On the variational principle for the topological entropy of certain non-compact sets
Takens, Floris; Verbitskiy, Evgeny

Published in:
Ergodic Theory and Dynamical Systems

DOI:
10.1017/S0143385702000913

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2003

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
On the variational principle for the topological entropy of certain non-compact sets

FLORIS TAKENS† and EVGENY VERBITSKIY‡
† Department of Mathematics, University of Groningen PO Box 800, 9700 AV, Groningen, The Netherlands
(e-mail: f.takens@math.rug.nl)
‡ Eurandom, Technical University of Eindhoven, PO Box 513, 5600 MB, Eindhoven, The Netherlands,
(e-mail: e.verbitskiy@tue.nl)

(Received 16 April 2001 and accepted in revised form 14 December 2001)

Abstract. For a continuous transformation $f$ of a compact metric space $(X, d)$ and any continuous function $\varphi$ on $X$ we consider sets of the form

$$K_\alpha = \left\{ x \in X : \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)) = \alpha \right\}, \quad \alpha \in \mathbb{R}.$$ 

For transformations satisfying the specification property we prove the following Variational Principle

$$h_{\text{top}}(f, K_\alpha) = \sup \left( h_\mu(f) : \mu \text{ is invariant and } \int \varphi \, d\mu = \alpha \right),$$

where $h_{\text{top}}(f, \cdot)$ is the topological entropy of non-compact sets. Using this result we are able to obtain a complete description of the multifractal spectrum for Lyapunov exponents of the so-called Manneville–Pomeau map, which is an interval map with an indifferent fixed point.

We also consider multi-dimensional multifractal spectra and establish a contraction principle.

1. Introduction

Often the problems of multifractal analysis of local (or pointwise) dimensions and entropies are reduced to consideration of the sets of the following form

$$K_\alpha = \left\{ x \in X : \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)) = \alpha \right\}, \quad \alpha \in \mathbb{R},$$
where $f : X \to X$ is some transformation and $\varphi : X \to \mathbb{R}$ is a function, sometimes called an observable. Typically, $f$ is a continuous transformation of some compact metric space $(X,d)$ and $\varphi$ is sufficiently smooth.

In particular, one is interested in the ‘size’ of these sets $K_\alpha$. The following characteristics of the sets $K_\alpha$ have been studied in the literature:

$$D_\varphi(\alpha) = \dim_H(K_\alpha), \quad E_\varphi(\alpha) = h_{\text{top}}(f, K_\alpha),$$

where $\dim_H(K_\alpha)$ and $h_{\text{top}}(f, K_\alpha)$ are the Hausdorff dimension and the topological entropy of $K_\alpha$, respectively. The precise definition of the topological entropy of non-compact sets will be given in §3, but for now the topological entropy should be viewed as a dimension-like characteristic, similar to the Hausdorff dimension. The functions $D_\varphi(\alpha)$ and $E_\varphi(\alpha)$ will be called the dimension and entropy multifractal spectra of $\varphi$.

Recently similar problems were considered in the relation with a definition of a rotational entropy $[7, 9]$.

Multifractal analysis studies various properties of the multifractal spectra $D_\varphi(\alpha)$, $E_\varphi(\alpha)$ as functions of $\alpha$, for example, smoothness and convexity, and relates these spectra to other characteristics of a dynamical system. In order to obtain non-trivial results one typically has to make two types of assumptions: firstly, on the dynamical system $(X,f)$, and secondly, on the properties of the observable function $\varphi$. For example,

- ([15], see also [16]) if $f$ is a sufficiently smooth expanding conformal map, and $\varphi$ is a Hölder continuous function, then $E_\varphi(\alpha)$ is real-analytic and concave;
- ([21]) if $f$ is an expansive homeomorphism with specification, and $\varphi$ has bounded variation, then $E_\varphi(\alpha)$ is $C^1$ and concave.

In both cases, $E_\varphi(\alpha)$ is a Legendre transform of a pressure function $P_\varphi(q) = P(q\varphi)$, where $P(\cdot)$ is the topological pressure.

The conditions on $\varphi$ in the examples above are meant to ensure the absence of phase transition, i.e. the existence and uniqueness of the equilibrium state for the potential $q\varphi$ for every $q \in \mathbb{R}$. The main goal of this paper is to relax such conditions and to obtain results for systems exhibiting phase transitions.

