

New approach to weighted topological entropy and pressure

MASAKI TSUKAMOTO

*Department of Mathematics, Kyushu University,
Moto-oka 744, Nishi-ku, Fukuoka 819-0395, Japan
(e-mail: tsukamoto@math.kyushu-u.ac.jp)*

(Received 10 August 2021 and accepted in revised form 14 December 2021)

Abstract. Motivated by fractal geometry of self-affine carpets and sponges, Feng and Huang [*J. Math. Pures Appl.* **106**(9) (2016), 411–452] introduced weighted topological entropy and pressure for factor maps between dynamical systems, and proved variational principles for them. We introduce a new approach to this theory. Our new definitions of weighted topological entropy and pressure are very different from the original definitions of Feng and Huang. The equivalence of the two definitions seems highly non-trivial. Their equivalence can be seen as a generalization of the dimension formula for the Bedford–McMullen carpet in purely topological terms.

Key words: dynamical system, weighted topological entropy, weighted topological pressure, variational principle, Bedford–McMullen carpet

2020 Mathematics Subject Classification: 37D35 (Primary); 37B40, 37A35, 37C45 (Secondary)

1. Introduction

1.1. *Weighted topological entropy and pressure.* The purpose of this paper is to introduce a new approach to *weighted topological entropy and pressure* introduced by Feng and Huang [FH16]. In this subsection, we describe their original theory. We explain our new approach in the next subsection.

We first quickly review the classical theory of entropy and pressure of dynamical systems. See the book of Walters [Wal82] for the details. A pair (X, T) is called a *dynamical system* if X is a compact metrizable space and $T : X \rightarrow X$ is a continuous map. We denote its *topological entropy* by $h_{\text{top}}(X, T)$. This is a topological invariant of dynamical systems, which counts the number of bits per iterate for describing the orbits of (X, T) .

One of the most basic theorems about topological entropy is the *variational principle*. We define $\mathcal{M}^T(X)$ as the set of invariant Borel probability measures on X . For each

measure $\mu \in \mathcal{M}^T(X)$, we denote its Kolomogorov–Sinai entropy by $h_\mu(T)$. Then the variational principle states that [Din70, Goodm71, Goodw69]

$$h_{\text{top}}(X, T) = \sup_{\mu \in \mathcal{M}^T(X)} h_\mu(T). \tag{1.1}$$

This theory can be generalized to *pressure*. Let (X, T) be a dynamical system with a continuous function $f : X \rightarrow \mathbb{R}$. Motivated by statistical mechanics, Ruelle [Rue73] (in some special cases) and Walters [Wal75] (for general systems) introduced the *topological pressure* $P(T, f)$ and proved the variational principle

$$P(T, f) = \sup_{\mu \in \mathcal{M}^T(X)} \left(h_\mu(T) + \int_X f \, d\mu \right). \tag{1.2}$$

The above (1.1) and (1.2) are classical and standard in ergodic theory. Recently, Feng and Huang [FH16] found an ingenious generalization of this classical theory. Motivated by fractal geometry of self-affine carpets and sponges [Bed84, KP96a, Mc84], they introduced *weighted versions* of entropy and pressure.

Let (X, T) and (Y, S) be dynamical systems. A map $\pi : X \rightarrow Y$ is called a *factor map* if π is a continuous surjection with $\pi \circ T = S \circ \pi$. We sometimes write $\pi : (X, T) \rightarrow (Y, S)$ for clarifying the maps T and S . For an invariant probability measure $\mu \in \mathcal{M}^T(X)$, we denote by $\pi_*\mu \in \mathcal{M}^S(Y)$ the push-forward of μ by π (this is defined by $\pi_*\mu(A) = \mu(\pi^{-1}A)$ for $A \subset Y$). Let $f : X \rightarrow \mathbb{R}$ be a continuous function, and let a_1, a_2 be two real numbers with $a_1 > 0$ and $a_2 \geq 0$. Feng and Huang [FH16, Question 1.1] asked (and then solved) the following question.

Question 1.1. How can one define a meaningful term $P^{(a_1, a_2)}(T, f)$ such that the following variational principle holds?

$$P^{(a_1, a_2)}(T, f) = \sup_{\mu \in \mathcal{M}^T(X)} \left(a_1 h_\mu(T) + a_2 h_{\pi_*\mu}(S) + \int_X f \, d\mu \right).$$

We describe their approach below. It is a modification of the definition of topological entropy given by Bowen [Bow73], which is in turn a modification of the standard definition of the Hausdorff dimension.

Here we explain only the case of $f \equiv 0$ for simplicity of the exposition. For the case of $f \not\equiv 0$, see their paper [FH16, §3.1] (they also studied the case where a sequence of factor maps $\pi_i : X_i \rightarrow X_{i+1}$ ($i = 1, 2, \dots, k$) is given. We think that our new approach can be also generalized to this setting. However, we concentrate on the simplest case in this paper).

Let d and d' be metrics on X and Y respectively. For $x \in X$, a natural number n and $\varepsilon > 0$, we define $B_n^{(a_1, a_2)}(x, \varepsilon) \subset X$ as the set of $y \in X$ satisfying the following two conditions:

$$\begin{aligned} d(T^j x, T^j y) &< \varepsilon \quad (0 \leq j < \lceil a_1 n \rceil); \\ d'(S^j \pi(x), S^j \pi(y)) &< \varepsilon \quad (0 \leq j < \lceil (a_1 + a_2)n \rceil). \end{aligned}$$

Here $\lceil u \rceil$ denotes the least integer not less than u . We call $B_n^{(a_1, a_2)}(x, \varepsilon)$ an (a_1, a_2) -weighted Bowen ball.

Let N be a natural number. We consider families of (a_1, a_2) -weighted Bowen balls $\{B_{n_j}^{(a_1, a_2)}(x_j, \varepsilon)\}_{j=1}^\infty$ satisfying

$$X = \bigcup_j B_{n_j}^{(a_1, a_2)}(x_j, \varepsilon), \quad n_j \geq N \text{ (for all } j \geq 1). \tag{1.3}$$

Let $s \geq 0$. We define $\Lambda_{N, \varepsilon}^{(a_1, a_2), s}(X)$ as the infimum of

$$\sum_j \exp(-sn_j),$$

where the infimum is taken over all families $\{B_{n_j}^{(a_1, a_2)}(x_j, \varepsilon)\}_{j=1}^\infty$ satisfying (1.3).

The quantity $\Lambda_{N, \varepsilon}^{(a_1, a_2), s}(X)$ is monotone in N . So we define

$$\Lambda_\varepsilon^{(a_1, a_2), s}(X) = \lim_{N \rightarrow \infty} \Lambda_{N, \varepsilon}^{(a_1, a_2), s}(X).$$

We vary the parameter s from 0 to ∞ . There exists a unique value of s , which we denote by $h_{\text{top}}^{(a_1, a_2)}(T, \varepsilon)$, where the value of $\Lambda_\varepsilon^{(a_1, a_2), s}(X)$ jumps from ∞ to 0:

$$\Lambda_\varepsilon^{(a_1, a_2), s}(X) = \begin{cases} 0 & (s > h_{\text{top}}^{(a_1, a_2)}(T, \varepsilon)), \\ \infty & (s < h_{\text{top}}^{(a_1, a_2)}(T, \varepsilon)). \end{cases}$$

Here, $h_{\text{top}}^{(a_1, a_2)}(T, \varepsilon)$ is monotone in ε . So we define the (a_1, a_2) -weighted topological entropy of $\pi : X \rightarrow Y$ by

$$h_{\text{top}}^{(a_1, a_2)}(\pi, T) = \lim_{\varepsilon \rightarrow 0} h_{\text{top}}^{(a_1, a_2)}(T, \varepsilon).$$

Feng and Huang [FH16, Theorem 1.4 and Corollary 1.5] solved Question 1.1 by this quantity.

THEOREM 1.2. (Feng and Huang, 2016) *In the above setting,*

$$h_{\text{top}}^{(a_1, a_2)}(\pi, T) = \sup_{\mu \in \mathcal{M}^T(X)} (a_1 h_\mu(T) + a_2 h_{\pi_* \mu}(S)).$$

1.2. New approach. In the previous subsection, we have described the original definition of weighted topological entropy introduced by Feng and Huang [FH16]. In this subsection, we explain our new approach. Our approach is a modification of the familiar definition of topological entropy (not the Hausdorff-dimension-like definition of [Bow73]).

First of all, notice that we can assume $a_1 + a_2 = 1$ in Question 1.1 because we can reduce the general case to this special case by a simple rescaling. So we study only this case. As in the previous subsection, here we explain the entropy case (that is, the case of $f \equiv 0$) for simplicity. We will explain the pressure case in §2.

Let (X, T) and (Y, S) be dynamical systems, and let $\pi : X \rightarrow Y$ be a factor map. Let d and d' be metrics on X and Y respectively. For a natural number N we define metrics d_N

and d'_N on X and Y respectively by

$$d_N(x_1, x_2) = \max_{0 \leq n < N} d(T^n x_1, T^n x_2), \quad d'_N(y_1, y_2) = \max_{0 \leq n < N} d'(S^n y_1, S^n y_2). \quad (1.4)$$

For $\varepsilon > 0$ and a non-empty subset $\Omega \subset X$, we define

$$\#(\Omega, N, \varepsilon) = \min \left\{ n \geq 1 \left| \begin{array}{l} \text{there exist open subsets } U_1, \dots, U_n \text{ of } X \text{ with} \\ \Omega \subset U_1 \cup \dots \cup U_n \text{ and } \text{diam}(U_k, d_N) < \varepsilon \\ \text{for all } 1 \leq k \leq n \end{array} \right. \right\}. \quad (1.5)$$

Here, $\text{diam}(U_k, d_N) = \sup_{x_1, x_2 \in U_k} d_N(x_1, x_2)$ is the diameter of U_k with respect to the metric d_N . When Ω is the empty set, we define $\#(\Omega, N, \varepsilon) = 0$. As is well known, the topological entropy of (X, T) is defined by

$$h_{\text{top}}(X, T) = \lim_{\varepsilon \rightarrow 0} \left(\lim_{N \rightarrow \infty} \frac{\log \#(X, N, \varepsilon)}{N} \right).$$

We will modify this definition.

Let $0 \leq w \leq 1$ be a real number. We set

$$\#^w(\pi, N, \varepsilon) = \min \left\{ \sum_{k=1}^n (\#(\pi^{-1}(V_k), N, \varepsilon))^w \left| \begin{array}{l} Y = V_1 \cup \dots \cup V_n \text{ is an open cover with} \\ \text{diam}(V_k, d'_N) < \varepsilon \text{ for all } 1 \leq k \leq n \end{array} \right. \right\}. \quad (1.6)$$

It is easy to check that this quantity is sub-multiplicative in N and monotone in ε . So we define the w -weighted topological entropy of $\pi : X \rightarrow Y$ by

$$h_{\text{top}}^w(\pi, T) = \lim_{\varepsilon \rightarrow 0} \left(\lim_{N \rightarrow \infty} \frac{\log \#^w(\pi, N, \varepsilon)}{N} \right).$$

This definition uses the metrics d and d' , but the value of $h_{\text{top}}^w(\pi, T)$ is a topological invariant (that is, independent of the choice of metrics).

The quantity $h_{\text{top}}^w(\pi, T)$ provides another solution to Question 1.1 for the case of $f \equiv 0$ and $(a_1, a_2) = (w, 1 - w)$. This is our main result for the weighted topological entropy.

THEOREM 1.3. (Variational principle for w -weighted topological entropy) For $0 \leq w \leq 1$,

$$h_{\text{top}}^w(\pi, T) = \sup_{\mu \in \mathcal{M}^T(X)} \{wh_{\mu}(T) + (1 - w)h_{\pi_*\mu}(S)\}.$$

As the above definition of $h_{\text{top}}^w(\pi, T)$ is close to the standard definition of topological entropy, the proof of this theorem is also close to a well-known proof of the standard variational principle. The basic structure of the proof is borrowed from the famous argument of Misiurewicz [Mis76]. At some technical points, we use the theory of principal extensions [DH13, Dow11].

By combining Theorems 1.2 and 1.3, we get a corollary.

