LOCALIZED VARIATIONAL PRINCIPLE FOR NON-BESICOVITCH METRIC SPACES

TAMARA KUCHERENKO

Abstract. We consider the localized entropy of a point $w \in \mathbb{R}^m$ which is computed by considering only those (n, ε)-separated sets whose statistical sums with respect to an m-dimensional potential Φ are "close" to a given value w. Previously, a local version of the variational principle was established for systems on non-Besicovitch compact metric spaces. We extend this result to all compact metric spaces.

1. Introduction

1.1. **Motivation.** Topological entropy characterizes the complexity of dynamical systems. It describes the exponential growth rate of the number of distinguishable orbits as time advances. A related property is the measure-theoretic entropy of an invariant probability measure μ. Roughly speaking, it measures the entropy of the system if sets of μ-measure zero are ignored.

The most important characterization of topological entropy in terms of measure-theoretic entropy is the variational principle, which asserts that topological entropy equals the supremum of the measure-theoretic entropies over all invariant Borel probability measures.

For each point in a rotation set we can associate local versions of the measure-theoretic and the topological entropies. The measure-theoretic entropy in this context has been studied extensively by Geller and Misiurewicz in [5] as well as by Jenkinson in [6]. On the other hand, the localized topological entropy was only recently introduced in [8]. We arrived at a result that can be understood as a localized variational principle with the localization arising from only using measures having rotation vectors close to the local focus point.

One assumption required from the metric space for the localized variational principle to work was the Besicovitch Covering Property (see e.g. [2, 10]). It was not clear whether this property is essential or just an artifact of the proof. Here we show that the later is true.

We say that a metric space (X, d) satisfies the Besicovitch Covering Property if there exists an integer N so that for each family $\mathcal{B}$ of closed balls, whose centers form a bounded subset of X, there is a subfamily $\mathcal{F}$ covering

---

Key words and phrases. topological entropy, generalized rotation sets, variational principle, localized entropy.

This work was partially supported by a grant from the PSC-CUNY (TRADA-45-278).

the set of centers of the balls in $\mathcal{B}$, and such that each point of X is contained in at most N balls from $\mathcal{F}$. A large variety of dynamical systems satisfies this property. Some examples are subshifts of finite type, hyperbolic systems and continuous maps on compact smooth Riemannian manifolds (see [2, 10]).

Whether or not a metric space satisfies the Besicovitch Covering Property hinges on the metric it is endowed with. For example, any uncountable complete separable metric space can be re-metricised with a bilipschitz equivalent metric so that the Besicovitch Covering Property is not satisfied [12]. One source of examples of spaces without the Besicovitch Covering Property are Heisenberg groups with Korányi distance [7, 15] or Carnot-Carathéodory distance [14], which are geometrically sub-Riemannian manifolds. It is still not clear whether for these spaces there exist equivalent metrics for which the Besicovitch Covering Property holds.

In [8] we prove a local version of the variational principle for compact metric spaces which satisfy the Besicovitch Covering Property. The aim of this note is to extend this result to all compact metric spaces.

1.2. **Basic definitions and statement of the results.** Let $f : X \to X$ be a continuous map on a compact metric space (X, d). We consider a continuous potential $\Phi = (\phi_1, \cdots, \phi_m) : X \to \mathbb{R}^m$. We denote by $\mathcal{M}(f)$ the set of all Borel f-invariant probability measures on X endowed with the weak* topology. Following [6], we define the *generalized rotation set* of Φ by

$$\mathrm{Rot}(\Phi) = \{\mathrm{rv}_\Phi(\mu) : \mu \in \mathcal{M}\},$$

where $\mathrm{rv}_\Phi(\mu) = \left(\int \phi_1 \, d\mu, \ldots, \int \phi_m \, d\mu\right)$ denotes the rotation vector of the measure μ. We call $\mathcal{M}_\Phi(w) = \{\mu \in \mathcal{M} : \mathrm{rv}_\Phi(\mu) = w\}$ the rotation class of w. We refer to [6, 8, 17] for further details about rotation sets.