A natural class of observable functions $\varphi$ would be the set of all continuous functions. Moreover, the set of all continuous functions is quite rich in the sense of possible phase transitions. For example [20, p. 52], for any set $\{\mu_1, \ldots, \mu_k\}$ of ergodic shift-invariant measures on $A^\mathbb{Z}$, where $A$ is a finite set, one can find a continuous function $\varphi$ such that all these measures $\mu_i$, $i = 1, \ldots, k$, are equilibrium states for $\varphi$. Nevertheless, Fan and Feng in [6] and Olivier in [14], in the case of symbolic dynamics, obtained results on the spectrum $E_\varphi(\alpha)$ for arbitrary continuous functions $\varphi$, similar to those mentioned above.

In fact, they were studying the dimension spectrum $D_\varphi(\alpha)$, but in the symbolic case for every $\alpha$ one has $E_\varphi(\alpha) = \#(A)D_\varphi(\alpha)$, where $\#(A)$ is the number of elements in $A$.

In this paper we study the entropy spectrum $E_\varphi(\alpha)$ for a continuous transformation $f$ on a compact metric space $(X,d)$ and an arbitrary continuous function $\varphi$. The main result of this paper (Theorem 5.1) states that if $f$ is a continuous transformation with specification property, then for any $\alpha$ with $K_\alpha \neq \emptyset$ one has

$$E_\varphi(\alpha) = H_\varphi(\alpha) = \Lambda_\varphi(\alpha),$$
where
\[ H_\varphi(\alpha) := \sup \left\{ h_\mu(f) : \mu \text{ is invariant and } \int \varphi \, d\mu = \alpha \right\}, \]
and \( \Lambda_\varphi(\alpha) \) is a special ‘ball’-counting dimension of \( K_\alpha \), similar to ones introduced in [6, 11].

Readers familiar with large deviations will recognize in \( H_\varphi(\alpha) \) the so-called rate function. Indeed, we use the large deviation results for dynamical systems with specification obtained by Young in [25].

The most intricate part of our proof is the equality \( E_\varphi(\alpha) = \gamma_{\text{Lambda}}(\sigma, \varphi)(\alpha) \). To show it we use a Moran fractal structure, inspired by one constructed in [6] for the symbolic case.

The Manneville–Pomeau map is a piecewise continuous map of a unit interval given by
\[ f_s(x) = x + x^s \mod 1, \quad 0 < s < 1. \]
This map has a unique indifferent fixed point \( x = 0 \), and is probably the simplest example of a non-uniformly hyperbolic dynamical system. Thermodynamic properties of this transformation are quite well understood, see [12, 13, 19, 22].

In [18], Pollicott and Weiss studied the multifractal spectrum for \( \varphi = \log f_s' \), i.e. the spectrum of Lyapunov exponents. They were able to obtain a partial description of this spectrum. Using our results we able to complete the picture, see §6 for details.

A straightforward modification of our proofs shows that the results are also valid in more general settings. Suppose \( f : X \to X \) is a continuous transformation with the specification property and \( \varphi = (\varphi_1, \ldots, \varphi_d) : X \to \mathbb{R}^d \) is a continuous function. For \( \alpha \in \mathbb{R}^d \) consider the set
\[ K_\alpha = \left\{ x \in X : \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi_j(f^i(x)) = \alpha_j, \ j = 1, \ldots, d \right\}. \]
Then
\[ E_\varphi(\alpha) = h_{\text{top}}(f, K_\alpha) = \sup \left\{ h_\mu(f) : \mu \text{ is invariant and } \int \varphi \, d\mu = \alpha \right\}. \tag{1} \]
In fact, even more is true. Suppose again that \( \varphi : X \to \mathbb{R}^d \) is a continuous function and \( \Psi : \text{Im}(\varphi) \to \mathbb{R}^m \) is a continuous map defined on \( \text{Im}(\varphi) = \{ \varphi(x) : x \in X \} \subseteq \mathbb{R}^d \). Define
\[ K_\beta^{\Psi} = \left\{ x \in X : \lim_{n \to \infty} \Psi \left( \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)) \right) = \beta \right\}. \]
Then, under an additional assumption that the entropy map is upper semi-continuous, for any \( \beta \) such that \( K_\beta^{\Psi} \neq \emptyset \) one has
\[ E_{\Psi}(\beta) = h_{\text{top}}(f, K_\beta^{\Psi}) = \sup \left\{ h_\mu(f) : \mu \text{ is invariant and } \Psi \left( \int \varphi \, d\mu \right) = \beta \right\}. \tag{2} \]

As an immediate consequence of (1) and (2) we obtain the following result, which we call the Contraction Principle for Multifractal Spectra, due to the clear analogy with the well-known Contraction Principle in Large Deviations:
\[ E_{\Psi}(\beta) = \sup_{\alpha : \Psi(\alpha) = \beta} E_\varphi(\alpha). \]
For more detail see §7.
Everywhere in the present paper \(#(C)\) denotes the cardinality of a set \(C\). Proofs of all lemmas are collected in §8.