COROLLARY 1.4. For $0 \leq w \leq 1$,

$$h_{\text{top}}^{(w, 1-w)}(\pi, T) = h_{\text{top}}^w(\pi, T).$$

Here the left-hand side is the weighted topological entropy $h_{\text{top}}^{(a_1, a_2)}(\pi, T)$ for $(a_1, a_2) = (w, 1 - w)$ defined in the previous subsection.

This corollary seems to be a very interesting statement. The author cannot see any direct way to prove it (without using the variational principles).

Problem 1.5. Can one prove the equality $h_{\text{top}}^{(w, 1-w)}(\pi, T) = h_{\text{top}}^w(\pi, T)$ without using measure theory?

The following example illustrates the importance of the equality $h_{\text{top}}^{(w, 1-w)}(\pi, T) = h_{\text{top}}^w(\pi, T)$.

Example 1.6. (Bedford–McMullen carpets) Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the circle, and let $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ be the torus. Let a and b be two natural numbers with $a \geq b \geq 2$. Set $A = \{0, 1, 2, \dots, a - 1\}$ and $B = \{0, 1, 2, \dots, b - 1\}$. Let $R \subset A \times B$ be a non-empty subset, and define

$$R' = \{y \in B \mid (x, y) \in R \text{ for some } x \in A\}.$$

We define $X \subset \mathbb{T}^2$ and $Y \subset \mathbb{T}$ by

$$X := \left\{ \left(\sum_{n=1}^{\infty} \frac{x_n}{a^n}, \sum_{n=1}^{\infty} \frac{y_n}{b^n} \right) \in \mathbb{T}^2 \mid (x_n, y_n) \in R \text{ for all } n \geq 1 \right\},$$

$$Y := \left\{ \sum_{n=1}^{\infty} \frac{y_n}{b^n} \in \mathbb{T} \mid y_n \in R' \text{ for all } n \geq 1 \right\}.$$

The space X is the famous Bedford–McMullen carpet [Bed84, Mc84]. We are going to explain that we can calculate the Hausdorff dimension of X (with respect to the natural metric on \mathbb{T}^2) by using Corollary 1.4.

We define continuous maps $T : X \rightarrow X$ and $S : Y \rightarrow Y$ by

$$T(x, y) = (ax, by), \quad S(y) = by.$$

Here, (X, T) and (Y, S) are dynamical systems. Let $\pi : X \rightarrow Y$ be the natural projection. Then π is a factor map between (X, T) and (Y, S) . We are interested in its weighted topological entropy. Set

$$w = \frac{\log b}{\log a} = \log_a b.$$

We have $0 \leq w \leq 1$. It directly follows from the definitions (the (a_1, a_2) -weighted Bowen ball $B_n^{(a_1, a_2)}(x, \varepsilon)$ for $a_1 = \log_a b$ and $a_2 = 1 - \log_a b$ is approximately a square of side length εb^{-n}) in §1.1 that the Hausdorff dimension of X is given by

$$\dim_H X = \frac{h_{\text{top}}^{(w, 1-w)}(\pi, T)}{\log b}.$$

From the equality $h_{\text{top}}^{(w,1-w)}(\pi, T) = h_{\text{top}}^w(\pi, T)$ in Corollary 1.4, we also have

$$\dim_H X = \frac{h_{\text{top}}^w(\pi, T)}{\log b}. \tag{1.7}$$

Now we calculate the w -weighted topological entropy $h^w(\pi, T)$.

CLAIM 1.7. For each $y \in B$, we define $t(y)$ as the number of $x \in A$ satisfying $(x, y) \in R$. Then

$$h_{\text{top}}^w(\pi, T) = \log \left(\sum_{y \in R'} t(y)^w \right).$$

Proof. First notice that in the definitions (1.5) and (1.6), we can use closed covers instead of open covers; this does not change their values. Here we will consider closed covers.

We define a metric d' on \mathbb{T} by

$$d'(x_1, x_2) = \min_{n \in \mathbb{Z}} |x_1 - x_2 - n|.$$

We define a metric d on \mathbb{T}^2 by

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

Let $\varepsilon > 0$ and take a natural number m with $b^{-m} < \varepsilon$. Let N be a natural number. For each $v \in (R')^{N+m}$, set

$$V_v = \left\{ \sum_{n=1}^{\infty} \frac{y_n}{b^n} \in Y \mid (y_1, \dots, y_{N+m}) = v \right\}.$$

These form a closed covering of Y with $\text{diam}(V_v, d'_N) < \varepsilon$. For each $(u, v) \in R^{N+m} \subset A^{N+m} \times B^{N+m}$ (where $u \in A^{N+m}$ and $v \in (R')^{N+m}$), we set

$$U_{(u,v)} = \left\{ \left(\sum_{n=1}^{\infty} \frac{x_n}{a^n}, \sum_{n=1}^{\infty} \frac{y_n}{b^n} \right) \in X \mid (x_1, \dots, x_{N+m}) = u, (y_1, \dots, y_{N+m}) = v \right\}.$$

These are closed subsets of X with $\text{diam}(U_{(u,v)}, d_N) < \varepsilon$ and

$$\pi^{-1}(V_v) = \bigcup_{\substack{u \in A^{N+m} \\ \text{with } (u,v) \in R^{N+m}}} U_{(u,v)}.$$

Hence, for $v = (v_1, \dots, v_{N+m}) \in (R')^{N+m}$,

$$\#(\pi^{-1}(V_v), N, \varepsilon) \leq t(v_1) \cdots t(v_{N+m}).$$

Therefore,

$$\#^w(\pi, N, \varepsilon) \leq \sum_{v_1, \dots, v_{N+m} \in R'} (t(v_1) \cdots t(v_{N+m}))^w = \left(\sum_{v \in R'} t(v)^w \right)^{N+m}.$$

Thus,

$$h_{\text{top}}^w(\pi, T) = \lim_{\varepsilon \rightarrow 0} \left(\lim_{N \rightarrow \infty} \frac{\log \#^w(\pi, N, \varepsilon)}{N} \right) \leq \log \left(\sum_{y \in B} t(y)^w \right).$$

Next, let $0 < \varepsilon < 1/a$. Fix $(p, q) \in R$. For a natural number N , we consider the following points in Y :

$$\sum_{n=1}^N \frac{v_n}{b^n} + \sum_{n=N+1}^{\infty} \frac{q}{b^n} \quad (v_1, \dots, v_N \in R'). \tag{1.8}$$

These points form an ε -separated set in Y with respect to the metric d'_N . We also consider the following points in X :

$$\left(\sum_{n=1}^N \frac{u_n}{a^n}, \sum_{n=1}^N \frac{v_n}{b^n} \right) + \sum_{n=N+1}^{\infty} \left(\frac{p}{a^n}, \frac{q}{b^n} \right) \quad ((u_1, v_1), \dots, (u_N, v_N) \in R). \tag{1.9}$$

These points form an ε -separated set in X with respect to the metric d_N .

Suppose $Y = V_1 \cup \dots \cup V_n$ is a covering with $\text{diam}(V_k, d'_N) < \varepsilon$. Then each V_k contains at most one point of (1.8). If V_k contains a point $\sum_{n=1}^N (v_n/b^n) + \sum_{n=N+1}^{\infty} (q/b^n)$, then $\pi^{-1}(V_k)$ contains $t(v_1) \dots t(v_N)$ points of the form (1.9) and hence

$$\#(\pi^{-1}(V_k), N, \varepsilon) \geq t(v_1) \dots t(v_N).$$

So

$$\#^w(\pi, N, \varepsilon) \geq \sum_{v_1, \dots, v_N \in R'} (t(v_1) \dots t(v_N))^w = \left(\sum_{v \in R'} t(v)^w \right)^N.$$

This shows

$$h_{\text{top}}^w(\pi, T) = \lim_{\varepsilon \rightarrow 0} \left(\lim_{N \rightarrow \infty} \frac{\log \#^w(\pi, N, \varepsilon)}{N} \right) \geq \log \left(\sum_{y \in R'} t(y)^w \right).$$

Notice that this proof of the claim is completely elementary. We have not used any sophisticated technique (in particular, measure theory). □

From (1.7) and Claim 1.7,

$$\dim_H X = \frac{\log(\sum_{y \in R'} t(y)^w)}{\log b} = \log_b \left(\sum_{y \in R'} t(y)^{\log_a b} \right). \tag{1.10}$$

This is a famous formula for the Hausdorff dimension of the Bedford–McMullen carpet [Bed84, Mc84]. Therefore, we conclude that the equality $h_{\text{top}}^{(1-w, w)}(\pi, T) = h_{\text{top}}^w(\pi, T)$ provides this famous formula fairly easily. This suggests that the equality $h_{\text{top}}^{(1-w, w)}(\pi, T) = h_{\text{top}}^w(\pi, T)$ is a rather deep statement. We can say that it is a *topological generalization* of the dimension formula for the Bedford–McMullen carpet.

Kenyon–Peres [KP96b, Theorems 1.1 and 3.2] generalized the formula (1.10) to closed T -invariant subsets of \mathbb{T}^2 which correspond to subshifts of finite type or sofic subshifts under the natural Markov partition. We can also prove their results from the equality $h_{\text{top}}^{(1-w,w)}(\pi, T) = h_{\text{top}}^w(\pi, T)$ as in the above.

The above example also illustrates that the two notions $h_{\text{top}}^{(a_1,a_2)}(\pi, T)$ and $h_{\text{top}}^w(\pi, T)$ have their own advantages. One of the great advantages of $h_{\text{top}}^{(a_1,a_2)}(\pi, T)$ is that its definition is intrinsically related to the Hausdorff dimension. So it can be directly applied to the study of geometric measure theory. The advantage of $h_{\text{top}}^w(\pi, T)$ is that its definition is elementary and hence (sometimes) easy to calculate.

In [FH16, pp. 441], Feng and Huang asked how to generalize their result to \mathbb{Z}^d -actions. It seems rather straightforward to generalize our new approach to \mathbb{Z}^d -actions and, possibly, actions of amenable groups.

Problem 1.8. Suppose that both $h_{\text{top}}^{(a_1,a_2)}(\pi, T)$ and $h_{\text{top}}^w(\pi, T)$ are generalized to group actions. Can one deduce any interesting consequence of their coincidence (like the above calculation of the Hausdorff dimension of the Bedford–McMullen carpet)?

We would like to mention the papers of Barral and Feng [BF09, BF12] and Feng [Fen11] (see also Yayama [Ya11a, Ya11b]). These papers studied Question 1.1 and related questions when (X, T) and (Y, S) are subshifts over finite alphabets. When (X, T) and (Y, S) are subshifts, the above definition of $h_{\text{top}}^w(\pi, T)$ (and its pressure version in §2) is essentially the same as that given in [BF09, Theorem 1.1] (see also [BF12, Theorem 3.1]). So we can say that the above definition generalizes the approach in [BF09, Theorem 1.1] from subshifts to general dynamical systems.

This paper studies only the abstract theory of $h_{\text{top}}^w(\pi, T)$ and its pressure version. However, the main motivation for the author to introduce these quantities is not to develop an abstract theory. The author naturally came up with the above definition of $h_{\text{top}}^w(\pi, T)$ when he studied the mean Hausdorff dimension of certain *infinite dimensional fractals*. (The mean Hausdorff dimension is a dynamical version of the Hausdorff dimension introduced in [LT19].) We plan to describe this connection in a separate paper.

2. Weighted topological pressure

In this section, we introduce our new definition of weighted topological pressure. For the original approach, see [FH16, §3.1].

Let $\pi : X \rightarrow Y$ be a factor map from a dynamical systems (X, T) to a dynamical system (Y, S) . Let $f : X \rightarrow \mathbb{R}$ be a continuous function.

Let d and d' be metrics on X and Y respectively. For a natural number N , we define new metrics d_N and d'_N on X and Y respectively by (1.4). We also define a continuous function $\mathbb{S}_N f : X \rightarrow \mathbb{R}$ by

$$\mathbb{S}_N f(x) = f(x) + f(Tx) + f(T^2x) + \dots + f(T^{N-1}x).$$

The metrics d_N, d'_N , and function $\mathbb{S}_N f$ are sometimes denoted by $d_N^T, (d')_N^S$, and $\mathbb{S}_N^T f$ respectively for clarifying the underlying dynamics.