For $w \in \mathrm{Rot}(\Phi)$, we define the *localized measure-theoretic entropy* at w (with respect to Φ and f) by

$$h_{\mathrm{m}}(w, f, \Phi) = \sup\{h_\mu(f) : \mu \in \mathcal{M}_\Phi(w)\}.$$

Alternatively, we can adopt Bowen-Dinaburg approach to define localized topological entropy. For $n \in \mathbb{N}$ and $\varepsilon > 0$, we say that a set $F \subset X$ is (n, ε)-separated if for all $x, y \in F$ with $x \neq y$ we have $d_n(x, y) \overset{\mathrm{def}}{=} \max_{k=0,\cdots,n-1} d(f^k(x), f^k(y)) \geq \varepsilon$. Note that d_n is a metric (called Bowen metric) that induces the same topology on X as d. For $x \in X$ and $n \in \mathbb{N}$, we denote by $\frac{1}{n} S_n(\Phi, f)(x)$ the m-dimensional Birkhoff average at x of length n with respect to Φ and f, where $S_n(\Phi, f)(x) = (S_n(\phi_1, f)(x), \ldots, S_n(\phi_m, f)(x))$ and

$$S_n(\phi_i, f)(x) = \sum_{k=0}^{n-1} \phi_i(f^k(x)).$$

Given $w \in \mathbb{R}^m$ and $r > 0$ we say a set $F \subset X$ is a (n, ε, r, w)-set for Φ and f if F is (n, ε)-separated set and for all $x \in F$ the Birkhoff average $\frac{1}{n} S_n(\Phi, f)(x)$ is contained in the open Euclidean ball $B(w, r)$ with center w

and radius r. For all $n \in \mathbb{N}$ and $\varepsilon, r > 0$ we pick a maximal (with respect to the inclusion) (n, ε, r, w)-set $F_n(\varepsilon, r, w)$.

Then the *localized topological entropy* at $w \in \mathbb{R}^m$ (with respect to Φ and f) is defined by

$$h_{\rm top}(w, \Phi, f) = \lim_{r\to 0} \lim_{\varepsilon \to 0} \limsup_{n\to\infty} \frac{1}{n} \log {\rm card} \ \ F_n(\varepsilon, r, w) \tag{1}$$

This definition is analogous to that of the classical topological entropy with the exception that we here only consider orbits with Birkhoff averages close to w. As in the case of the classical topological entropy, this definition is independent of the choice of the sets $F_n(\varepsilon, r, w)$.

Note that the definition of $h_{\rm top}(w, \Phi, f)$ is only meaningful if $B(w, r)$ contains statistical averages for infinitely many n and arbitrarily small r. In case a point $w \in \mathbb{R}^m$ satisfies this property then it must be contained in the rotation set ${\rm Rot}(\Phi)$. However, this is not sufficient, see [8] for the precise condition. Here we are only interested in those points of the rotation set which can be approximated by rotation vectors of ergodic measures. For such points the localized topological entropy is well defined.

The classical variational principle (without localization) states that the topological and the measure-theoretic versions of the entropy coincide. However, it turns out that in the case of localized entropy the measure-theoretic and topological entropies may differ, and strict inequalities can occur in both directions [9]. On the other hand, the following result gives a fairly complete description of the assumptions needed to still have a variational principle.

Localized Variational Principle. [8] *Let $f : X \to X$ be a continuous map on a compact metric space X which satisfies the Besicovitch Covering Property. Let $\Phi : X \to \mathbb{R}^m$ be continuous and let $w \in {\rm Rot}(\Phi)$ be such that the map $v \mapsto h_m(v, f, \Phi)$ is continuous at w and $h_{\rm m}(w, f, \Phi)$ is approximated by ergodic measures. Then*

$$h_{\rm top}(w, f, \Phi) = h_{\rm m}(w, f, \Phi).$$

Here, we say that $h_{\rm m}(w, f, \Phi)$ is *approximated by ergodic measures* if there exists a sequence of ergodic measurs $(\mu_n)_{n\in\mathbb{N}}$ such that ${\rm rv}_\Phi(\mu_n) \to w$ and $h_{\mu_n}(f) \to h_{\rm m}(w, f, \Phi)$ as $n \to \infty$. The assumption that $h_{\rm m}(f, \Phi, w)$ is approximated by ergodic measures cannot be dropped in the previous theorem. Indeed, there are examples which do not satisfy this assumption and $h_{\rm top}(w, f, \Phi) < h_{\rm m}(w, f, \Phi)$ holds. On the other hand, without the assumption that $v \mapsto h_m(v, f, \Phi)$ is continuous at w, we arrive at the opposite inequality [9]. We recall that the continuity of $v \mapsto h_m(f, \Phi, v)$ holds for all points w if the entropy map $\mu \mapsto h_\mu(f)$ is upper semi-continuous. In particular, this is true if f is expansive [1], a C^∞ map on a compact smooth Riemannian manifold [11], or satisfies the entropy-expansiveness [4].