2. **Multifractal spectrum of continuous functions**

Let \(f : X \to X\) be a continuous transformation of a compact metric space \((X, d)\). Throughout this paper we will assume that \(f\) has finite topological entropy. Suppose \(\varphi : X \to \mathbb{R}\) is a continuous function. For \(\alpha \in \mathbb{R}\) define

\[
K_\alpha = \left\{ x \in X : \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)) = \alpha \right\}.
\]  

We introduce the following notation

\[
\mathcal{L}_\varphi = \{ \alpha \in \mathbb{R} : K_\alpha \neq \emptyset \}.
\]

**Lemma 2.1.** The set \(\mathcal{L}_\varphi\) is a non-empty bounded subset of \(\mathbb{R}\).

**Definition 2.1.** A continuous transformation \(f : X \to X\) satisfies the specification property if for any \(\varepsilon > 0\) there exists an integer \(m = m(\varepsilon)\) such that for arbitrary finite intervals \(I_j = [a_j, b_j] \subseteq \mathbb{N}, j = 1, \ldots, k\), such that

\[
\text{dist}(I_i, I_j) \geq m(\varepsilon), \quad i \neq j,
\]

and any \(x_1, \ldots, x_k\) in \(X\) there exists a point \(x \in X\) such that

\[
d(f^{p+a_j}x, f^p x_j) < \varepsilon \quad \text{for all} \quad p = 0, \ldots, b_j - a_j \text{ and every } j = 1, \ldots, k.
\]

Following the present day tradition we do not require that \(x\) is periodic. Specification implies topological mixing. Moreover, by the Blokh Theorem [2], for continuous transformations of the interval these two conditions are equivalent. Using this equivalence and the results of Jakobson [8], we conclude that for the logistic family \(f_\varepsilon(x) = \varepsilon x(1 - x)\) the specification property holds for a set of parameters of positive Lebesgue measure.

The specification property allows us to connect together arbitrary pieces of orbits. Suppose now that for two values \(\alpha_1, \alpha_2\) the corresponding sets \(K_{\alpha_1}, K_{\alpha_2}\) are not empty. Using the specification property we are able to construct points with ergodic averages, converging to any number \(\alpha \in (\alpha_1, \alpha_2)\). Hence, \(\mathcal{L}_\varphi\) is a convex set. This implies the following.

**Lemma 2.2.** If \(f : X \to X\) satisfies the specification property, then \(\mathcal{L}_\varphi\) is an interval.

We recall that the entropy spectrum \(\mathcal{E}_\varphi(\cdot)\) of \(\varphi\) is the map assigning to each \(\alpha \in \mathcal{L}_\varphi\) the value

\[
\mathcal{E}_\varphi(\alpha) = h_{\text{top}}(f, K_\alpha).
\]  

The definition and some fundamental facts about the topological entropy of non-compact sets are collected in the following section.
3. Topological entropy of non-compact sets

The generalization of the topological entropy to non-compact or non-invariant sets goes back to Bowen [3]. Later Pesin and Pitskel [17] generalized the notion of the topological pressure to the case of non-compact sets. In this paper we use an equivalent definition of the topological entropy, which can be found in [16].

3.1. Definition of the topological entropy. Once again, let \((X, d)\) be a compact metric space and \(f : X \to X\) be a continuous transformation. For any \(n \in \mathbb{N}\) we define a new metric \(d_n\) on \(X\)

\[
d_n(x, y) = \max\{d(f^k(x), f^k(y)) : k = 0, \ldots, n - 1\},
\]

and for every \(\varepsilon > 0\) we denote by \(B_n(x, \varepsilon)\) an open ball of radius \(\varepsilon\) in the metric \(d_n\) around \(x\), i.e.

\[
B_n(x, \varepsilon) = \{y \in X : d_n(x, y) < \varepsilon\}.
\]

