B) < \epsilon_3$, there is a $u_B \in f^k X$ such that if $y \in B \cap f^k X$, then $d(u_B, y) < \epsilon_2(u_B, k, \epsilon_1)$. In particular,

$$d(f^k x_A, u_B) < \epsilon_2(u_B, k, \epsilon_1).$$

Hence, there is a point $\phi_1(A) \in f^{-k}u_B$ such that $d(x_A, \phi_1(A)) < \epsilon_1$, so

$$d(f^i x_A, f^i \phi_1(A)) < \epsilon$$

for $0 \leq i < n$.

Let E_C be a maximal (n, ϵ)-separated set in $f^{-k}u_B$.

Since E_C spans $f^{-k}u_B$, there is a point $\phi_2(A) \in E_C$ such that, for $0 \leq i < n$,

$$d(f^i \phi_1(A), f^i \phi_2(A)) \leq \epsilon,$$

and, hence,

$$d(f^i x_A, f^i \phi_2(A)) \leq 2\epsilon.$$

We have defined a map ϕ_2 from α_C^n into E_C. Hence,

$$\operatorname{card}(\alpha_C^n) \leq \operatorname{card} E_C \cdot \Big(\max_{y \in E_C} \phi_2^{-1}(y) \Big).$$

To prove (25), we will show

$$\operatorname{card}(\phi_2^{-1}(y)) \leq 2^n \tag{26}$$

for each $y \in E_C$.

Let $A, \tilde{A}$ be such that $\phi_2(x_A) = \phi_2(\tilde{A})$.

Then, for $0 \leq i < n$, we have

$$d(f^i x_A, f^i x_{\tilde{A}}) \leq 4\epsilon.$$

It follows that $f^i x_A$ is in the 4ϵ ball about $f^i x_{\tilde{A}}$. Since each 4ϵ ball meets at most two elements of α, it follows that there are at most 2^n choices for $A, \tilde{A}$. This proves (25) and finishes the proof of Step 1.

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

Step 2. We have

$$h_{\mathrm{pre}}(f) \leq \sup_{\mu \in \mathcal{M}(f)} h_{\mathrm{pre},\mu}(f). \tag{27}$$

Given $\epsilon > 0$, we wish to produce an f-invariant probability measure μ such that

$$h_{\mathrm{pre},\mu} \geq h_{\mathrm{pre}}(f, \epsilon). \tag{28}$$

Again we want to follow the idea of the Misiurewicz proof of the standard variational principle (as in [**9**, pp. 269–271]), but we must make important modifications. Although it is not strictly necessary, the reader will find it helpful in what follows to be familiar with the proof in the aforementioned reference. We will motivate our construction by making frequent reference to that proof. When it is convenient to refer to the arguments in [**9**], we simply call them the *SVP* arguments.

Choose sequences $n_i \to \infty, k_i > n_i, x_i \in X$, such that

$$h_{\mathrm{pre}}(f, \epsilon) = \lim_{i \to \infty} \frac{1}{n_i} \log r(n_i, \epsilon, f^{-k_i} x_i, f).$$

Let E_i denote a maximal (n_i, ϵ)-separated set in $f^{-k_i} x_i$ such that $\operatorname{card} E_i = r(n_i, \epsilon, f^{-k_i} x_i, f)$. Thus, we have

$$h_{\mathrm{pre}}(f, \epsilon) = \lim_{i \to \infty} \frac{1}{n_i} \log \operatorname{card} E_i. \tag{29}$$

Letting δ_x denote the point mass at point $x \in X$, let

$$\mu_i = \frac{1}{\operatorname{card} E_i} \sum_{x \in E_i} \delta_x,$$

let

$$\nu_i = \frac{1}{n_i} \sum_{j=0}^{n_i - 1} f_\star^j \mu_i,$$

and let

$$\nu = \lim_{i \to \infty} \nu_i,$$

where we have passed to subsequences assuring that the last limit of measures exists.

Let α be a finite ν-continuity partition such that $\operatorname{diam}(\alpha) < \epsilon$. That is, for each $A \in \alpha$, $\mu(\partial A) = 0$ and $\operatorname{diam}(A) < \epsilon$.