For $\varepsilon > 0$ and a non-empty subset $\Omega \subset X$, we define

$$P(\Omega, f, N, \varepsilon) = \inf \left\{ \sum_{k=1}^n \exp \left(\sup_{U_k} \mathbb{S}_N f \right) \mid \left. \begin{array}{l} \text{there exist open subsets } U_1, \dots, U_n \text{ of } X \text{ with} \\ \Omega \subset U_1 \cup \dots \cup U_n \text{ and } \text{diam}(U_k, d_N) < \varepsilon \\ \text{for all } 1 \leq k \leq n \end{array} \right\}. \tag{2.1}$$

(When U_k is the empty set, we assume that the term $\exp(\sup_{U_k} \mathbb{S}_N f)$ is zero.) We sometimes denote $P(\Omega, f, N, \varepsilon)$ by $P_T(\Omega, f, N, \varepsilon)$ for clarifying the map T . When Ω is the empty set, we define $P(\Omega, f, N, \varepsilon) = 0$. It is well known that the topological pressure of (X, T, f) is given by

$$P(T, f) = \lim_{\varepsilon \rightarrow 0} \left(\lim_{N \rightarrow \infty} \frac{\log P(X, f, N, \varepsilon)}{N} \right).$$

We will modify this definition. Let $0 \leq w \leq 1$ be a real number. We set

$$P^w(\pi, f, N, \varepsilon) = \inf \left\{ \sum_{k=1}^n (P(\pi^{-1}(V_k), f, N, \varepsilon))^w \mid \left. \begin{array}{l} Y = V_1 \cup \dots \cup V_n \text{ is an open cover with} \\ \text{diam}(V_k, d'_N) < \varepsilon \text{ for all } 1 \leq k \leq n \end{array} \right\}. \tag{2.2}$$

We sometimes denote this by $P_T^w(\pi, f, N, \varepsilon)$.

The quantity $P^w(\pi, f, N, \varepsilon)$ is sub-multiplicative in N and monotone in ε . So we define the *w-weighted topological pressure* by

$$P^w(\pi, T, f) = \lim_{\varepsilon \rightarrow 0} \left(\lim_{N \rightarrow \infty} \frac{\log P^w(\pi, f, N, \varepsilon)}{N} \right).$$

The value of $P^w(\pi, T, f)$ is independent of the choices of the metrics d and d' . So it provides a topological invariant. We sometimes use the notation $P^w(\pi, X, T, Y, S, f)$ instead of $P^w(\pi, T, f)$ for clarifying all the data involved.

Now we state our main result of the paper.

THEOREM 2.1. (Variational principle for w -weighted topological pressure) *For any $0 \leq w \leq 1$,*

$$P^w(\pi, T, f) = \sup_{\mu \in \mathcal{M}^T(X)} \left(wh_\mu(T) + (1 - w)h_{\pi_*\mu}(S) + w \int_X f \, d\mu \right).$$

When $f \equiv 0$, we have $P^w(\pi, T, f) = h_{\text{top}}^w(\pi, T)$. So Theorem 1.3 in §1.2 follows from Theorem 2.1. The proof of Theorem 2.1 occupies the rest of the paper.

For the simplicity of the notation, we write

$$P_{\text{var}}^w(\pi, T, f) := \sup_{\mu \in \mathcal{M}^T(X)} \left(wh_\mu(T) + (1 - w)h_{\pi_*\mu}(S) + w \int_X f \, d\mu \right). \tag{2.3}$$

(Here *var* is the abbreviation of *variational*.) Then our main purpose is to prove the equality

$$P^w(\pi, T, f) = P^w_{\text{var}}(\pi, T, f).$$

In the rest of this section, we gather some elementary properties of w -weighted topological pressure. Here we always assume that $\pi : (X, T) \rightarrow (Y, S)$ is a factor map between dynamical systems with a continuous function $f : X \rightarrow \mathbb{R}$. We take $0 \leq w \leq 1$. Let d and d' be metrics on X and Y respectively.

LEMMA 2.2. *Let m be a natural number.*

$$P^w(\pi, T^m, \mathbb{S}_m f) = mP^w(\pi, T, f).$$

Here the left-hand side is $P^w(\pi, X, T^m, Y, S^m, \mathbb{S}_m^T f)$.

Proof. Let ε be a positive number. There exists $0 < \delta < \varepsilon$ such that

$$\begin{aligned} d(x_1, x_2) < \delta &\implies d_m^T(x_1, x_2) < \varepsilon \quad (x_1, x_2 \in X), \\ d'(y_1, y_2) < \delta &\implies (d')_m^S(y_1, y_2) < \varepsilon \quad (y_1, y_2 \in Y). \end{aligned}$$

Then for any natural number N ,

$$\begin{aligned} d_N^{T^m}(x_1, x_2) < \delta &\implies d_{mN}^T(x_1, x_2) < \varepsilon \quad (x_1, x_2 \in X), \\ (d')_N^{S^m}(y_1, y_2) < \delta &\implies (d')_{mN}^S(y_1, y_2) < \varepsilon \quad (y_1, y_2 \in Y). \end{aligned}$$

Because $\mathbb{S}_N^{T^m}(\mathbb{S}_m^T f) = \mathbb{S}_{mN}^T f$, for any subset $\Omega \subset X$,

$$P_{T^m}(\Omega, \mathbb{S}_m^T f, N, \varepsilon) \leq P_T(\Omega, f, mN, \varepsilon) \leq P_{T^m}(\Omega, \mathbb{S}_m^T f, N, \delta).$$

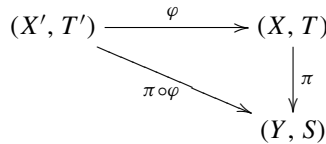
Then,

$$P_{T^m}^w(\pi, \mathbb{S}_m^T f, N, \varepsilon) \leq P_T^w(\pi, f, mN, \varepsilon) \leq P_{T^m}^w(\pi, \mathbb{S}_m^T f, N, \delta).$$

Thus,

$$P^w(\pi, T^m, \mathbb{S}_m^T f) = mP^w(\pi, T, f). \quad \square$$

LEMMA 2.3. *Let (X', T') be a dynamical system, and let $\varphi : (X', T') \rightarrow (X, T)$ be a factor map.*



Then

$$P^w(\pi, T, f) \leq P^w(\pi \circ \varphi, T', f \circ \varphi).$$

Here the right-hand side is $P^w(\pi \circ \varphi, X', T', Y, S, f \circ \varphi)$.

Proof. Let \tilde{d} be a metric on X' . For any $\varepsilon > 0$ there exists $0 < \delta < \varepsilon$ satisfying

$$\tilde{d}(x_1, x_2) < \delta \implies d(\varphi(x_1), \varphi(x_2)) < \varepsilon.$$

Then for any $N > 0$

$$\tilde{d}_N(x_1, x_2) < \delta \implies d_N(\varphi(x_1), \varphi(x_2)) < \varepsilon.$$

From this, we have for any $\Omega \subset X'$

$$P_T(\varphi(\Omega), f, N, \varepsilon) \leq P_{T'}(\Omega, f \circ \varphi, N, \delta).$$

For any $V \subset Y$

$$\varphi((\pi \circ \varphi)^{-1}(V)) = \pi^{-1}(V).$$

So

$$P_T(\pi^{-1}(V), f, N, \varepsilon) \leq P_{T'}((\pi \circ \varphi)^{-1}(V), f \circ \varphi, N, \delta).$$

Then

$$P_T^w(\pi, f, N, \varepsilon) \leq P_{T'}^w(\pi \circ \varphi, f \circ \varphi, N, \delta).$$

Therefore

$$P^w(\pi, T, f) \leq P^w(\pi \circ \varphi, T', f \circ \varphi). \quad \square$$

The next lemma is a bit complicated. It might be better for some readers to look at Remark 2.5 below before reading the lemma. It will provide a clearer perspective.

LEMMA 2.4. *Let (Y', S') be a dynamical system and let $\phi : (Y', S') \rightarrow (Y, S)$ be a factor map. Define the fiber product,*

$$X \times_Y Y' = \{(x, y) \in X \times Y' \mid \pi(x) = \phi(y)\}.$$

Now, $(X \times_Y Y', T \times S')$ becomes a dynamical system. We define factor maps $\varphi : X \times_Y Y' \rightarrow X$ and $\Pi : X \times_Y Y' \rightarrow Y'$ by

$$\varphi(x, y) = x, \quad \Pi(x, y) = y.$$

The diagram is as follows:

$$\begin{array}{ccc} (X \times_Y Y', T \times S') & \xrightarrow{\varphi} & (X, T) \\ \Pi \downarrow & & \downarrow \pi \\ (Y', S') & \xrightarrow{\phi} & (Y, S) \end{array}$$

Then,

$$P^w(\pi, T, f) \leq P^w(\Pi, T \times S', f \circ \varphi).$$

Here the right-hand side is $P^w(\Pi, X \times_Y Y', T \times S', Y', S', f \circ \varphi)$.

Proof. The point of the proof is that for any subset $A \subset Y'$, we have

$$\pi^{-1}(\phi(A)) = \varphi(\Pi^{-1}(A)).$$

Let \tilde{d} be a metric on Y' and we define a metric ρ on $X \times_Y Y'$ by

$$\rho((x_1, y_1), (x_2, y_2)) = \max(d(x_1, x_2), \tilde{d}(y_1, y_2)).$$

Let ε be a positive number. We have

$$\rho((x_1, y_1), (x_2, y_2)) < \varepsilon \implies d(x_1, x_2) < \varepsilon.$$

Then for any natural number N and any subset $\Omega \subset X \times_Y Y'$,

$$P_T(\varphi(\Omega), f, N, \varepsilon) \leq P_{T \times S'}(\Omega, f \circ \varphi, N, \varepsilon).$$

In particular, for any subset $A \subset Y'$,

$$\begin{aligned} P_T(\pi^{-1}(\phi(A)), f, N, \varepsilon) &= P_T(\varphi(\Pi^{-1}(A)), f, N, \varepsilon) \\ &\leq P_{T \times S'}(\Pi^{-1}(A), f \circ \varphi, N, \varepsilon). \end{aligned} \tag{2.4}$$

There exists $0 < \delta < \varepsilon$ such that

$$\tilde{d}(y_1, y_2) < \delta \implies d'(\phi(y_1), \phi(y_2)) < \varepsilon.$$

Now we claim that

$$P_T^w(\pi, f, N, \varepsilon) \leq P_{T \times S'}^w(\Pi, f \circ \varphi, N, \delta).$$

Indeed, take any positive number C with

$$P_{T \times S'}^w(\Pi, f \circ \varphi, N, \delta) < C.$$

Then there exists an open covering $Y' = V_1 \cup \dots \cup V_n$ such that $\text{diam}(V_k, \tilde{d}_N) < \delta$ for all $1 \leq k \leq n$ and

$$\sum_{k=1}^n (P_{T \times S'}(\Pi^{-1}(V_k), f \circ \varphi, N, \delta))^w < C.$$

We can find compact subsets $A_k \subset V_k$ satisfying $Y' = A_1 \cup \dots \cup A_n$. We have

$$\begin{aligned} \sum_{k=1}^n (P_T(\pi^{-1}(\phi(A_k)), f, N, \varepsilon))^w &\leq \sum_{k=1}^n (P_{T \times S'}(\Pi^{-1}(A_k), f \circ \varphi, N, \varepsilon))^w \quad \text{by (2.4)} \\ &\leq \sum_{k=1}^n (P_{T \times S'}(\Pi^{-1}(A_k), f \circ \varphi, N, \delta))^w \quad \text{by } \delta < \varepsilon \\ &\leq \sum_{k=1}^n (P_{T \times S'}(\Pi^{-1}(V_k), f \circ \varphi, N, \delta))^w \quad \text{by } A_k \subset V_k \\ &< C. \end{aligned}$$

Each $\phi(A_k)$ is a closed subset of Y with $\text{diam}(\phi(A_k), d'_N) < \varepsilon$. By the definition (2.1), there exist open subsets $W_k \supset \phi(A_k)$ of Y for $1 \leq k \leq n$ such that $\text{diam}(W_k, d'_N) < \varepsilon$ and

$$\sum_{k=1}^n (P(\pi^{-1}(W_k), f, N, \varepsilon))^w < C.$$

Noticing $Y = W_1 \cup \dots \cup W_n$, we have

$$P_T^w(\pi, f, N, \varepsilon) < C.$$

Because C is an arbitrary number larger than $P_{T \times S'}^w(\Pi, f \circ \varphi, N, \delta)$, this shows

$$P_T^w(\pi, f, N, \varepsilon) \leq P_{T \times S'}^w(\Pi, f \circ \varphi, N, \delta).$$

Thus, we conclude

$$P^w(\pi, T, f) \leq P^w(\Pi, T \times S', f \circ \varphi). \quad \square$$

Remark 2.5. Let (X', T') and (Y', S') be dynamical systems, and let $\pi' : X' \rightarrow Y'$ be a factor map. Suppose there exist factor maps $\varphi : (X', T') \rightarrow (X, T)$ and $\phi : (Y', S') \rightarrow (Y, S)$ satisfying $\pi \circ \varphi = \phi \circ \pi'$.

$$\begin{array}{ccc} (X', T') & \xrightarrow{\varphi} & (X, T) \\ \pi' \downarrow & & \downarrow \pi \\ (Y', S') & \xrightarrow{\phi} & (Y, S) \end{array}$$

Then,

$$P^w(\pi, T, f) \leq P^w(\pi', T', f \circ \varphi). \quad (2.5)$$

Here the right-hand side is $P^w(\pi', X', T', Y', S', f \circ \varphi)$. Lemmas 2.3 and 2.4 are special cases of this statement. We can prove (2.5) by using the variational principle (Theorem 2.1). However, it seems difficult to prove it in an elementary way. We will not use (2.5) in the paper.