We define a function \(m(Z, s, N, \varepsilon)\) as

\[
m(Z, s, N, \varepsilon) = \inf_{\Gamma} \sum_i \exp(-sn_i),
\]

where the infimum is taken over all collections \(\Gamma = \{B_n(x_i, \varepsilon)\}\) covering \(Z\), such that \(n(\Gamma) \geq N\). The quantity \(m(Z, s, N, \varepsilon)\) does not decrease with \(N\), hence the following limit exists:

\[
m(Z, s, \varepsilon) = \lim_{N \to \infty} m(Z, s, N, \varepsilon) = \sup_{N>0} m(Z, s, N, \varepsilon).
\]

It is easy to show that there exists a critical value of the parameter \(s\), which we will denote by \(h_{\text{top}}(f, Z, \varepsilon)\), where \(m(Z, s, \varepsilon)\) jumps from \(+\infty\) to 0, i.e.

\[
m(Z, s, \varepsilon) = \begin{cases} +\infty, & s < h_{\text{top}}(f, Z, \varepsilon), \\ 0, & s > h_{\text{top}}(f, Z, \varepsilon). \end{cases}
\]

There are no restrictions on the value \(m(Z, s, \varepsilon)\) for \(s = h_{\text{top}}(f, Z, \varepsilon)\). It can be infinite, zero, or positive and finite. One can show [16] that the following limit exists

\[
h_{\text{top}}(f, Z) = \lim_{\varepsilon \to 0} h_{\text{top}}(f, Z, \varepsilon).
\]

We will call \(h_{\text{top}}(f, Z)\) the topological entropy of \(f\) restricted to \(Z\) or, simply, the topological entropy of \(Z\), when there is no confusion about \(f\).

3.2. Properties of the topological entropy. Here we recall some of the basic properties and important results on the topological entropy of non-compact or non-invariant sets.

**Theorem 3.1.** [16] The topological entropy as defined above satisfies the following:

1. \(h_{\text{top}}(f, Z_1) \leq h_{\text{top}}(f, Z_2)\) for any \(Z_1 \subseteq Z_2 \subseteq X\);
2. \(h_{\text{top}}(f, Z) = \sup_i h_{\text{top}}(f, Z_i),\) where \(Z = \bigcup_{i=1}^{\infty} Z_i \subseteq X\).
The next theorem establishes a relation between the topological entropy of a set and the measure-theoretic entropies of measures, concentrated on this set, generalizing the classical result for compact sets.

**Theorem 3.2.** (Bowen [3]) Let \( f : X \to X \) be a continuous transformation of a compact metric space. Suppose \( \mu \) is an invariant measure and \( Z \subseteq X \) is such that \( \mu(Z) = 1 \), then

\[
h_{\text{top}}(f, Z) \geq h_\mu(f),
\]

where \( h_\mu(f) \) is the measure-theoretic entropy.

Suppose we are given an invariant measure \( \mu \). A point \( x \) is called generic for \( \mu \) if the sequence of probability measures

\[
\delta_{x,n} = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{f^k(x)},
\]

where \( \delta_y \) is the Dirac measure at \( y \), converges to \( \mu \) in the weak topology. Denote by \( G_\mu \) the set of all generic points for \( \mu \). If \( \mu \) is an ergodic invariant measure, then by the Ergodic Theorem \( \mu(G_\mu) = 1 \). Applying the previous theorem we immediately conclude that \( h_{\text{top}}(f, G_\mu) \geq h_\mu(f) \). In fact, the opposite inequality is valid as well.

**Theorem 3.3.** (Bowen [3]) Let \( \mu \) be an ergodic invariant measure, then

\[
h_{\text{top}}(f, G_\mu) = h_\mu(f).
\]

Pesin and Pitskel in [17] have proved the following variational principle for non-compact sets.

**Theorem 3.4.** Suppose \( f : X \to X \) is a continuous transformation of a compact metric space \( (X, d) \) and \( Z \subseteq X \) is an invariant set. Denote by \( \mathcal{M}_f(Z) \) the set of all invariant measures \( \mu \) such that \( \mu(Z) = 1 \). For any \( x \in X \) denote by \( V(x) \) the set of all limit points of the sequence \( \{\delta_{x,n}\} \). Assume that for every \( x \in Z \) one has

\[
V(x) \cap \mathcal{M}_f(Z) \neq \emptyset.
\]

Then \( h_{\text{top}}(f, Z) = \sup_{\mu \in \mathcal{M}_f(Z)} h_\mu(f) \).