Motivated by the SVP arguments in [**9**], we wish to show that, for every positive integer m,

$$H_\nu(\alpha^m \mid \mathcal{B}^-) \geq m h_{\mathrm{pre}}(f, \epsilon). \tag{30}$$

Indeed, from this, we have

$$h_{\mathrm{pre},\nu}(f) \geq h_{\mathrm{pre},\nu}(\alpha, f) = \inf_m \frac{H_\nu(\alpha^m \mid \mathcal{B}^-)}{m} \geq h_{\mathrm{pre}}(f, \epsilon),$$

as required.

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

If we were in the SVP case and not conditioning on $\mathcal{B}^-$, the result that $H_\nu(\alpha^m) \geq mh(f,\epsilon)$ follows from the following two facts:

$$H_\nu(\alpha^m) = \lim_{i\to\infty} H_{\nu_i}(\alpha^m) \tag{31}$$

and

$$H_{\nu_i}(\alpha^m) \geq \frac{1}{n_i}\sum_{\ell=0}^{n_i-1} H_{f_\star^\ell \mu_i}(\alpha^m) = \frac{1}{n_i}\sum_{\ell=0}^{n_i-1} H_{\mu_i}(f^{-\ell}\alpha^m). \tag{32}$$

Statement (31) follows from the fact that the map $\mu \to \mu(A)$ is continuous at ν, and statement (32) follows from the concavity of the function $t \to -t\log t$ arising in the entropy formula

$$H_{\nu_i}(\alpha^m) = \sum_{A\in\alpha^m} -\nu_i(A)\log\nu_i(A). \tag{33}$$

We need to establish the analogs of (31) and (32) after conditioning on $\mathcal{B}^-$. To do this we will make use of the Rokhlin theory of measurable partitions (see [**5**] for definitions and basic facts about this theory). This theory concerns itself with complete measure spaces, so we will actually deal with a slightly larger σ-algebra $\mathcal{B}^-_{\mathcal{C}}$ in which ν is complete.

Let $\mathcal{C}$ denote the subcollection of $\mathcal{B}^-$ consisting of ν-null sets. That is, $E \in \mathcal{C}$ if and only if $E \in \mathcal{B}^-$ and $\nu(E) = 0$. Since the countable union of elements in $\mathcal{C}$ is again in $\mathcal{C}$, it follows that for any σ-algebra $\mathcal{A}$ of subsets of X, there is an enlarged σ-algebra $\mathcal{A}_{\mathcal{C}}$ defined by $A \in \mathcal{A}_{\mathcal{C}}$ if and only if there are sets B, M, N such that $A = B \cup M$, $B \in \mathcal{A}$, $N \in \mathcal{C}$ and $M \subseteq N$. The σ-algebra $\mathcal{B}^-_{\mathcal{C}}$ is simply the standard ν-completion of $\mathcal{B}^-$. We will also consider the σ-algebras $\mathcal{B}^k = (f^{-k}\mathcal{B})_{\mathcal{C}}$ for $k \geq 1$.

Since $f^{-1}(\mathcal{C}) \subseteq \mathcal{C}$, we have that

$$\mathcal{B}^1 \supseteq \mathcal{B}^2 \supseteq \cdots.$$

Letting $\mathcal{B}^\infty = \bigcap_{k\geq 1}\mathcal{B}^k$, it is easy to check that

$$\mathcal{B}^- \subset \mathcal{B}^-_{\mathcal{C}} \subset \mathcal{B}^\infty,$$

and

$$f^{-\ell}\mathcal{B}^k \subset \mathcal{B}^{\ell+k} \quad \text{for all } \ell \geq 1.$$

Note that the σ-algebras $\mathcal{B}^k$, $\mathcal{B}^\infty$ are not necessarily ν-complete.

LEMMA 6.1. *Under the above notation, we have*

$$H_\nu(\alpha^m \mid \mathcal{B}^-) \geq \limsup_{i\to\infty} H_{\nu_i}(\alpha^m \mid \mathcal{B}^\infty), \tag{34}$$

and

$$H_{\nu_i}(\alpha^m \mid \mathcal{B}^{k_i}) \geq \frac{1}{n_i}\sum_{\ell=0}^{n_i-1} H_{f_\star^\ell \mu_i}(\alpha^m \mid \mathcal{B}^{k_i}). \tag{35}$$

Remark. In the course of the proof of the lemma, we will see that while the proof of estimate (34) is rather straightforward, that of (35) is considerably more complicated. It is not true that

$$H_{\nu_i}(\alpha_0^{m-1} \mid \mathcal{A}) \geq \frac{1}{n_i}\sum_{\ell=0}^{n_i-1} H_{f_\star^\ell \mu_i}(\alpha_0^{m-1} \mid \mathcal{A}) \tag{36}$$

for all average measures ν_i and all invariant sub-σ-algebras $\mathcal{A}$. Indeed, statement (35) holds because the measure ν_i is closely related to the measurable partition $\{f^{-k_i}x : x \in X\}$.