All conditions except the Besicovitch property are necessary for the conclusion of this theorem. However, the Besicovitch property appears to be nonessential. Moreover, the inequality $h_{\text{top}}(w, f, \Phi) \leq h_{\text{m}}(w, f, \Phi)$ is proven in [8] without any additional assumptions on the metric space X. Here we present an alternate proof of the opposite inequality that does not rely on the Besicovitch property.

A paper by Feng De-Jun and Huang Wen [3] discusses variational principle for topological entropies of compact, but not necessarily invariant subsets of X. To obtain their result the authors have to consider a broader range of Borel probability measures than f-invariant measures. The difference between their approach and what is presented here brings about an interesting problem. For a fixed point $w \in \text{Rot}(\Phi)$ we may consider a set of points in X whose Birkhoff averages with respect to Φ are getting in a suitable sense close to w. In case the set arising in this context is compact, we can determine whether its topological entropies considered in [3] coincide with $h_{\text{top}}(w, f, \Phi)$.

2. Localized Variational Principle for Entropy

This section is devoted to the proof of the following theorem.

Theorem 1. *Let $f : X \to X$ be a continuous map on a compact metric space X. Let $\Phi : X \to \mathbb{R}^m$ be continuous and let $w \in \text{Rot}(\Phi)$ be such that the map $v \mapsto h_m(v, f, \Phi)$ is continuous at w and $h_{\text{m}}(w, f, \Phi)$ is approximated by ergodic measures. Then $h_{\text{top}}(w, f, \Phi) = h_{\text{m}}(w, f, \Phi)$.*

Note that the inequality $h_{\text{top}}(w, f, \Phi) \leq h_{\text{m}}(w, f, \Phi)$ was proven in [8]. The proof of the opposite inequality $h_{\text{m}}(w, f, \Phi) \leq h_{\text{top}}(w, f, \Phi)$ relies on the following three lemmas.

We fix $w \in \text{Rot}(\Phi)$ and $r > 0$. We denote by

$$h(r, w, \Phi, f) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log \text{card } F_n(\varepsilon, r, w) \tag{2}$$

Here $F_n(\varepsilon, r, w)$ stands for a maximal (n, ε, w, r)-set. Then the localized topological entropy at w (with respect to f and Φ) is

$$h_{\text{top}}(w, f, \Phi) = \lim_{r \to 0} h(r, w, f, \Phi) \tag{3}$$

Lemma 1. *Let X be a metric space, $f : X \to X$ and $\Phi : X \to \mathbb{R}^m$. For $k \in \mathbb{N}$ denote by $\Phi_k = \frac{1}{k} S_k(\Phi, f)$. Then for any $n \in \mathbb{N}$ we have*

$$\frac{1}{n} S_n(\Phi_k, f^k) = \frac{1}{kn} S_{kn}(\Phi, f)$$

The equality in this lemma is proved by standard algebraic manipulation and thus we omit it here.

Lemma 2. *Let $f : X \to X$ be a continuous map on a compact metric space, $\Phi : X \to \mathbb{R}^m$ be a continuous potential and $w \in \text{Rot}(\Phi)$. For any $r > 0$ and $k \in \mathbb{N}$ we have*

$$h(r, w, f^k, \Phi_k) = k \cdot h(r, w, f, \Phi),$$

where $\Phi_k = \frac{1}{k} S_k(\Phi, f)$.

Proof. For an $\varepsilon > 0$ let F be any (n, ε, r, w)-set with respect to Φ_k and f^k. We will show that F is also a (kn, ε, r, w)-set with respect to Φ and f. For any $x, y \in F$ we have

$$\max_{0 \le i \le kn-1} d(f^i(x), f^i(y)) \ge \max_{0 \le j \le n-1} d(f^{jk}(x), f^{jk}(y)) > \varepsilon.$$

Moreover,

$$\frac{1}{kn} S_{kn}(\Phi, f)(x) = \frac{1}{n} S_n(\Phi_k, f^k)(x) \in B(r, w).$$

Since every (n, ε, r, w)-set with respect to Φ_k and f^k is a (kn, ε, r, w)-set with respect to Φ and f, we obtain

$$\operatorname{card} F_n(\varepsilon, r, w, f^k, \Phi_k) \le \operatorname{card} F_{kn}(\varepsilon, r, w, f, \Phi).$$

Therefore,

$$\frac{1}{n} \log \operatorname{card} F_n(\varepsilon, r, w, f^k, \Phi_k) \le k \cdot \frac{1}{kn} \log \operatorname{card} F_{kn}(\varepsilon, r, w, f, \Phi).$$

Passing to the upper limit as $n \to \infty$ and to the limit as $\varepsilon \to 0$ we obtain $h(r, w, f^k, \Phi_k) \le kh(r, w, f, \Phi)$.