Finally, we mention two basic results on calculus, which underpin many arguments of this paper.

LEMMA 2.6.

(1) For $0 \leq w \leq 1$ and non-negative numbers x, y ,

$$(x + y)^w \leq x^w + y^w.$$

(2) Let p_1, \dots, p_n be non-negative numbers with $p_1 + \dots + p_n = 1$. For any real numbers x_1, \dots, x_n ,

$$\sum_{i=1}^n (-p_i \log p_i + p_i x_i) \leq \log \sum_{i=1}^n e^{x_i}.$$

In particular (letting $x_1 = \dots = x_n = 0$),

$$-\sum_{i=1}^n p_i \log p_i \leq \log n.$$

Proof. (1) is completely elementary. (2) is proved in [Wal82, §9.3, Lemma 9.9]. □

3. *Kolmogorov–Sinai entropy*

In this section, we review basic definitions on Kolmogorov–Sinai entropy. For the details, see the book of Walters [Wal82].

Let (X, μ) be a probability measure space, namely X is a set equipped with a σ -algebra and μ is a probability measure defined on it. In our later applications, X is always a compact metrizable space with the standard Borel σ -algebra.

Let $\mathcal{A} = \{A_1, A_2, \dots, A_n\}$ be a finite measurable partition of X , namely each A_i is a measurable subset of X and

$$X = \bigcup_{i=1}^n A_i, \quad A_i \cap A_j = \emptyset \quad (i \neq j).$$

We define the *Shannon entropy* of \mathcal{A} by

$$H_\mu(\mathcal{A}) = - \sum_{i=1}^n \mu(A_i) \log \mu(A_i),$$

where we assume $0 \log 0 = 0$.

For another finite measurable partition $\mathcal{A}' = \{A'_1, A'_2, \dots, A'_m\}$, we set

$$\mathcal{A} \vee \mathcal{A}' = \{A_i \cap A'_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}.$$

This is a finite measurable partition of X . We define the *conditional entropy* by

$$H_\mu(\mathcal{A} \mid \mathcal{A}') = - \sum_{\substack{1 \leq j \leq m \\ \text{with } \mu(A'_j) > 0}} \mu(A'_j) \left\{ \sum_{i=1}^n \frac{\mu(A_i \cap A'_j)}{\mu(A'_j)} \log \frac{\mu(A_i \cap A'_j)}{\mu(A'_j)} \right\}.$$

Here, in the first summation, we have considered only the index j satisfying $\mu(A'_j) > 0$. We have [Wal82, Theorem 4.3(i)]

$$H_\mu(\mathcal{A} \vee \mathcal{A}') = H_\mu(\mathcal{A}') + H_\mu(\mathcal{A} \mid \mathcal{A}').$$

We write $\mathcal{A}' < \mathcal{A}$ if $\mathcal{A} \vee \mathcal{A}' = \mathcal{A}$. This is equivalent to the condition that for every $A \in \mathcal{A}$, there exists $A' \in \mathcal{A}'$ containing A . If $\mathcal{A}' < \mathcal{A}$, then

$$H_\mu(\mathcal{A} \mid \mathcal{A}') = H_\mu(\mathcal{A}) - H_\mu(\mathcal{A}')$$

and $H_\mu(\mathcal{A}') \leq H_\mu(\mathcal{A})$.

LEMMA 3.1.

- (1) $H_\mu(\mathcal{A})$ is subadditive in \mathcal{A} . Namely, for two finite measurable partitions \mathcal{A} and \mathcal{A}' of X ,

$$H_\mu(\mathcal{A} \vee \mathcal{A}') \leq H_\mu(\mathcal{A}) + H_\mu(\mathcal{A}').$$

- (2) $H_\mu(\mathcal{A})$ is concave in μ . Namely, for $0 \leq t \leq 1$ and two probability measures μ and μ' on X ,

$$H_{(1-t)\mu+t\mu'}(\mathcal{A}) \geq (1-t)H_\mu(\mathcal{A}) + tH_{\mu'}(\mathcal{A}).$$

Proof. See [Wal82, Theorem 4.3(viii)] and [Wal82, §8.1 Remark] for the proofs of (1) and (2) respectively. \square

Let $T : X \rightarrow X$ be a measurable map satisfying $T_*\mu = \mu$. Let \mathcal{A} be a finite measurable partition of X . For a natural number N , we define a new measurable partition \mathcal{A}^N of X by

$$\mathcal{A}^N = \mathcal{A} \vee T^{-1}\mathcal{A} \vee T^{-2}\mathcal{A} \vee \dots \vee T^{-(N-1)}\mathcal{A}.$$

We define the entropy $h_\mu(T, \mathcal{A})$ by

$$h_\mu(T, \mathcal{A}) = \lim_{N \rightarrow \infty} \frac{H_\mu(\mathcal{A}^N)}{N}.$$

Finally, we define the *Kolmogorov–Sinai entropy* of the measure-preserving transformation T by

$$h_\mu(T) = \sup\{h_\mu(T, \mathcal{A}) \mid \mathcal{A} \text{ is a finite measurable partition of } X\}.$$

We will need the following lemma later. See theorem 4.12(iv) of the book [Wal82, §4.5] for the proof.

LEMMA 3.2. *If \mathcal{A} and \mathcal{A}' are two finite measurable partitions of X , then*

$$h_\mu(T, \mathcal{A}) \leq h_\mu(T, \mathcal{A}') + H_\mu(\mathcal{A} \mid \mathcal{A}').$$

4. *Proof of $P_{\text{var}}^w(\pi, T, f) \leq P^w(\pi, T, f)$*

Let $\pi : (X, T) \rightarrow (Y, S)$ be a factor map between dynamical systems and let $f : X \rightarrow \mathbb{R}$ be a continuous function. The purpose of this section is to prove a half of the variational principle.

PROPOSITION 4.1. *For any $0 \leq w \leq 1$ and $\mu \in \mathcal{M}^T(X)$,*

$$wh_\mu(T) + (1 - w)h_{\pi_*\mu}(S) + w \int_X f \, d\mu \leq P^w(\pi, T, f).$$

Therefore, $P_{\text{var}}^w(\pi, T, f) \leq P^w(\pi, T, f)$.

Proof. Set $\nu = \pi_*\mu$. This is an invariant probability measure on Y . We will prove

$$wh_\mu(T) + (1 - w)h_\nu(S) + w \int_X f \, d\mu \leq P^w(\pi, T, f) + 1 + 2 \log 2. \tag{4.1}$$

If this is proved, then we will get the above statement by the standard *amplification trick*. Namely, for each natural number m , we apply (4.1) to $\pi : (X, T^m) \rightarrow (Y, S^m)$ with a continuous function $\mathbb{S}_m f : X \rightarrow \mathbb{R}$:

$$wh_\mu(T^m) + (1 - w)h_\nu(S^m) + w \int_X \mathbb{S}_m f \, d\mu \leq P^w(\pi, T^m, \mathbb{S}_m f) + 1 + 2 \log 2.$$

We have $h_\mu(T^m) = mh_\mu(T)$, $h_\nu(S^m) = mh_\nu(S)$, $\int_X \mathbb{S}_m f \, d\mu = m \int_X f \, d\mu$ and

$$P^w(\pi, T^m, \mathbb{S}_m f) = mP^w(\pi, T, f) \quad (\text{Lemma 2.2}).$$

Hence,

$$wh_\mu(T) + (1 - w)h_\nu(S) + w \int_X f d\mu \leq P^w(\pi, T, f) + \frac{1 + 2 \log 2}{m}.$$

Letting $m \rightarrow \infty$, we get the statement. So it is enough to prove (4.1).

Let $\mathcal{A} = \{A_1, \dots, A_\alpha\}$ be a finite measurable partition of Y and let \mathcal{B} be a finite measurable partition of X . We will prove that

$$wh_\mu(T, \mathcal{B}) + (1 - w)h_\nu(S, \mathcal{A}) + w \int_X f d\mu \leq P^w(\pi, T, f) + 1 + 2 \log 2. \tag{4.2}$$

For each A_a in \mathcal{A} ($1 \leq a \leq \alpha$), we take a compact subset $C_a \subset A_a$ satisfying

$$\sum_{a=1}^\alpha \nu(A_a \setminus C_a) < \frac{1}{\log \alpha}. \tag{4.3}$$

We set $C_0 = Y \setminus (C_1 \cup \dots \cup C_\alpha)$ and $\mathcal{C} = \{C_0, C_1, C_2, \dots, C_\alpha\}$.

CLAIM 4.2. \mathcal{C} is a finite measurable partition of Y satisfying

$$h_\nu(S, \mathcal{A}) < h_\nu(S, \mathcal{C}) + 1.$$

Proof. From Lemma 3.2,

$$h_\nu(S, \mathcal{A}) \leq h_\nu(S, \mathcal{C}) + H_\nu(\mathcal{A} | \mathcal{C}).$$

Because $C_a \subset A_a$ for $1 \leq a \leq \alpha$,

$$H_\nu(\mathcal{A} | \mathcal{C}) = \nu(C_0) \sum_{a=1}^\alpha \left(-\frac{\nu(A_a \cap C_0)}{\nu(C_0)} \log \frac{\nu(A_a \cap C_0)}{\nu(C_0)} \right) \leq \nu(C_0) \log \alpha.$$

The last term is smaller than one by (4.3). □

We consider $\mathcal{B} \vee \pi^{-1}(\mathcal{C})$, which has the form

$$\mathcal{B} \vee \pi^{-1}(\mathcal{C}) = \{B_{ab} \mid 0 \leq a \leq \alpha, 1 \leq b \leq \beta_a\}, \quad \pi^{-1}(C_a) = \bigcup_{b=1}^{\beta_a} B_{ab} \quad (0 \leq a \leq \alpha).$$

For each B_{ab} ($0 \leq a \leq \alpha, 1 \leq b \leq \beta_a$), we take a compact subset $D_{ab} \subset B_{ab}$ such that

$$\sum_{a=0}^\alpha \log \beta_a \left(\sum_{b=1}^{\beta_a} \mu(B_{ab} \setminus D_{ab}) \right) < 1. \tag{4.4}$$

We set

$$D_{a0} = \pi^{-1}(C_a) \setminus \bigcup_{b=1}^{\beta_a} D_{ab} \quad (0 \leq a \leq \alpha).$$

We define

$$\mathcal{D} = \{D_{ab} \mid 0 \leq a \leq \alpha, 0 \leq b \leq \beta_a\}.$$

CLAIM 4.3. \mathcal{D} is a finite measurable partition of X with $\pi^{-1}(\mathcal{C}) \prec \mathcal{D}$ and

$$h_\mu(T, \mathcal{B}) \leq h_\mu(T, \mathcal{D}) + 1.$$

Proof. $\pi^{-1}(\mathcal{C}) \prec \mathcal{D}$ is obvious by the construction.

$$\begin{aligned} h_\mu(T, \mathcal{B}) &\leq h_\mu(T, \mathcal{B} \vee \pi^{-1}(\mathcal{C})) \\ &\leq h_\mu(T, \mathcal{D}) + H_\mu(\mathcal{B} \vee \pi^{-1}(\mathcal{C}) \mid \mathcal{D}) \quad \text{by Lemma 3.2.} \end{aligned}$$

Because $D_{ab} \subset B_{ab}$ for $0 \leq a \leq \alpha$ and $1 \leq b \leq \beta_a$,

$$\begin{aligned} H_\mu(\mathcal{B} \vee \pi^{-1}(\mathcal{C}) \mid \mathcal{D}) &= \sum_{a=0}^\alpha \mu(D_{a0}) \sum_{b=1}^{\beta_a} \left(- \frac{\mu(D_{a0} \cap B_{ab})}{\mu(D_{a0})} \log \frac{\mu(D_{a0} \cap B_{ab})}{\mu(D_{a0})} \right) \\ &\leq \sum_{a=0}^\alpha \mu(D_{a0}) \log \beta_a \\ &< 1 \quad \text{by (4.4).} \end{aligned} \quad \square$$

We will prove that

$$wh_\mu(T, \mathcal{D}) + (1 - w)h_\nu(S, \mathcal{C}) + w \int_X f \, d\mu \leq P^w(\pi, T, f) + 2 \log 2.$$

If this is proved, then (4.2) will follow from Claims 4.2 and 4.3.