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

Proof of (30) assuming Lemma 6.1. Since $\mathcal{B}^- \subseteq \mathcal{B}^\infty$, we have that

$$H_\nu(\alpha^m \mid \mathcal{B}^-) \geq H_\nu(\alpha^m \mid \mathcal{B}^\infty). \tag{37}$$

Further, since $\mathcal{B}^\infty$ is the decreasing limit of the subalgebras $\mathcal{B}^k$, we clearly have

$$H_\nu(\alpha^m \mid \mathcal{B}^\infty) \geq H_\nu(\alpha^m \mid \mathcal{B}^{k_i}),$$

for all i.

Also, since μ_i is supported on $f^{-k_i}x_i$, the canonical system of conditional measures induced by μ_i on the measurable partition $\{f^{-k_i}x : x \in X\}$ reduces to a single measure on the set $f^{-k_i}x_i$ which we may identify with μ_i.

Now, each element $A \in \mathcal{B}^{k_i}$ can be expressed as the disjoint union $A = B \cup C$ with

$$B \in f^{-k_i}\mathcal{B} = \{f^{-k_i}x : x \in X\},$$

and $C \in \mathcal{C}$. Since μ_i is supported on elements of $f^{-k_i}\mathcal{B}$, we have $\mu_i(C) = 0$. Hence, for any finite partition γ, we have

$$H_{\mu_i}(\gamma \mid \mathcal{B}^{k_i}) = H_{\mu_i}(\gamma \mid f^{-k_i}x_i).$$

Since each element of $\alpha^m \mid f^{-k_i}x_i$ has at most one point, we have

$$H_{\mu_i}(\alpha^m \mid f^{-k_i}x_i) = \log \operatorname{card} E_i.$$

In the sequel, for non-negative integers $a \leq b$, let us denote by $[a, b)$ the set of integers j such that $a \leq j < b$.

For $0 \leq j < m$, let

$$C(i, j) = \{n \in [0, n_i) : n \equiv j \bmod (\mathrm{m})\},$$

and let

$$\rho(i) = \max\ C(i, j).$$

We use the standard notation

$$\alpha_s^t = \bigvee_{j=s}^{j=t} f^{-j}\alpha, \quad \alpha^s = \alpha_0^{s-1}.$$

Since we can express $[0, n_i)$ as the disjoint union

$$[0, n_i) = \bigsqcup_{0 \leq j < m} C(i, j), \tag{38}$$

and, for each fixed $0 \leq j < m$, we can write

$$\alpha^{n_i-1} = \alpha^j \vee \left(\bigvee_{t \in C(i,j)} f^{-t}\alpha^m\right) \vee \alpha_{n_i-\rho(i)}^{n_i-1}, \tag{39}$$

we get

$$\begin{aligned}\log \operatorname{card} E_i &= H_{\mu_i}(\alpha^{n_i-1} \mid f^{-k_i}x_i) \\ &\leq H_{\mu_i}(\alpha^j \mid f^{-k_i}x_i) + H_{\mu_i}\left(\bigvee_{t \in C(i,j)} f^{-t}\alpha^m \;\middle|\; f^{-k_i}x_i\right) \\ &\quad + H_{\mu_i}(\alpha_{n_i-\rho(i)}^{n_i-1} \mid f^{-k_i}x_i).\end{aligned}$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

Thus,

$$H_{\mu_i}\Big(\bigvee_{t\in C(i,j)} f^{-t}\alpha^m \;\Big|\; f^{-k_i}x_i\Big) \geq \log \operatorname{card} E_i - 2m \log \operatorname{card} \alpha. \tag{40}$$