To prove the opposite inequality we fix an $\varepsilon > 0$ and $n \in \mathbb{N}$ and let F be any (kn, ε, r, w)-set with respect to Φ and f. We use the uniform continuity of f on X to find a $0 < \delta < \varepsilon$ such that for $i = 0, ..., k$ we have $d(f^i(x), f^i(y)) < \frac{\varepsilon}{2}$ whenever $d(x, y) < \delta$. Denote be F_n any maximal (n, δ, r, w)-set with respect to Φ_k and f^k. If $x \in F$ then $\frac{1}{n} S_n(f^k, \Phi_k)(x) \in B(w, r)$ by Lemma 1. The maximality of F_n implies the existence of $y_x \in F_n$ such that

$$\max_{0 \le j \le n-1} d(f^{jk}(x), f^{jk}(y_x)) < \delta.$$

Then $d(f^i(x), f^i(y_x)) < \frac{\varepsilon}{2}$ for all $0 \le i \le kn - 1$. The map $x \in F \mapsto y_x \in F_n$ is injective. Indeed, if for some $x_1, x_2 \in F$ we have $y_{x_1} = y_{x_2}$ then the triangle inequality yields

$$\max_{0 \le i \le kn-1} d(f^i(x_1), f^i(x_2)) < \varepsilon.$$

This contradicts the fact that F is (kn, ε)-separated. Therefore, $\operatorname{card} F \le \operatorname{card} F_n$. Since F was arbitrary (kn, ε, r, w)-set with respect to Φ and f, we obtain

$$k \cdot \frac{1}{kn} \log \operatorname{card} F_{kn}(\varepsilon, r, w, f, \Phi) \le \frac{1}{n} \log \operatorname{card} F_n(\delta, r, w, f^k, \Phi_k).$$

Letting $n \to \infty$ and $\varepsilon \to 0$ we obtain the desired inequality. $\square$

Remark. *The consequence of this lemma is an analog of the power rule for classical entropy* $h_{\text{top}}(w, f^k, \Phi_k) = k \cdot h_{\text{top}}(w, f, \Phi)$.

Before formulating the next lemma we recall the standard definition of the entropy of a measure $\mu \in \mathcal{M}(f)$. Let $\mathcal{A} = \{A_1, A_2, ..., A_k\}$ be a finite partition of X. Then the entropy of the partition $\mathcal{A}$ is

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

Note that the convexity of the function $x \mapsto x\log x$ implies $H_\mu(\mathcal{A}) \leq \log \operatorname{card}(\mathcal{A})$.

The join of the partitions $f^{-j}(\mathcal{A}) = \{f^{-j}(A_1), ..., f^{-j}(A_k)\}$ is the partition $\mathcal{A}^n = \bigvee_{j=0}^{n-1} f^{-j}(\mathcal{A})$, which consists of all sets of the form $\cap_{j=0}^{n-1} f^{-j}(A_{i_j})$ with $A_{i_j} \in \mathcal{A}$. The entropy of f with respect to $\mathcal{A}$ is

$$h_\mu(f, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\mathcal{A}^n)$$

Finally, the entropy of the measure μ with respect to f is

$$h_\mu(f) = \sup\{h_\mu(f, \mathcal{A}) : \mathcal{A} \text{ is a finite partition of } X\}$$

Lemma 3. *Let $f : X \to X$ be a continuous map on a compact metric space, $\Phi : X \to \mathbb{R}^m$ be a continuous potential, $w \in \operatorname{Rot}(\Phi)$ and $r > 0$. Suppose that $\mu \in \mathcal{M}(f)$ is such that $\operatorname{rv}_\Phi(\mu) \in B(r, w)$ and for μ-almost all $x \in X$ $\lim_{n\to\infty} \frac{1}{n} S_n(f, \Phi)(x) = \operatorname{rv}_\Phi(\mu)$. Then*

$$h_\mu(f) \leq h(r, w, f, \Phi) + \log 2 + 1.$$

Proof. Let $\mathcal{A} = \{A_1, A_2, ..., A_k\}$ be any Borel partition of X. Choose $\varepsilon > 0$ such that $\varepsilon < \frac{1}{k\log k}$. Since the sequence $\frac{1}{n} S_n(f, \Phi)$ converges μ-almost everywhere to $\operatorname{rv}_\Phi(\mu)$, we use Egoroff's theorem to find subsets $D_i \subset A_i$, $(i = 1, ..., k)$ with the following properties