From the definition of the entropy,

$$\begin{aligned} wh_\mu(T, \mathcal{D}) + (1 - w)h_\nu(S, \mathcal{C}) &= \lim_{N \rightarrow \infty} \left(w \cdot \frac{H_\mu(\mathcal{D}^N)}{N} + (1 - w) \cdot \frac{H_\nu(\mathcal{C}^N)}{N} \right) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \{H_\nu(\mathcal{C}^N) + w(H_\mu(\mathcal{D}^N) - H_\nu(\mathcal{C}^N))\}. \end{aligned}$$

Because $\nu = \pi_*\mu$, we have $H_\nu(\mathcal{C}^N) = H_\mu(\pi^{-1}(\mathcal{C}^N))$. Because $\pi^{-1}(\mathcal{C}^N) \prec \mathcal{D}^N$,

$$H_\mu(\mathcal{D}^N) - H_\mu(\pi^{-1}(\mathcal{C}^N)) = H_\mu(\mathcal{D}^N \mid \pi^{-1}(\mathcal{C}^N)).$$

So,

$$wh_\mu(T, \mathcal{D}) + (1 - w)h_\nu(S, \mathcal{C}) = \lim_{N \rightarrow \infty} \frac{1}{N} \{H_\nu(\mathcal{C}^N) + w \cdot H_\mu(\mathcal{D}^N \mid \pi^{-1}(\mathcal{C}^N))\}.$$

We have

$$\int_X f \, d\mu = \frac{1}{N} \int_X \mathbb{S}_N f \, d\mu.$$

Therefore,

$$\begin{aligned} wh_\mu(T, \mathcal{D}) + (1 - w)h_\nu(S, \mathcal{C}) + w \int_X f \, d\mu &= \lim_{N \rightarrow \infty} \frac{1}{N} \{H_\nu(\mathcal{C}^N) + w \cdot H_\mu(\mathcal{D}^N \mid \pi^{-1}(\mathcal{C}^N)) + w \int_X \mathbb{S}_N f \, d\mu\}. \end{aligned} \quad (4.5)$$

For $C \in \mathcal{C}^N$, we define

$$\mathcal{D}_C^N = \{D \in \mathcal{D}^N \mid D \cap \pi^{-1}(C) \neq \emptyset\} = \{D \in \mathcal{D}^N \mid D \subset \pi^{-1}(C), D \neq \emptyset\}.$$

Then,

$$\pi^{-1}(C) = \bigcup_{D \in \mathcal{D}_C^N} D.$$

For $C \in \mathcal{C}^N$ with $\nu(C) > 0$ and $D \in \mathcal{D}_C^N$, we set

$$\mu(D \mid C) = \frac{\mu(D)}{\nu(C)} = \frac{\mu(D)}{\mu(\pi^{-1}(C))}.$$

For each $C \in \mathcal{C}^N$ with $\nu(C) > 0$, we have

$$\sum_{D \in \mathcal{D}_C^N} \mu(D \mid C) = 1.$$

CLAIM 4.4. *We have the following inequality:*

$$H_\nu(\mathcal{C}^N) + w \cdot H_\mu(\mathcal{D}^N \mid \pi^{-1}(\mathcal{C}^N)) + w \int_X \mathbb{S}_N f \, d\mu \leq \log \sum_{C \in \mathcal{C}^N} \left(\sum_{D \in \mathcal{D}_C^N} e^{\sup_D \mathbb{S}_N f} \right)^w.$$

Proof. We have

$$\begin{aligned} \int_X \mathbb{S}_N f \, d\mu &= \sum_{D \in \mathcal{D}^N} \int_D \mathbb{S}_N f \, d\mu \leq \sum_{D \in \mathcal{D}^N} \mu(D) \sup_D \mathbb{S}_N f \\ &= \sum_{\substack{C \in \mathcal{C}^N \\ \text{with } \nu(C) > 0}} \nu(C) \left(\sum_{D \in \mathcal{D}_C^N} \mu(D \mid C) \sup_D \mathbb{S}_N f \right). \end{aligned}$$

Hence,

$$\begin{aligned} &H_\mu(\mathcal{D}^N \mid \pi^{-1}(\mathcal{C}^N)) + \int_X \mathbb{S}_N f \, d\mu \\ &\leq \sum_{\substack{C \in \mathcal{C}^N \\ \text{with } \nu(C) > 0}} \nu(C) \left\{ \sum_{D \in \mathcal{D}_C^N} \left(-\mu(D \mid C) \log \mu(D \mid C) + \mu(D \mid C) \sup_D \mathbb{S}_N f \right) \right\}. \end{aligned}$$

By Lemma 2.6(2),

$$\sum_{D \in \mathcal{D}_C^N} \left(-\mu(D \mid C) \log \mu(D \mid C) + \mu(D \mid C) \sup_D \mathbb{S}_N f \right) \leq \log \sum_{D \in \mathcal{D}_C^N} e^{\sup_D \mathbb{S}_N f}.$$

So

$$H_\mu(\mathcal{D}^N \mid \pi^{-1}(\mathcal{C}^N)) + \int_X \mathbb{S}_N f \, d\mu \leq \sum_{C \in \mathcal{C}^N} \nu(C) \left(\log \sum_{D \in \mathcal{D}_C^N} e^{\sup_D \mathbb{S}_N f} \right).$$

Therefore,

$$\begin{aligned}
 & H_\nu(\mathcal{C}^N) + w \cdot H_\mu(\mathcal{D}^N \mid \pi^{-1}(\mathcal{C}^N)) + w \int_X \mathbb{S}_N f \, d\mu \\
 & \leq \sum_{C \in \mathcal{C}^N} \left\{ -\nu(C) \log \nu(C) + \nu(C) \log \left(\sum_{D \in \mathcal{D}_C^N} e^{\sup_D \mathbb{S}_N f} \right)^w \right\} \\
 & \leq \log \sum_{C \in \mathcal{C}^N} \left(\sum_{D \in \mathcal{D}_C^N} e^{\sup_D \mathbb{S}_N f} \right)^w \quad \text{by Lemma 2.6(2).} \quad \square
 \end{aligned}$$

We take metrics d and d' on X and Y respectively. Recall that C_a ($1 \leq a \leq \alpha$) are mutually disjoint compact subsets of Y and that D_{ab} ($0 \leq a \leq \alpha, 1 \leq b \leq \beta_a$) are mutually disjoint compact subsets of X . Hence, we can take $\varepsilon > 0$ such that:

(a) for any $y \in C_a$ and $y' \in C_{a'}$ with distinct $1 \leq a, a' \leq \alpha$,

$$\varepsilon < d'(y, y');$$

(b) for any $x \in D_{ab}$ and $x' \in D_{ab'}$ with $0 \leq a \leq \alpha$ and distinct $1 \leq b, b' \leq \beta_a$,

$$\varepsilon < d(x, x').$$

CLAIM 4.5. *Let N be a natural number.*

(1) *If a subset $V \subset Y$ has $\text{diam}(V, d'_N) < \varepsilon$, then the number of $C \in \mathcal{C}^N$ having non-empty intersection with V is at most 2^N :*

$$|\{C \in \mathcal{C}^N \mid C \cap V \neq \emptyset\}| \leq 2^N.$$

(2) *If a subset $U \subset X$ has $\text{diam}(U, d_N) < \varepsilon$, then for each $C \in \mathcal{C}^N$, the number of $D \in \mathcal{D}_C^N$ having non-empty intersection with U is at most 2^N :*

$$|\{D \in \mathcal{D}_C^N \mid D \cap U \neq \emptyset\}| \leq 2^N.$$

Proof. (1) For each $0 \leq k < N$, the set $S^k V$ may have non-empty intersection with C_0 and at most one set in $\{C_1, C_2, \dots, C_\alpha\}$. The above statement follows from this.

(2) Suppose $C \in \mathcal{C}^N$ has the form

$$C = C_{a_0} \cap S^{-1}C_{a_1} \cap S^{-2}C_{a_2} \cap \dots \cap S^{-(N-1)}C_{a_{N-1}},$$

with $0 \leq a_0, \dots, a_{N-1} \leq \alpha$. Recall that $\{D_{a_k 0}, D_{a_k 1}, D_{a_k 2}, \dots, D_{a_k \beta_{a_k}}\}$ is a partition of $\pi^{-1}(C_{a_k})$. Then any set $D \in \mathcal{D}_C^N$ has the form

$$D = D_{a_0 b_0} \cap T^{-1}D_{a_1 b_1} \cap T^{-2}D_{a_2 b_2} \cap \dots \cap T^{-(N-1)}D_{a_{N-1} b_{N-1}},$$

with $0 \leq b_k \leq \beta_{a_k}$ for $0 \leq k \leq N - 1$.

For each $0 \leq k < N$, the set $T^k U$ may have non-empty intersection with $D_{a_k 0}$ and, at most, one set in $\{D_{a_k 1}, D_{a_k 2}, \dots, D_{a_k \beta_{a_k}}\}$. Now the above statement follows from this. □

Let N be a natural number. Suppose we are given an open cover $Y = V_1 \cup \dots \cup V_n$ with $\text{diam}(V_i, d'_N) < \varepsilon$ for all $1 \leq i \leq n$. Moreover, suppose that for each $1 \leq i \leq n$, we

are given an open cover $\pi^{-1}(V_i) = U_{i1} \cup U_{i2} \cup \dots \cup U_{im_i}$ with $\text{diam}(U_{ij}, d_N) < \varepsilon$ for all $1 \leq j \leq m_i$. We are going to prove

$$\log \sum_{C \in \mathcal{C}^N} \left(\sum_{D \in \mathcal{D}_C^N} e^{\sup_D \mathbb{S}_N f} \right)^w \leq 2N \log 2 + \log \sum_{i=1}^n \left(\sum_{j=1}^{m_i} e^{\sup_{U_{ij}} \mathbb{S}_N f} \right)^w \tag{4.6}$$

Suppose this is proved. Then by Claim 4.4,

$$\begin{aligned} H_v(\mathcal{C}^N) + w \cdot H_\mu(\mathcal{D}^N \mid \pi^{-1}(\mathcal{C}^N)) + w \int_X \mathbb{S}_N f \, d\mu \\ \leq 2N \log 2 + \log \sum_{i=1}^n \left(\sum_{j=1}^{m_i} e^{\sup_{U_{ij}} \mathbb{S}_N f} \right)^w \end{aligned}$$

Taking the infimum over $\{V_i\}$ and $\{U_{ij}\}$ satisfying the above assumptions, we have

$$H_v(\mathcal{C}^N) + w \cdot H_\mu(\mathcal{D}^N \mid \pi^{-1}(\mathcal{C}^N)) + w \int_X \mathbb{S}_N f \, d\mu \leq 2N \log 2 + \log P^w(\pi, f, N, \varepsilon).$$

Divide this by N and let $N \rightarrow \infty$. Recalling (4.5), we get

$$wh_\mu(T, \mathcal{D}) + (1 - w)h_v(S, \mathcal{C}) + w \int_X f \, d\mu \leq 2 \log 2 + \lim_{N \rightarrow \infty} \frac{\log P^w(\pi, f, N, \varepsilon)}{N}.$$

Letting $\varepsilon \rightarrow 0$, we get the desired result:

$$wh_\mu(T, \mathcal{D}) + (1 - w)h_v(S, \mathcal{C}) + w \int_X f \, d\mu \leq 2 \log 2 + P^w(\pi, T, f).$$

So the rest of the work is to prove (4.6).