Using (35) we have

$$\begin{aligned}
H_{\nu_i}(\alpha^m \mid \mathcal{B}^\infty) &\geq H_{\nu_i}(\alpha^m \mid \mathcal{B}^{k_i}) \\
&\geq H_{(1/n_i)\sum_{\ell=0}^{n_i-1} f_\star^\ell \mu_i}(\alpha^m \mid \mathcal{B}^{k_i}) \\
&\geq \frac{1}{n_i}\sum_{\ell=0}^{n_i-1} H_{f_\star^\ell \mu_i}(\alpha^m \mid \mathcal{B}^{k_i}) \\
&= \frac{1}{n_i}\sum_{\ell=0}^{n_i-1} H_{\mu_i}(f^{-\ell}\alpha^m \mid f^{-\ell}\mathcal{B}^{k_i}) \\
&\geq \frac{1}{n_i}\sum_{\ell=0}^{n_i-1} H_{\mu_i}(f^{-\ell}\alpha^m \mid \mathcal{B}^{k_i}) \\
&\geq \frac{1}{n_i}\sum_{0\leq j<m}\sum_{\ell\in C(i,j)} H_{\mu_i}(f^{-\ell}\alpha^m \mid \mathcal{B}^{k_i}) \\
&\geq \frac{m}{n_i}\inf_{0\leq j<m} H_{\mu_i}\Big(\bigvee_{\ell\in C(i,j)} f^{-\ell}\alpha^m \;\Big|\; \mathcal{B}^{k_i}\Big) \\
&= \frac{m}{n_i}\inf_{0\leq j<m} H_{\mu_i}\Big(\bigvee_{\ell\in C(i,j)} f^{-\ell}\alpha^m \;\Big|\; f^{-k_i}x_i\Big) \\
&\geq \frac{m}{n_i}[\log \operatorname{card} E_i - 2m \log \operatorname{card} \alpha].
\end{aligned}$$

From these inequalities, we get that

$$\begin{aligned}
\limsup_{i\to\infty} H_{\nu_i}(\alpha^m \mid \mathcal{B}^\infty) &\geq \limsup_{i\to\infty} \frac{m}{n_i}[\log \operatorname{card} E_i - 2m \log \operatorname{card} \alpha] \\
&= \lim_{i\to\infty} \frac{m}{n_i}[\log \operatorname{card} E_i - 2m \log \operatorname{card} \alpha] \\
&= m h_{\mathrm{pre}}(f, \epsilon).
\end{aligned}$$

This and (34) proves (30).

Proof of Lemma 6.1. The triple $(X, \mathcal{B}_\mathcal{C}^-, \nu)$ is a Lebesgue space together with a (possibly empty), at most countable, set of atoms. Hence, we can find an increasing sequence $\beta_1 \leq \beta_2 \leq \cdots$ of finite partitions consisting of elements of $\mathcal{B}_\mathcal{C}^-$ such that

$$\mathcal{B}_\mathcal{C}^- = \bigvee_{j=1}^{\infty} \beta_j$$

and each β_j is a ν-continuity partition.

Then,

$$H_\nu(\alpha^m \mid \mathcal{B}_\mathcal{C}^-) = \inf_j H_\nu(\alpha^m \mid \beta_j). \tag{41}$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

Also, since α^m and all of the β_j's are ν-continuity partitions, we have

$$H_\nu(\alpha^m \vee \beta_j) = \lim_{i\to\infty} H_{\nu_i}(\alpha^m \vee \beta_j)$$

and

$$H_\nu(\beta_j) = \lim_{i\to\infty} H_{\nu_i}(\beta_j).$$

This and the standard formulas

$$H_\nu(\alpha^m \mid \beta_j) = H_\nu(\alpha^m \vee \beta_j) - H_\nu(\beta_j)$$

and

$$H_{\nu_i}(\alpha^m \mid \beta_j) = H_{\nu_i}(\alpha^m \vee \beta_j) - H_{\nu_i}(\beta_j)$$

give

$$H_\nu(\alpha^m \mid \beta_j) = \lim_{i\to\infty} H_{\nu_i}(\alpha^m \mid \beta_j). \tag{42}$$

Now, letting $\epsilon_1 > 0$ and using (41), we can pick j such that

$$H_\nu(\alpha^m \mid \mathcal{B}_\mathcal{C}^-) > H_\nu(\alpha^m \mid \beta_j) - \epsilon_1.$$

Then, using (42), we have

$$\begin{aligned} H_\nu(\alpha^m \mid \mathcal{B}_\mathcal{C}^-) &\geq H_\nu(\alpha^m \mid \beta_j) - \epsilon_1 \\ &= \lim_{i\to\infty} H_{\nu_i}(\alpha^m \mid \beta_j) - \epsilon_1 \\ &\geq \limsup_{i\to\infty} H_{\nu_i}(\alpha^m \mid \mathcal{B}_\mathcal{C}^-) - \epsilon_1 \\ &\geq \limsup_{i\to\infty} H_{\nu_i}(\alpha^m \mid \mathcal{B}^\infty) - \epsilon_1. \end{aligned}$$

Since $H_\nu(\alpha^m \mid \mathcal{B}^-) \geq H_\nu(\alpha^m \mid \mathcal{B}_\mathcal{C}^-)$ and ϵ_1 was arbitrary, we get (34).