- D_i is compact
- $\mu(A_i \setminus D_i) < \varepsilon$
- $\frac{1}{n} S_n(f, \Phi) \to \operatorname{rv}_\Phi(\mu)$ uniformly on D_i

Let $D_0 = X \setminus \cup_{i=1}^{k} D_i$. Then $\mathcal{D} = \{D_0, D_1, ..., D_k\}$ is also a partition of X. Denote by $r_\mu = r - \|\operatorname{rv}_\Phi(\mu) - w\|$ and by

$$l = 4\left\lceil \frac{\sup\{\|\Phi(x)\| : x \in X\}}{r_\mu} \right\rceil.$$

We consider the join of the partitions

$$\mathcal{D}^{ln} = \bigvee_{j=0}^{ln-1} f^{-j}(\mathcal{D}).$$

We split $\mathcal{D}^{ln}$ into "good sets" $\mathcal{G}^n$ and "bad sets" $\mathcal{B}_i^n$, $(i = 0, ..., l-1)$ in the following way.

$$\mathcal{B}_0^n = D_0 \cap f^{-1}(D_0) \cap ... \cap f^{-ln+1}(D_0)$$
$$\mathcal{B}_i^n = \{D \in \mathcal{D}^{ln} \setminus \cup_{j=0}^{i-1} \mathcal{B}_j^n : D \subset D_0 \cap f^{-1}(D_0) \cap ... \cap f^{-(l-i)n+1}(D_0)\}$$

Then $\mathcal{G}^n = \mathcal{D}^{ln} \setminus \cup_{i=0}^{l-1} \mathcal{B}_i^n$. We also denote by

$$E_n = \left\{ x \in X : \frac{1}{m} S_m(f, \Phi) \in B(r, w) \text{ for any } m > ln \right\}$$

Since $\frac{1}{n} S_n(f, \Phi)$ converges uniformly on D_i $(i \neq 0)$ to the rotation vector of μ, there is $N_\mu > 0$ such that for any $n > N_\mu$ and any $x \in \cup_{i=1}^k D_i$ we have

$$\left\| \frac{1}{n} S_n(f, \Phi)(x) - \mathrm{rv}_\Phi(\mu) \right\| < \frac{r_\mu}{2}.$$

From now on we will consider $n > N_\mu$.

First we will show that for such n any set $G \in \mathcal{G}^n$ is a subset of E_n. Pick any $x \in G$. Then there is $s < n$ such that $f^s(x) \in D_i$ for some $i \neq 0$. Then for any $m > ln$ we have

$$\begin{aligned} \left\| \frac{1}{m} S_m(f, \Phi)(x) - w \right\| \leq & \left\| \frac{1}{m} S_m(f, \Phi)(x) - \frac{1}{m} S_m(f, \Phi)(f^s(x)) \right\| \\ & + \left\| \frac{1}{m} S_m(f, \Phi)(f^s(x)) - w \right\| \end{aligned}$$

To estimate the first term we note that

$$\begin{aligned} \|S_m(f, \Phi)(x) - S_m(f, \Phi)(f^s(x))\| &\leq \|\Phi(x)\| + \|\Phi(f(x))\| + \cdots + \|\Phi(f^{s-1}(x))\| \\ &\quad + \|\Phi(f^m(x))\| + \cdots + \|\Phi(f^{m+s-1}(x))\| \\ &\leq 2s \cdot \sup\{\|\Phi(x)\| : x \in X\} \\ &\leq 2s \cdot \frac{lr_\mu}{4} \\ &\leq \frac{lnr_\mu}{2} \end{aligned}$$

To estimate the second term we use the fact that $f^s(x) \in \cup_{i=1}^k D_i$ and $m > N_\mu$ implies $\frac{1}{m} S_m(f, \Phi)(f^s(x)) \in B(\frac{r_\mu}{2}, \mathrm{rv}_\Phi(\mu))$. Therefore,

$$\left\| \frac{1}{m} S_m(f, \Phi)(f^s(x)) - w \right\| \leq \frac{r_\mu}{2} + \|w - \mathrm{rv}_\Phi(\mu)\|$$

Combining these two estimates we obtain

$$\begin{aligned} \left\| \frac{1}{m} S_m(f, \Phi)(x) - w \right\| &\leq \frac{1}{m} \cdot \frac{lnr_\mu}{2} + \frac{r_\mu}{2} + \|w - \mathrm{rv}_\Phi(\mu)\| \\ &< r_\mu + \|w - \mathrm{rv}_\Phi(\mu)\| \\ &\leq r \end{aligned}$$

Therefore, $x \in E_n$.