For $D \in \mathcal{D}^N$, we have

$$e^{\sup_D \mathbb{S}_N f} \leq \sum_{U_{ij} \cap D \neq \emptyset} e^{\sup_{U_{ij}} \mathbb{S}_N f}.$$

Here the sum is taken over the index (i, j) such that U_{ij} has non-empty intersection with D .

Let $C \in \mathcal{C}^N$. We define \mathcal{V}_C as the set of $1 \leq i \leq n$ such that $V_i \cap C \neq \emptyset$. By Claim 4.5(2),

$$\sum_{D \in \mathcal{D}_C^N} e^{\sup_D \mathbb{S}_N f} \leq 2^N \sum_{i \in \mathcal{V}_C} \sum_{j=1}^{m_i} e^{\sup_{U_{ij}} \mathbb{S}_N f}.$$

Then (recall $0 \leq w \leq 1$),

$$\begin{aligned} \left(\sum_{D \in \mathcal{D}_C^N} e^{\sup_D \mathbb{S}_N f} \right)^w &\leq 2^{Nw} \left(\sum_{i \in \mathcal{V}_C} \sum_{j=1}^{m_i} e^{\sup_{U_{ij}} \mathbb{S}_N f} \right)^w \\ &\leq 2^{Nw} \sum_{i \in \mathcal{V}_C} \left(\sum_{j=1}^{m_i} e^{\sup_{U_{ij}} \mathbb{S}_N f} \right)^w \quad \text{by Lemma 2.6(1).} \end{aligned}$$

Hence,

$$\sum_{C \in \mathcal{C}^N} \left(\sum_{D \in \mathcal{D}_C^N} e^{\sup_D \mathbb{S}_N f} \right)^w \leq 2^{Nw} \sum_{C \in \mathcal{C}^N} \left\{ \sum_{i \in \mathcal{V}_C} \left(\sum_{j=1}^{m_i} e^{\sup_{U_{ij}} \mathbb{S}_N f} \right)^w \right\}.$$

By Claim 4.5(1), for each $1 \leq i \leq n$, the number of $C \in \mathcal{C}^N$ satisfying $i \in \mathcal{V}_C$ is at most 2^N . So the right-hand side is bounded from above by

$$2^{Nw} \cdot 2^N \sum_{i=1}^n \left(\sum_{j=1}^{m_i} e^{\sup_{U_{ij}} \mathbb{S}_N f} \right)^w.$$

Therefore,

$$\sum_{C \in \mathcal{C}^N} \left(\sum_{D \in \mathcal{D}_C^N} e^{\sup_D \mathbb{S}_N f} \right)^w \leq 2^{Nw} \cdot 2^N \sum_{i=1}^n \left(\sum_{j=1}^{m_i} e^{\sup_{U_{ij}} \mathbb{S}_N f} \right)^w.$$

Taking the logarithm,

$$\begin{aligned} \log \sum_{C \in \mathcal{C}^N} \left(\sum_{D \in \mathcal{D}_C^N} e^{\sup_D \mathbb{S}_N f} \right)^w &\leq (N + Nw) \log 2 + \log \sum_{i=1}^n \left(\sum_{j=1}^{m_i} e^{\sup_{U_{ij}} \mathbb{S}_N f} \right)^w \\ &\leq 2N \log 2 + \log \sum_{i=1}^n \left(\sum_{j=1}^{m_i} e^{\sup_{U_{ij}} \mathbb{S}_N f} \right)^w. \end{aligned}$$

This is the estimate (4.6). So we have finished the proof of the proposition. □

5. Zero-dimensional principal extension

In this section, we prepare some definitions and results on principal extensions. The main reference is the book of Downarowicz [Dow11].

Let $\pi : (X, T) \rightarrow (Y, S)$ be a factor map between dynamical systems. Let d be a metric on X . We define the *topological conditional entropy* of π by

$$h_{\text{top}}(X, T \mid Y, S) = \lim_{\varepsilon \rightarrow 0} \left(\lim_{N \rightarrow \infty} \frac{\sup_{y \in Y} \log \#(\pi^{-1}(y), N, \varepsilon)}{N} \right).$$

Here, $\#(\pi^{-1}(y), N, \varepsilon)$ is the number defined by (1.5). It is easy to check that the quantity

$$\sup_{y \in Y} \log \#(\pi^{-1}(y), N, \varepsilon)$$

is sub-additive in N and monotone in ε . So the above limits exist. This definition of the topological conditional entropy arises from [Dow11, Lemma 6.8.2].

The factor map π is said to be *principal* if $h_{\text{top}}(X, T \mid Y, S) = 0$. In the case that this condition holds, the dynamical system (X, T) is called a *principal extension* of (Y, S) .

The next theorem shows an important consequence of this condition. This is proved in [Dow11, Corollary 6.8.9]. (See also the paper of Ledrappier and Walters [LW77].)

THEOREM 5.1. *A principal factor map preserves Kolmogorov–Sinai entropy. Namely, if $\pi : (X, T) \rightarrow (Y, S)$ is a principal factor map between dynamical systems, then for any*

invariant probability measure $\mu \in \mathcal{M}^T(X)$,

$$h_\mu(T) = h_{\pi_*\mu}(S).$$

Remark 5.2. Indeed, [Dow11, Corollary 6.8.9] proves the following more precise result. Let $\pi : (X, T) \rightarrow (Y, S)$ be a factor map with $h_{\text{top}}(Y, S) < \infty$. Then π is a principal factor map if and only if $h_\mu(T) = h_{\pi_*\mu}(S)$ for all $\mu \in \mathcal{M}^T(X)$.

LEMMA 5.3. *Let $(X, T), (Y, S), (Y', S')$ be dynamical systems. Let $\pi : X \rightarrow Y$ be a factor map and let $\phi : Y' \rightarrow Y$ be a principal factor map. We define the fiber product (see Lemma 2.4)*

$$X \times_Y Y' = \{(x, y) \in X \times Y' \mid \pi(x) = \phi(y)\}.$$

So, $(X \times_Y Y', T \times S')$ becomes a dynamical system. We define factor maps $\varphi : X \times_Y Y' \rightarrow X, T$ and $\Pi : X \times_Y Y' \rightarrow Y', S'$ by

$$\varphi(x, y) = x, \quad \Pi(x, y) = y.$$

$$\begin{array}{ccc} (X \times_Y Y', T \times S') & \xrightarrow{\varphi} & (X, T) \\ \Pi \downarrow & & \downarrow \pi \\ (Y', S') & \xrightarrow{\phi: \text{principal}} & (Y, S) \end{array}$$

Then φ is a principal factor map. (The map Π is not used in this statement, but we have introduced it for convenience in what follows.)

Proof. Let d and d' be metrics on X and Y' respectively. We define a metric ρ on $X \times_Y Y'$ by

$$\rho((x_1, y_1), (x_2, y_2)) = \max(d(x_1, x_2), d'(y_1, y_2)).$$

For any natural number N and $x \in X$, the metric space

$$(\varphi^{-1}(x), \rho_N)$$

is isometric to $(\phi^{-1}(\pi(x)), d'_N)$. Therefore, for any $\varepsilon > 0$,

$$\#(\varphi^{-1}(x), N, \varepsilon) = \#(\phi^{-1}(\pi(x)), N, \varepsilon).$$

So (recall that a factor map is always surjective),

$$\sup_{x \in X} \#(\varphi^{-1}(x), N, \varepsilon) = \sup_{x \in X} \#(\phi^{-1}(\pi(x)), N, \varepsilon) = \sup_{y \in Y} \#(\phi^{-1}(y), N, \varepsilon).$$

Thus,

$$h_{\text{top}}(X \times_Y Y', T \times S' \mid X, T) = h_{\text{top}}(Y', S' \mid Y, S) = 0. \quad \square$$

The next theorem is a key technical result. This is proved in [Dow11, Theorem 7.6.1]. (See also [DH13].) Here recall that a compact metrizable space is said to be *zero-dimensional* if clopen subsets form an open basis of the topology (a subset of a

topological space is called *clopen* if it is closed and open). For example, the Cantor set $\{0, 1\}^{\mathbb{N}}$ is zero-dimensional. A dynamical system (X, T) is said to be zero-dimensional if X is a zero-dimensional compact metrizable space.

THEOREM 5.4. *Every dynamical system has a zero-dimensional principal extension. Namely, for any dynamical system (X, T) , there exist a dynamical system (X', T') and a factor map $\phi : X' \rightarrow X$ such that X' is zero-dimensional and ϕ is principal.*

Recall that we have defined two terms $P^w(\pi, T, f)$ and $P_{\text{var}}^w(\pi, T, f)$ in §2.

COROLLARY 5.5. *Let $\pi : (X, T) \rightarrow (Y, S)$ be a factor map between dynamical systems with a continuous function $f : X \rightarrow \mathbb{R}$. There exists a factor map $\pi' : (X', T') \rightarrow (Y', S')$ with a continuous function $f' : X' \rightarrow \mathbb{R}$ satisfying the following two conditions.*

- (1) X' and Y' are zero-dimensional.
- (2) For any $0 \leq w \leq 1$, we have

$$P^w(\pi, T, f) \leq P^w(\pi', T', f'), \quad P_{\text{var}}^w(\pi', T', f') \leq P_{\text{var}}^w(\pi, T, f).$$

Proof. By Theorem 5.4, there exists a zero-dimensional principal extension $\phi : (Y', S') \rightarrow (Y, S)$. We consider the fiber product $(X \times_Y Y', T \times S')$ and the projections $\varphi : X \times_Y Y' \rightarrow X$ and $\Pi : X \times_Y Y' \rightarrow Y'$ as in Lemma 5.3. Then φ is a principal factor map.

$$\begin{array}{ccc} (X \times_Y Y', T \times S') & \xrightarrow{\varphi} & (X, T) \\ \Pi \downarrow & & \downarrow \pi \\ (Y', S') & \xrightarrow{\phi: \text{principal}} & (Y, S) \end{array}$$

By Lemma 2.4, for any $0 \leq w \leq 1$,

$$P^w(\pi, T, f) \leq P^w(\Pi, T \times S', f \circ \varphi).$$

Here, the right-hand side is $P^w(\Pi, X \times_Y Y', T \times S', Y', S', f \circ \varphi)$. By Theorem 5.1, for any invariant probability measure $\mu \in \mathcal{M}^{T \times S'}(X \times_Y Y')$,

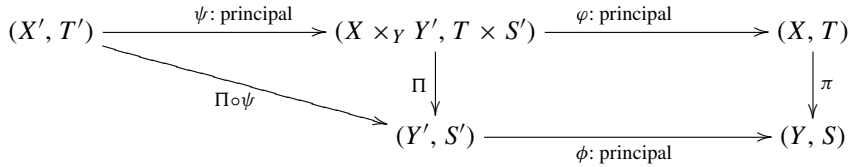
$$h_{\mu}(T \times S') = h_{\varphi_*\mu}(T), \quad h_{\Pi_*\mu}(S') = h_{\phi_*\Pi_*\mu}(S) = h_{\pi_*\varphi_*\mu}(S).$$

Then,

$$\begin{aligned} & P_{\text{var}}^w(\Pi, T \times S', f \circ \varphi) \\ &= \sup_{\mu \in \mathcal{M}^{T \times S'}(X \times_Y Y')} \left\{ wh_{\mu}(T \times S') + (1 - w)h_{\Pi_*\mu}(S') + w \int_{X \times_Y Y'} f \circ \varphi \, d\mu \right\} \\ &= \sup_{\mu \in \mathcal{M}^{T \times S'}(X \times_Y Y')} \left\{ wh_{\varphi_*\mu}(T) + (1 - w)h_{\pi_*\varphi_*\mu}(S) + w \int_X f \, d(\varphi_*\mu) \right\} \\ &\leq P_{\text{var}}^w(\pi, T, f). \end{aligned} \tag{5.1}$$

(Here, we prove $P_{\text{var}}^w(\Pi, T \times S', f \circ \varphi) \leq P_{\text{var}}^w(\pi, T, f)$. Indeed we can prove the equality $P_{\text{var}}^w(\Pi, T \times S', f \circ \varphi) = P_{\text{var}}^w(\pi, T, f)$ because the map $\varphi_* : \mathcal{M}^{T \times S'}(X \times_Y Y') \rightarrow \mathcal{M}^T(X)$ is surjective. However, we do not need this.)