Next, we proceed to the proof of (35).

The measure ν_i is supported on the set

$$\bar{X} = \bigcup_{\ell=0}^{n_i-1} f^{-k_i}(f^\ell x_i).$$

Case 1. For every pair of integers $0 \leq r < s \leq k_i$, $f^r x_i \neq f^s x_i$.

In this case, let $\xi = \{f^{-k_i}(f^\ell x_i) : 0 \leq \ell < n_i\}$. This is a finite partition of $\bar{X}$, and each of the measures $f_\star^\ell \mu_i$ is supported on the element $f^{-k_i}(f^\ell x_i)$ of the partition ξ. Moreover, $f_\star^\ell \mu_i$ coincides with the conditional measure $\nu_{i,C}$ of ν_i on the element $C = f^{-k_i}(f^\ell x_i) \in \xi$. Note that $\nu_i(C) = 1/n_i$ for each $C \in \xi$.

We have

$$\begin{aligned} H_{\nu_i}(\alpha^m \mid \mathcal{B}^{k_i}) &= H_{\nu_i}(\alpha^m \mid \xi) \\ &= \sum_{C\in\xi} H_{\nu_{i,C}}(\alpha^m \mid C)\nu_i(C) \\ &= \frac{1}{n_i}\sum_{\ell=0}^{n_i-1} H_{f_\star^\ell \mu_i}(\alpha^m \mid f^{-k_i}(f^\ell x_i)) \\ &= \frac{1}{n_i}\sum_{\ell=0}^{n_i-1} H_{f_\star^\ell \mu_i}(\alpha^m \mid \mathcal{B}^{k_i}). \end{aligned}$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

So, in case 1, we actually have equality in (35).

Case 2. There are integers $0 \le s < t < n_i$ such that $f^s x_i = f^t x_i$.

In this case, let τ_1 be the least positive integer such that there is a $t > \tau_1$ such that $f^t x_i = f^{\tau_1} x_i$, and let τ be the least integer such that $\tau > \tau_1$ and $f^{\tau_1} x_i = f^{\tau_1+\tau} x_i$.

Then, $\{f^{-k_i}(f^\ell x_i) : 0 \le \ell < n_i\}$ is the disjoint union

$$\bigsqcup_{0\le \ell < \tau_1+\tau} f^{-k_i}(f^\ell x_i).$$

This time, let ξ be the partition

$$\xi = \{C = f^{-k_i}(f^\ell x_i) : 0 \le \ell < \tau_1 + \tau\},$$

and, again, let $\nu_{i,C}$ denote the conditional measure of ν_i on C.

For each $\tau_1 \le \ell < \tau_1+\tau$, let $D(\ell)$ be the set of integers in $[\tau_1, n_i)$ which are congruent to ℓ mod τ, and let $d(\ell) = \operatorname{card} D(\ell)$.

Then, for $0 \le \ell < \tau_1$, $f^\ell_\star \mu_i$ is supported on the single element $f^{-k_i}(f^\ell x_i)$ of $\mathcal{B}^{k_i}$ which gives

$$H_{f^\ell \mu_i}(\alpha^m \mid f^{-k_i}(f^\ell x_i)) = H_{f^\ell \mu_i}(\alpha^m \mid \mathcal{B}^{k_i}). \tag{43}$$

In addition, we have

$$\nu_{i, f^{-k_i}(f^\ell x_i)} = f^\ell_\star \mu_i \tag{44}$$

and

$$\nu_i(f^{-k_i}(f^\ell x_i)) = \frac{1}{n_i}. \tag{45}$$

On the other hand, for $\tau_1 \le \ell < \tau_1 + \tau$, and $s \in D(\ell)$, $f^s_\star \mu_i$ is supported on the single element $f^{-k_i}(f^\ell x_i)$ of $\mathcal{B}^{k_i}$ which gives

$$H_{f^s_\star \mu_i}(\alpha^m \mid f^{-k_i}(f^\ell x_i)) = H_{f^s_\star \mu_i}(\alpha^m \mid \mathcal{B}^{k_i}). \tag{46}$$