Now we will show that the cardinality of $\mathcal{G}^n$ is comparable to the cardinality of $\mathcal{D}^{nl}$. If a set $B \in \mathcal{B}_i^n$, $(0 < i \leq l)$ then

$$B = D_0 \cap f^{-1}(D_0) \cap ... \cap f^{-(l-i)n+1}(D_0) \cap f^{-(l-i)n}(D_{j_0}) \cap ... \cap f^{-ln+1}(D_{j_{in}})$$

If B is not empty then the set $D_{j_0} \cap ... \cap f^{-in+1}(D_{j_{in}})$ is also not empty. Moreover, different sets $B \in \mathcal{B}_i^n$ correspond to different sets of the form above. By construction of $\mathcal{B}_i^n$ there is a set G in $\mathcal{G}^n$ such that $G \subset D_{j_0} \cap ... \cap f^{-in+1}(D_{j_{in}})$. Therefore,

$$\operatorname{card} \mathcal{G}^n \geq \max_{0 \leq i \leq l} \{\mathcal{B}_i^n\}$$

Since the families $\mathcal{B}_i^n$ are disjoint and $\mathcal{D}^{ln} = \cup_{i=0}^{l} \mathcal{B}_i^n \cup \mathcal{G}^n$, we obtain

$$\operatorname{card} \mathcal{D}^{ln} \leq (l+2) \operatorname{card} \mathcal{G}^n \tag{4}$$

Consider

$$\mathcal{C} = \{D_0 \cup D_1, D_0 \cup D_2, ..., D_0 \cup D_k\}.$$

Since the sets $D_i \, (1 \leq i \leq k)$ are compact, $\mathcal{C}$ is an open cover of X. We denote by $\mathcal{C}_n$ a subfamily of the join $\bigvee_{j=0}^{ln-1} f^{-j}(\mathcal{C})$ which covers E_n and has minimal cardinality. Next we will show that

$$\operatorname{card} \mathcal{G}^n \leq 2^{ln} \operatorname{card} \mathcal{C}_n \tag{5}$$

Let $G \in \mathcal{G}^n$,

$$G = D_{i_1} \cap f^{-1}(D_{i_2}) \cap ... \cap f^{-ln+1}(D_{i_{ln}}).$$

Then $G \subset E_n$ and thus there is a set $C \in \mathcal{C}_n$ such that $G \cap C \neq \varnothing$. The set C is of the form

$$C = (D_0 \cup D_{j_1}) \cap f^{-1}(D_0 \cup D_{j_2}) \cap ... \cap f^{-ln+1}(D_0 \cup D_{j_{ln}}).$$

Since $C \cap G$ is not empty, for $s = 1, ..., ln$ we must have $D_{i_s} \cap (D_0 \cup D_{j_s}) \neq \varnothing$. Since the sets $\{D_i\}_{i=0}^k$ form a partition of X, either $i_s = 0$ or $i_s = j_s$. Also, in this case $G \subset C$. This implies that any set in $\mathcal{G}^n$ is a subset of some set $C \in \mathcal{C}_n$. Moreover, each set in $\mathcal{C}_n$ can contain at most 2^{ln} sets from $\mathcal{G}^n$. We conclude that $\operatorname{card} \mathcal{G}^n \leq 2^{ln} \operatorname{card} \mathcal{C}_n$.

Let δ be a Lebesque number of the cover $\mathcal{C}$, that is any subset of X of diameter less than or equal to δ lies in some member of $\mathcal{C}$. Then δ is also a Lebesque number of the cover $\bigvee_{j=0}^{ln-1} f^{-j}(\mathcal{C})$ in the d_n-metric. Since $\mathcal{C}_n$ is a minimal cover of E_n, every set $C \in \mathcal{C}_n$ contains a point $x_C \in E_n$ which is not in any other element of $\mathcal{C}_n$. Then the ball in the d_{ln}-metric centered at x_C of diameter δ is contained in C. Therefore, points $\{x_C : C \in \mathcal{C}_n\}$ form a $(\delta/2, ln)$-separated set of E_n.