By applying Theorem 5.4 to the system $(X \times_Y Y', T \times S')$, there exists a zero-dimensional principal extension $\psi : (X', T') \rightarrow (X \times_Y Y', T \times S')$.



By Lemma 2.3,

$$P^w(\Pi, T \times S', f \circ \varphi) \leq P^w(\Pi \circ \psi, T', f \circ \varphi \circ \psi).$$

Here, the right-hand side is $P^w(\Pi \circ \psi, X', T', Y', S', f \circ \varphi \circ \psi)$. As in the above (5.1), by Theorem 5.1,

$$P_{\text{var}}^w(\Pi \circ \psi, T', f \circ \varphi \circ \psi) \leq P_{\text{var}}^w(\Pi, T \times S', f \circ \varphi).$$

So we conclude

$$P^w(\pi, T, f) \leq P^w(\Pi \circ \psi, T', f \circ \varphi \circ \psi), \quad P_{\text{var}}^w(\Pi \circ \psi, T', f \circ \varphi \circ \psi) \leq P_{\text{var}}^w(\pi, T, f).$$

Set $\pi' := \Pi \circ \psi : (X', T') \rightarrow (Y', S')$ and $f' := f \circ \varphi \circ \psi : X' \rightarrow \mathbb{R}$. These satisfy the required conditions. □

6. Completion of the proof of the variational principle

In this section, we prove $P^w(\pi, T, f) \leq P_{\text{var}}^w(\pi, T, f)$ and complete the proof of the variational principle. First, we consider the case of zero-dimensional dynamical systems. Later, we will reduce the general case to this zero-dimensional case.

PROPOSITION 6.1. *Let $\pi : (X, T) \rightarrow (Y, S)$ be a factor map between zero-dimensional dynamical systems. Then, for any $0 \leq w \leq 1$ and a continuous function $f : X \rightarrow \mathbb{R}$,*

$$P^w(\pi, T, f) \leq P_{\text{var}}^w(\pi, T, f).$$

Proof. Let $\varepsilon > 0$. We will prove that there exists $\mu \in \mathcal{M}^T(X)$ satisfying

$$wh_{\mu}(T) + (1 - w)h_{\pi_*\mu}(S) + w \int_X f d\mu \geq \lim_{N \rightarrow \infty} \frac{\log P^w(\pi, f, N, \varepsilon)}{N}.$$

We take metrics d and d' on X and Y respectively. Let $Y = A_1 \cup \dots \cup A_\alpha$ be a clopen partition (that is, A_a are mutually disjoint clopen subsets of Y) with $\text{diam}(A_a, d') < \varepsilon$ for all $1 \leq a \leq \alpha$. Here we have used $\dim Y = 0$.

From $\dim X = 0$, for each $1 \leq a \leq \alpha$, we can also take a clopen partition

$$\pi^{-1}(A_a) = \bigcup_{b=1}^{\beta_a} B_{ab} \quad \text{with } \text{diam}(B_{ab}, d) < \varepsilon \text{ for all } 1 \leq b \leq \beta_a.$$

Set $\mathcal{A} = \{A_1, \dots, A_\alpha\}$ and $\mathcal{B} = \{B_{ab} \mid 1 \leq a \leq \alpha, 1 \leq b \leq \beta_a\}$. These are clopen partitions of Y and X respectively. We have $\pi^{-1}(\mathcal{A}) \prec \mathcal{B}$.

Let N be a natural number. We have $\pi^{-1}(\mathcal{A}^N) \prec \mathcal{B}^N$. For each non-empty $A \in \mathcal{A}^N$, we define

$$\mathcal{B}_A^N = \{B \in \mathcal{B}^N \mid B \cap \pi^{-1}(A) \neq \emptyset\} = \{B \in \mathcal{B}^N \mid B \subset \pi^{-1}(A), B \neq \emptyset\}.$$

We have

$$\pi^{-1}(A) = \bigcup_{B \in \mathcal{B}_A^N} B.$$

We set

$$Z_{N,A} = \sum_{B \in \mathcal{B}_A^N} e^{\sup_B \mathbb{S}_N f}.$$

Define

$$Z_N = \sum_{A \in \mathcal{A}^N} (Z_{N,A})^w.$$

Here, the sum is taken over only non-empty $A \in \mathcal{A}^N$. When we consider below a sum over $A \in \mathcal{A}^N$ (or $B \in \mathcal{B}^N$), we always assume that A (or B) is not empty.

We have

$$P^w(\pi, f, N, \varepsilon) \leq Z_N.$$

So it is enough to prove that there exists $\mu \in \mathcal{M}^T(X)$ satisfying

$$wh_\mu(T, \mathcal{B}) + (1 - w)h_{\pi_*\mu}(S, \mathcal{A}) + w \int_X f d\mu \geq \lim_{N \rightarrow \infty} \frac{\log Z_N}{N},$$

where the limit in the right-hand side exists because Z_N is sub-multiplicative in N .

Let N be a natural number. For non-empty $B \in \mathcal{B}^N$, we denote by $\mathcal{A}^N(B)$ the unique element of \mathcal{A}^N containing $\pi(B)$. For non-empty $A \in \mathcal{A}^N$, we have $\mathcal{A}^N(B) = A$ for all $B \in \mathcal{B}_A^N$.

For each non-empty set B in \mathcal{B}^N , we take a point $x_B \in B$ satisfying $\mathbb{S}_N f(x_B) = \sup_B \mathbb{S}_N f$. (Such a point exists because B is closed.) We define a probability measure

on X by

$$\begin{aligned} \sigma_N &= \frac{1}{Z_N} \sum_{B \in \mathcal{B}^N} (Z_{N, \mathcal{A}^N(B)})^{w-1} e^{\mathbb{S}_N f(x_B)} \cdot \delta_{x_B} \\ &= \frac{1}{Z_N} \sum_{A \in \mathcal{A}^N} \sum_{B \in \mathcal{B}_A^N} (Z_{N,A})^{w-1} e^{\mathbb{S}_N f(x_B)} \cdot \delta_{x_B}. \end{aligned}$$

Here, δ_{x_B} is the delta probability measure at the point x_B and σ_N is not an invariant measure in general. We set

$$\mu_N = \frac{1}{N} \sum_{n=0}^{N-1} T_*^n \sigma_N.$$

We can take a subsequence $\{\mu_{N_k}\}$ converging to an invariant probability measure μ on X in the weak* topology. We will prove that this measure μ satisfies

$$w h_\mu(T) + (1 - w) h_{\pi_* \mu}(S) + w \int_X f d\mu \geq \lim_{N \rightarrow \infty} \frac{\log Z_N}{N}.$$

CLAIM 6.2. For every natural number N ,

$$w H_{\sigma_N}(\mathcal{B}^N) + (1 - w) H_{\pi_* \sigma_N}(\mathcal{A}^N) + w \int_X \mathbb{S}_N f d\sigma_N = \log Z_N.$$

Proof. We have

$$\pi_* \sigma_N = \frac{1}{Z_N} \sum_{B \in \mathcal{B}^N} (Z_{N, \mathcal{A}^N(B)})^{w-1} e^{\mathbb{S}_N f(x_B)} \cdot \delta_{\pi(x_B)}.$$

For each non-empty $A \in \mathcal{A}^N$,

$$\begin{aligned} \pi_* \sigma_N(A) &= \frac{1}{Z_N} \sum_{B \in \mathcal{B}_A^N} (Z_{N, \mathcal{A}^N(B)})^{w-1} e^{\mathbb{S}_N f(x_B)} \\ &= \frac{1}{Z_N} (Z_{N,A})^w \quad \text{by } \mathcal{A}^N(B) = A \text{ for } B \in \mathcal{B}_A^N. \end{aligned}$$

Then,

$$H_{\pi_* \sigma_N}(\mathcal{A}^N) = \log Z_N - w \sum_{A \in \mathcal{A}^N} \frac{(Z_{N,A})^w}{Z_N} \log Z_{N,A}. \tag{6.1}$$

For non-empty $B \in \mathcal{B}^N$,

$$\sigma_N(B) = \frac{(Z_{N, \mathcal{A}^N(B)})^{w-1}}{Z_N} e^{\mathbb{S}_N f(x_B)}.$$

Then,

$$\begin{aligned}
 H_{\sigma_N}(\mathcal{B}^N) &= - \sum_{B \in \mathcal{B}^N} \frac{(Z_{N, \mathcal{A}^N(B)})^{w-1}}{Z_N} e^{\mathbb{S}_N f(x_B)} \log \left(\frac{(Z_{N, \mathcal{A}^N(B)})^{w-1}}{Z_N} e^{\mathbb{S}_N f(x_B)} \right) \\
 &= \frac{\log Z_N}{Z_N} \underbrace{\sum_{B \in \mathcal{B}^N} (Z_{N, \mathcal{A}^N(B)})^{w-1} e^{\mathbb{S}_N f(x_B)}}_{(I)} \\
 &\quad - \frac{w-1}{Z_N} \underbrace{\sum_{B \in \mathcal{B}^N} (Z_{N, \mathcal{A}^N(B)})^{w-1} e^{\mathbb{S}_N f(x_B)} \log Z_{N, \mathcal{A}^N(B)}}_{(II)} \\
 &\quad - \underbrace{\sum_{B \in \mathcal{B}^N} \frac{(Z_{N, \mathcal{A}^N(B)})^{w-1}}{Z_N} e^{\mathbb{S}_N f(x_B)} \mathbb{S}_N f(x_B)}_{(III)}.
 \end{aligned}$$

We calculate the term (I) by

$$(I) = \sum_{A \in \mathcal{A}^N} \sum_{B \in \mathcal{B}_A^N} (Z_{N,A})^{w-1} e^{\mathbb{S}_N f(x_B)} = \sum_{A \in \mathcal{A}^N} (Z_{N,A})^{w-1} \cdot Z_{N,A} = Z_N.$$

The term (II) is calculated by

$$(II) = \sum_{A \in \mathcal{A}^N} \sum_{B \in \mathcal{B}_A^N} (Z_{N,A})^{w-1} e^{\mathbb{S}_N f(x_B)} \log Z_{N,A} = \sum_{A \in \mathcal{A}^N} (Z_{N,A})^w \log Z_{N,A}.$$

For the term (III), we consider

$$\int_X \mathbb{S}_N f \, d\sigma_N = \frac{1}{Z_N} \sum_{B \in \mathcal{B}^N} (Z_{N, \mathcal{A}^N(B)})^{w-1} e^{\mathbb{S}_N f(x_B)} \mathbb{S}_N f(x_B) = (III).$$

Thus,

$$H_{\sigma_N}(\mathcal{B}^N) + \int_X \mathbb{S}_N f \, d\sigma_N = \log Z_N - \frac{w-1}{Z_N} \sum_{A \in \mathcal{A}^N} (Z_{N,A})^w \log Z_{N,A}.$$

Combining this with (6.1),

$$w H_{\sigma_N}(\mathcal{B}^N) + (1-w) H_{\pi_* \sigma_N}(\mathcal{A}^N) + w \int_X \mathbb{S}_N f \, d\sigma_N = \log Z_N. \quad \square$$

CLAIM 6.3. Let M and N be natural numbers. We have

$$\begin{aligned}
 \frac{1}{M} H_{\mu_N}(\mathcal{B}^M) &\geq \frac{1}{N} H_{\sigma_N}(\mathcal{B}^N) - \frac{2M \log |\mathcal{B}|}{N}, \\
 \frac{1}{M} H_{\pi_* \mu_N}(\mathcal{A}^M) &\geq \frac{1}{N} H_{\pi_* \sigma_N}(\mathcal{A}^N) - \frac{2M \log |\mathcal{A}|}{N}.
 \end{aligned}$$

Here, $|\mathcal{A}|$ and $|\mathcal{B}|$ are the cardinalities of \mathcal{A} and \mathcal{B} respectively.