In addition, we have

$$\nu_{i, f^{-k_i}(f^\ell x_i)} = \frac{1}{d(\ell)} \sum_{s\in D(\ell)} f^s_\star \mu_i \tag{47}$$

and

$$\nu_i(f^{-k_i}(f^\ell x_i)) = \frac{d(\ell)}{n_i}. \tag{48}$$

From concavity of $s \to -s \log s$, we get, for $\tau_1 \le \ell < \tau_1 + \tau$,

$$H_{\nu_{i, f^{-k_i}(f^\ell x_i)}}(\alpha^m \mid f^{-k_i}(f^\ell x_i)) \ge \frac{1}{d(\ell)} \sum_{s\in D(\ell)} H_{f^s_\star \mu_i}(\alpha^m \mid f^{-k_i}(f^\ell x_i)). \tag{49}$$

Applying the above statements gives

$$\begin{aligned} H_{\nu_i}(\alpha^m \mid \mathcal{B}^{k_i}) &= H_{\nu_i}(\alpha^m \mid \xi) \\ &= \sum_{C\in\xi} H_{\nu_{i,C}}(\alpha^m \mid C)\nu_i(C) \\ &= \sum_{0\le \ell < \tau_1} H_{\nu_{i, f^{-k_i}(f^\ell x_i)}}(\alpha^m \mid f^{-k_i}(f^\ell x_i))\nu_i(f^{-k_i}(f^\ell x_i)) \\ &\quad + \sum_{\tau_1 \le \ell < \tau_1+\tau} H_{\nu_{i, f^{-k_i}(f^\ell x_i)}}(\alpha^m \mid f^{-k_i}(f^\ell x_i))\nu_i(f^{-k_i}(f^\ell x_i)) \end{aligned}$$

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press

$$\begin{aligned}
&\geq \frac{1}{n_i} \sum_{0\leq \ell<\tau_1} H_{f_\star^\ell \mu_i}(\alpha^m \mid f^{-k_i}(f^\ell x_i)) \\
&\quad + \frac{1}{n_i} \sum_{\tau_1\leq \ell<\tau_1+\tau} \sum_{s\in D(\ell)} H_{f_\star^s \mu_i}(\alpha^m \mid f^{-k_i}(f^\ell x_i)) \\
&= \frac{1}{n_i} \sum_{0\leq \ell<\tau_1} H_{f_\star^\ell \mu_i}(\alpha^m \mid \mathcal{B}^{k_i}) \\
&\quad + \frac{1}{n_i} \sum_{\tau_1\leq \ell<\tau_1+\tau} \sum_{s\in D(\ell)} H_{f_\star^s \mu_i}(\alpha^m \mid \mathcal{B}^{k_i}) \\
&= \frac{1}{n_i} \sum_{\ell=0}^{n_i-1} H_{f_\star^\ell \mu_i}(\alpha^m \mid \mathcal{B}^{k_i}),
\end{aligned}$$

completing the proofs of (35), Lemma 6.1 and Theorem 2.5.

References

[1] R. Bowen. Entropy-expansive maps. *Trans. Amer. Math. Soc.* **164** (1972), 323–333.

[2] M. Denker, C. Grillenberger and K. Sigmund. *Ergodic Theory on Compact Spaces (Lecture Notes in Mathematics, 527)*. Springer, Berlin, 1976.

[3] D. Fiebig, F. Fiebig and Z. Nitecki. Entropy and preimage sets. *Preprint*, 2003.

[4] M. Hurley. On topological entropy of maps. *Ergod. Th. & Dynam. Sys.* **15** (1995), 557–568.

[5] Ya. G. Sinai, I. P. Cornfeld and S. V. Fomin. *Ergodic Theory*. Springer, Berlin, 1982.

[6] R. Langevin and F. Przytycki. Entropie de l'image inverse d'une application. *Bull. Soc. Math. France* **120** (1991), 237–250.

[7] R. Langevin and P. Walczak. Entropie d'une dynamique. *C. R. Math. Acad. Sci. Paris Ser I.* **312** (1991), 141–144.

[8] Z. Nitecki and F. Przytycki. Preimage entropy for mappings. *Int. J. Bifurcation Chaos* **9**(9) (1999), 1815–1843.

[9] K. Petersen. *Ergodic Theory*. Cambridge University Press, Cambridge, 1983.

https://doi.org/10.1017/S0143385704000240 Published online by Cambridge University Press