Recall that $F_{ln}(\delta/2, w, r)$ denotes the maximal $(\delta/2, ln)$-separated set with the property that $\frac{1}{ln} S_{ln}(f, \Phi)(x) \in B(r, w)$ for any $x \in F_{ln}(\delta/2, w, r)$. We see that

$$\operatorname{card} \mathcal{C}_n = \operatorname{card} \{x_C : C \in \mathcal{C}_n\} \leq \operatorname{card} F_{ln}(\delta/2, w, r) \tag{6}$$

Combining this last inequality with (4) and (5) we obtain

$$\begin{aligned} \operatorname{card} \mathcal{D}^{ln} &\leq (l+2) \operatorname{card} \mathcal{G}^n \\ &\leq 2^{ln}(l+2) \operatorname{card} \mathcal{C}_n \\ &\leq 2^{ln}(l+2) \operatorname{card} F_{ln}(\delta/2, w, r) \end{aligned}$$

Using the fact that $H_\mu(\mathcal{D}^{ln}) \le \log \operatorname{card} \mathcal{D}^{ln}$ we estimate

$$\begin{aligned} h_\mu(f, \mathcal{D}) &= \lim_{n\to\infty} \frac{1}{ln} H_\mu(\mathcal{D}^{ln}) \\ &\le \lim_{n\to\infty} \frac{1}{ln} \left(ln \log 2 + \log(l+2) + \log \operatorname{card} F_{ln}(\delta/2, w, r) \right) \\ &\le \log 2 + \lim_{n\to\infty} \frac{1}{ln} \log \operatorname{card} F_{ln}(\delta/2, w, r) \end{aligned}$$

We let $\delta \to 0$ in the last inequality and obtain

$$h_\mu(f, \mathcal{D}) \le \log 2 + h(r, w, f, \Phi).$$

Now it is left to compare $h_\mu(f, \mathcal{A})$ and $h_\mu(f, \mathcal{D})$. It is a standard argument to show that $h_\mu(f, \mathcal{A}) \le h_\mu(f, \mathcal{D}) + 1$ (see [16] or [13]). We will outline it here for the sake of completeness. We have ([16, Th 4.12])

$$h_\mu(f, \mathcal{A}) \le h_\mu(f, \mathcal{D}) + H_\mu(\mathcal{A}|\mathcal{D}),$$

where $H_\mu(\mathcal{A}|\mathcal{D})$ is the conditional entropy of $\mathcal{A}$ given $\mathcal{D}$ defined by

$$H_\mu(\mathcal{A}|\mathcal{D}) = -\sum_{i=0}^{k} \sum_{j=1}^{k} \mu(D_i) \frac{\mu(D_i \cap A_j)}{\mu(D_i)} \log \frac{\mu(D_i \cap A_j)}{\mu(D_i)}$$

Since for $i \ne 0$ either $\mu(D_i \cap A_j) = 1$ (when $j = i$) or $\mu(D_i \cap A_j) = 0$ (when $j \ne i$), we get

$$H_\mu(\mathcal{A}|\mathcal{D}) = \mu(D_0) \left[-\sum_{j=1}^{k} \frac{\mu(D_0 \cap A_j)}{\mu(D_0)} \log \frac{\mu(D_0 \cap A_j)}{\mu(D_0)} \right]$$

The expression in the brackets above is the entropy of the cover $\mathcal{A}$ restricted to the set D_0, and thus it is bounded by the $\log \operatorname{card} \mathcal{A} = \log k$. Therefore, $H_\mu(\mathcal{A}|\mathcal{D}) \le \mu(D_0) \log k \le k\varepsilon \log k \le 1$.

We arrive at the inequality $h_\mu(f, \mathcal{A}) \le 1 + \log 2 + h(r, w, f, \Phi)$. Since $\mathcal{A}$ was an arbitrarily chosen Borel partition we obtain $h_\mu(f) \le 1 + \log 2 + h(r, w, f, \Phi)$. □

Now we are ready to prove the main theorem.