Proof. This is rather standard. (See the proof of the standard variational principle in [Wal82, §8.2].) Here we provide the proof for \mathcal{B}^M . The case of \mathcal{A}^M is the same.

From the concavity of the entropy function (Lemma 3.1(2)), for $\mu_n = (1/N) \sum_{n=0}^{N-1} T_*^n \sigma_N$,

$$H_{\mu_N}(\mathcal{B}^M) \geq \frac{1}{N} \sum_{n=0}^{N-1} H_{T_*^n \sigma_N}(\mathcal{B}^M) = \frac{1}{N} \sum_{n=0}^{N-1} H_{\sigma_N}(T^{-n} \mathcal{B}^M). \tag{6.2}$$

Let $N = qM + r$ with $0 \leq r < M$,

$$\begin{aligned} \sum_{n=0}^{N-1} H_{\sigma_N}(T^{-n} \mathcal{B}^M) &= \sum_{t=0}^{M-1} \sum_{s=0}^q H_{\sigma_N}(T^{-sM-t} \mathcal{B}^M) - \sum_{n=qM+r}^{qM+M-1} H_{\sigma_N}(T^{-n} \mathcal{B}^M) \\ &\geq \sum_{t=0}^{M-1} \sum_{s=0}^q H_{\sigma_N}(T^{-sM-t} \mathcal{B}^M) - M \log |\mathcal{B}^M| \\ &\geq \sum_{t=0}^{M-1} \sum_{s=0}^q H_{\sigma_N}(T^{-sM-t} \mathcal{B}^M) - M^2 \log |\mathcal{B}|. \end{aligned} \tag{6.3}$$

We estimate $\sum_{s=0}^q H_{\sigma_N}(T^{-sM-t} \mathcal{B}^M)$ from below for each t . We have

$$T^{-sM-t} \mathcal{B}^M = \bigvee_{m=0}^{M-1} T^{-(sM+t+m)} \mathcal{B}.$$

When we fix $0 \leq t \leq M - 1$ and move $0 \leq s \leq q$ and $0 \leq m \leq M - 1$, the number $sM + t + m$ moves over

$$t, t + 1, t + 2, \dots, t + (q + 1)M - 1 \quad \text{without multiplicity.}$$

Hence,

$$\begin{aligned} &\sum_{s=0}^q H_{\sigma_N}(T^{-sM-t} \mathcal{B}^M) + \sum_{n=0}^{t-1} H_{\sigma_N}(T^{-n} \mathcal{B}) \\ &\geq H_{\sigma_N} \left(\bigvee_{n=0}^{t+(q+1)M-1} T^{-n} \mathcal{B} \right) \quad \text{by Lemma 3.1(1)} \\ &\geq H_{\sigma_N}(\mathcal{B}^N) \quad \text{by } t + (q + 1)M \geq (q + 1)M > N. \end{aligned}$$

Therefore,

$$\begin{aligned} \sum_{s=0}^q H_{\sigma_N}(T^{-sM-t} \mathcal{B}^M) &\geq H_{\sigma_N}(\mathcal{B}^N) - \sum_{n=0}^{t-1} H_{\sigma_N}(T^{-n} \mathcal{B}) \\ &\geq H_{\sigma_N}(\mathcal{B}^N) - t \log |\mathcal{B}| \\ &\geq H_{\sigma_N}(\mathcal{B}^N) - M \log |\mathcal{B}| \quad \text{by } t < M. \end{aligned}$$

Thus,

$$\sum_{t=0}^{M-1} \sum_{s=0}^q H_{\sigma_N}(T^{-sM-t} \mathcal{B}^M) \geq M \cdot H_{\sigma_N}(\mathcal{B}^N) - M^2 \log |\mathcal{B}|.$$

So by (6.3),

$$\begin{aligned} \sum_{n=0}^{N-1} H_{\sigma_N}(T^{-n} \mathcal{B}^M) &\geq \sum_{t=0}^{M-1} \sum_{s=0}^q H_{\sigma_N}(T^{-sM-t} \mathcal{B}^M) - M^2 \log |\mathcal{B}| \\ &\geq M \cdot H_{\sigma_N}(\mathcal{B}^N) - 2M^2 \log |\mathcal{B}|. \end{aligned}$$

From (6.2), we conclude that

$$\frac{1}{M} H_{\mu_N}(\mathcal{B}^M) \geq \frac{1}{NM} \sum_{n=0}^{N-1} H_{\sigma_N}(T^{-n} \mathcal{B}^M) \geq \frac{1}{N} H_{\sigma_N}(\mathcal{B}^N) - \frac{2M \log |\mathcal{B}|}{N}. \quad \square$$

We have

$$\int_X f \, d\mu_N = \frac{1}{N} \int_X \sum_{n=0}^{N-1} f \circ T^n \, d\sigma_N = \frac{1}{N} \int_X \mathbb{S}_N f \, d\sigma_N.$$

Claim 6.3 implies

$$\begin{aligned} &\frac{w}{M} H_{\mu_N}(\mathcal{B}^M) + \frac{1-w}{M} H_{\pi_*\mu_N}(\mathcal{A}^M) + w \int_X f \, d\mu_N \\ &\geq \frac{w}{N} H_{\sigma_N}(\mathcal{B}^N) + \frac{1-w}{N} H_{\pi_*\sigma_N}(\mathcal{A}^N) + \frac{w}{N} \int_X \mathbb{S}_N f \, d\sigma_N - \frac{2M(\log |\mathcal{A}| + \log |\mathcal{B}|)}{N} \\ &= \frac{\log Z_N}{N} - \frac{2M(\log |\mathcal{A}| + \log |\mathcal{B}|)}{N} \quad \text{by Claim 6.2.} \end{aligned}$$

Because $\mu_{N_k} \rightarrow \mu$ as $k \rightarrow \infty$, letting $N = N_k \rightarrow \infty$,

$$\frac{w}{M} H_{\mu}(\mathcal{B}^M) + \frac{1-w}{M} H_{\pi_*\mu}(\mathcal{A}^M) + w \int_X f \, d\mu \geq \lim_{N \rightarrow \infty} \frac{\log Z_N}{N}.$$

Here, we have used the clopenness of the elements of \mathcal{A}^M and \mathcal{B}^M . Finally, letting $M \rightarrow \infty$, we get

$$wh_{\mu}(T, \mathcal{B}) + (1-w)h_{\pi_*\mu}(S, \mathcal{A}) + w \int_X f \, d\mu \geq \lim_{N \rightarrow \infty} \frac{\log Z_N}{N}. \quad \square$$

Now we can prove the main result (Theorem 2.1). We repeat the statement for the convenience of readers.

THEOREM 6.4. (= Theorem 2.1) *Let $\pi : (X, T) \rightarrow (Y, S)$ be a factor map between dynamical systems. Then for any $0 \leq w \leq 1$ and a continuous function $f : X \rightarrow \mathbb{R}$,*

$$P^w(\pi, T, f) = P_{\text{var}}^w(\pi, T, f).$$

Proof. We already proved in Proposition 4.1 that

$$P_{\text{var}}^w(\pi, T, f) \leq P^w(\pi, T, f).$$

By Corollary 5.5, there exists a factor map $\pi' : (X', T') \rightarrow (Y', S')$ between zero-dimensional dynamical systems with a continuous function $f' : X' \rightarrow \mathbb{R}$ such that

$$P^w(\pi, T, f) \leq P^w(\pi', T', f'), \quad P_{\text{var}}^w(\pi', T', f') \leq P_{\text{var}}^w(\pi, T, f).$$

By Proposition 6.1,

$$P^w(\pi', T', f') \leq P_{\text{var}}^w(\pi', T', f').$$

Therefore,

$$P^w(\pi, T, f) \leq P_{\text{var}}^w(\pi, T, f).$$

So we conclude that

$$P^w(\pi, T, f) = P_{\text{var}}^w(\pi, T, f). \quad \square$$

Remark 6.5. The book of Downarowicz [Dow11] systematically develops the idea of using zero-dimensional dynamical systems in the study of entropy theory. The above proof is influenced by this idea. We also notice that it seems difficult to use this *zero-dimensional trick* in the proof of Proposition 4.1 in §4 because it is difficult to prove that principal extensions preserve weighted topological pressure without using the variational principle. A similar remark is given in [Dow11, Remark 7.6.12] about the proof of the standard variational principle.

Acknowledgements. M. Tsukamoto was supported by JSPS KAKENHI JP21K03227.

REFERENCES

- [Bed84] T. Bedford. Crinkly curves, Markov partitions and box dimension in self-similar sets. *PhD Thesis*, University of Warwick, 1984.
- [BF09] J. Barral and D.-J. Feng. Weighted thermodynamic formalism and applications. *Preprint*, 2009, arXiv:0909.4247.
- [BF12] J. Barral and D.-J. Feng. Weighted thermodynamic formalism on subshifts and applications. *Asian J. Math.* **16** (2012), 319–352.
- [Bow73] R. Bowen. Topological entropy for noncompact subsets. *Trans. Amer. Math. Soc.* **184** (1973), 125–136.
- [DH13] T. Downarowicz and D. Huczek. Zero-dimensional principal extensions. *Acta Appl. Math.* **126** (2013), 117–129.
- [Din70] E. I. Dinaburg. A correlation between topological entropy and metric entropy. *Dokl. Akad. Nauk SSSR* **190** (1970), 19–22.
- [Dow11] T. Downarowicz. *Entropy in Dynamical Systems*. Cambridge University Press, Cambridge, 2011.
- [Fen11] D.-J. Feng. Equilibrium states for factor maps between subshifts. *Adv. Math.* **226** (2011), 2470–2502.
- [FH16] D.-J. Feng and W. Huang. Variational principle for weighted topological pressure. *J. Math. Pures Appl. (9)* **106** (2016), 411–452.
- [Goodm71] T. N. T. Goodman. Relating topological entropy and measure entropy. *Bull. Lond. Math. Soc.* **3** (1971), 176–180.
- [Goodw69] L. W. Goodwyn. Topological entropy bounds measure-theoretic entropy. *Proc. Amer. Math. Soc.* **23** (1969), 679–688.
- [KP96a] R. Kenyon and Y. Peres. Measures of full dimension on affine-invariant sets. *Ergod. Th. & Dynam. Sys.* **16** (1996), 307–323.
- [KP96b] R. Kenyon and Y. Peres. Hausdorff dimensions of sofic affine-invariant sets. *Israel J. Math.* **94** (1996), 157–178.

- [LT19] E. Lindenstrauss and M. Tsukamoto. Double variational principle for mean dimension. *Geom. Funct. Anal.* **29** (2019), 1048–1109.
- [LW77] F. Ledrappier and P. Walters. A relativised variational principle for continuous transformations. *J. Lond. Math. Soc. (2)* **16** (1977), 568–576.
- [Mc84] C. McMullen. The Hausdorff dimension of general Sierpinski carpets. *Nagoya Math. J.* **96** (1984), 1–9.
- [Mis76] M. Misiurewicz. A short proof of the variational principle for \mathbb{Z}_+^N actions on a compact space. *Int. Conf. on Dyn. Syst. Math. Phys. (Rennes, 1975) (Astérisque, 40)*. Société Mathématique de France, Paris, 1976, pp. 145–157.
- [Rue73] D. Ruelle. Statistical mechanics on a compact set with \mathbb{Z}^n action satisfying expansiveness and specification. *Trans. Amer. Math. Soc.* **185** (1973), 237–251.
- [Wal75] P. Walters. A variational principle for the pressure of continuous transformations. *Amer. J. Math.* **17** (1975), 937–971.
- [Wal82] P. Walters. *An Introduction to Ergodic Theory*. Springer-Verlag, New York, 1982.
- [Ya11a] Y. Yayama. Existence of a measurable saturated compensation function between subshifts and its applications. *Ergod. Th. & Dynam. Sys.* **31** (2011), 1563–1589.
- [Ya11b] Y. Yayama. Application of a relative variational principle to dimension of nonconformal expanding maps. *Stoch. Dyn.* **11** (2011), 643–679.