Proof of Theorem 1. Since $h_{\mathrm{m}}(w, f, \Phi)$ is approximated by ergodic measures, for any $\varepsilon > 0$ and $r > 0$ there is an ergodic measure $\mu = \mu(\varepsilon, r)$ such that $\mathrm{rv}_\Phi(\mu) \in B(r, w)$ and $|h_\mu(f) - h_{\mathrm{m}}(w, f, \Phi)| < \varepsilon$. Since μ is ergodic, we can apply Lemma 3 and obtain

$$h_\mu(f) \le 1 + \log 2 + h(r, w, f, \Phi).$$

Fix any $k \in \mathbb{N}$. As before, denote $\Phi_k = \frac{1}{k} S_k(f, \Phi)$. Then $\mathrm{rv}_{\Phi_k}(\mu) = \mathrm{rv}_\Phi(\mu)$ and by Lemma 1 we have

$$\frac{1}{n} S_n(f^k, \Phi_k)(x) = \frac{1}{kn} S_{kn}(f, \Phi)(x) \to \mathrm{rv}_{\Phi_k}(\mu) \quad \text{for } \mu\text{-almost all } x.$$

Therefore, measure μ satisfies the assumptions of Lemma 3 for the maps f^k and Φ_k. We obtain

$$h_\mu(f^k) \leq 1 + \log 2 + h(r, w, f^k, \Phi_k)$$

Application of the power rule for measure-theoretic entropy of μ on the left-hand side and Lemma 2 on the right gives

$$kh_\mu(f) \leq 1 + \log 2 + kh(r, w, f, \Phi)$$

Since $k \in \mathbb{N}$ was arbitrary, we obtain $h_\mu(f) \leq h(r, w, f, \Phi)$. By the choice of measure μ we have $h_{\rm m}(w, f, \Phi) - \varepsilon \leq h(r, w, f, \Phi)$. Finally, letting ε and r approach 0 we obtain the desired inequality

$$h_{\rm m}(w, f, \Phi) \leq h_{\rm top}(w, f, \Phi).$$

This completes the proof of the theorem. □

References

[1] R. Bowen *Entropy expansive maps*, Trans. Am. Math. Soc. **164** (1972), 323–331.

[2] H. Federer, *Geometric measure theory*, Springer-Verlag Berlin Heidelberg New York, 1996.

[3] De-Jung Feng and Wen Huang, *Variational principle for topological entropies of subsets*, J. Funct. Anal. **263** (2012) 2228–2254.

[4] T. Fisher, L. Diaz, M. Pacifico and J. Vieitez, *Entropy-expansiveness for partially hyperbolic diffeomorphisms*, Discrete Contin. Dyn. Syst. **32** (2012), 4195–4207.

[5] W. Geller and M. Misiurewicz, *R*otation and entropy, Trans. American Math. Soc. **351** (1999), 2927–2948.

[6] O. Jenkinson, *Rotation, entropy, and equilibrium states*, Trans. American Math. Soc. **353** (2001), 3713–3739.

[7] A. Kornyi and H. M. Reimann, *Foundations for the theory of quasiconformal mappings on the Heisenberg group*, Adv. Math. **111(1)** (1995), 1–87.

[8] T. Kucherenko and C. Wolf, *Geometry and entropy of generalized rotation sets*, Israel Journal of Mathematics **199** (2014), 791–829.

[9] T. Kucherenko and C. Wolf, *Localized Pressure and equilibrium states*, submitted (2013).

[10] P. Loeb, *On the Besicovitch covering theorem*, SUT J. Math. **25** (1989), 51–55.

[11] S. Newhouse, *Continuity properties of entropy*, Ann. of Math. (2) **129** (1989), 215–235.

[12] D. Preiss, *Dimension of metrics and differentiation of measures* In General topology and its relations to modern analysis and algebra V (Prague 1981) Sigma Ser. Pure Math. 3 pp. 565568 Heldermann, Berlin 1983.

[13] F. Przytycki and M. Urbanski, *Conformal fractals: ergodic theory methods*, London Mathematical Society Lecture Note Series, 371. Cambridge University Press, Cambridge, 2010. x+354 pp.

[14] S. Rigot *Counter example to the Besicovitch covering property for some Carnot groups equipped with their Carnot-Carathéodory metric*, Mathematische Zeitschrift **248(4)** (2004), 827–848.

[15] E. Sawyer and R. L. Wheeden, *Weighted inequalities for fractional integrals on Euclidean and homogeneous spaces*, Amer. J. Math. **114(4)** (1992), 813–874.

[16] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics 79, Springer, 1981.

[17] K. Ziemian, *Rotation sets for subshifts of finite type*, Fund. Math. **146** (1995), 189–201.

Department of Mathematics, The City College of New York, New York, NY, 10031, USA
E-mail address: tkucherenko@ccny.cuny.edu