The conditions of this theorem are very difficult to check in any specific situation. However, there is no hope for improving the above result for general sets \( Z \). There are examples [16, 17] of sets for which the condition \( V(x) \cap \mathcal{M}_f(Z) \neq \emptyset \) does not hold for all \( x \in Z \), and one has a strict inequality

\[
h_{\text{top}}(f, Z) > \sup\{h_\mu(f) : \mu \in \mathcal{M}_f(X) \text{ and } \mu(Z) = 1\}.
\]

In this paper we restrict our attention to the sets of a special form: namely, the sets \( K_\alpha \) given by (3). For these particular sets we prove a variational principle for the topological entropy, provided the transformation \( f \) satisfies the specification property.
THEOREM 3.5. Suppose $f : X \to X$ is a continuous transformation with the specification property. Let $\varphi \in C(X, \mathbb{R})$ and assume that for some $\alpha \in \mathbb{R}$

$$K_\alpha = \left\{ x \in X : \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)) = \alpha \right\} \neq \emptyset,$$

then

$$h_{\text{top}}(f, K_\alpha) = \sup \left\{ h_\mu(f) : \mu \text{ is invariant and } \int \varphi \, d\mu = \alpha \right\}.$$

This result is a corollary of Theorem 5.1, which we establish below.

Remark 3.1. Under the conditions of the above theorem, it is possible that for a certain parameter value $\alpha$ there exists a unique invariant probability measure $\mu_\alpha$ with $\int \varphi \, d\mu = \alpha$, such that

$$h_{\text{top}}(f, K_\alpha) = h_{\mu_\alpha}(f).$$

Hence, $\mu_\alpha$ is a measure of maximal entropy among all invariant measures $\mu$ with $\int \varphi \, d\mu = \alpha$. However, it is also possible, that $\mu_\alpha(K_\alpha) = 0$. This situation, for example, occurs in the family of Manneville–Pomeau maps, see §6 for more details.

3.3. Entropy Distribution Principle. The following statement will allow us to estimate the topological entropies of the sets from below, without constructing probability measures, which are invariant and concentrated on a given set. It is sufficient to consider probability measures, which are not necessarily invariant, but which satisfy some specific 'uniformity condition'. We call this result the Entropy Distribution Principle, by the clear analogy with the well-known Mass Distribution Principle [5].

THEOREM 3.6. (Entropy Distribution Principle) Let $f : X \to X$ be a continuous transformation. Suppose a set $Z \subseteq X$ and a constant $s \geq 0$ are such that for any sufficiently small $\varepsilon > 0$ one can find a Borel probability measure $\mu = \mu_\varepsilon$ satisfying:

1. $\mu_\varepsilon(Z) > 0$;
2. $\mu_\varepsilon(B_n(x, \varepsilon)) \leq C(\varepsilon)e^{-ns}$ for some constant $C(\varepsilon) > 0$ and every ball $B_n(x, \varepsilon)$ such that $B_n(x, \varepsilon) \cap Z \neq \emptyset$.

Then $h_{\text{top}}(f, Z) \geq s$.

Proof. We are going to show that $h_{\text{top}}(f, Z, \varepsilon) \geq s$ for every sufficiently small $\varepsilon > 0$. Indeed, choose $\varepsilon > 0$ such that conditions (1) and (2) hold. Let $\Gamma = \{B_n(x_i, \varepsilon)\}$ be some cover of $Z$. Without loss of generality, we may assume that $B_n(x_i, \varepsilon) \cap Z \neq \emptyset$ for every $i$. Then

$$\sum_i \exp(-s \epsilon_i) \geq C(\varepsilon)^{-1} \sum_i \mu_\varepsilon(B_n(x_i, \varepsilon)) \geq C(\varepsilon)^{-1} \mu_\varepsilon \left( \bigcup_i B_n(x_i, \varepsilon) \right) \geq C(\varepsilon)^{-1} \mu_\varepsilon(Z) > 0.$$

Therefore, $m(Z, s, \varepsilon) > 0$ and hence $h_{\text{top}}(f, Z, \varepsilon) \geq s$. □
4. **Upper estimates of $E_{\phi}(\alpha)$**

In this section we are going to define two auxiliary quantities $H_{\phi}(\alpha)$ and $\Lambda_{\phi}(\alpha)$. These quantities will be used to give an upper estimate on the multifractal spectrum $E_{\phi}(\alpha)$.

4.1. **Definition of $H_{\phi}(\alpha)$**. Let us introduce some notation:

- $M(X)$: the set of all Borel probability measures on $X$,
- $M_f(X)$: the set of all $f$-invariant Borel probability measures on $X$,
- $M^e_f(X)$: the set of all ergodic $f$-invariant Borel probability measures on $X$,
- $M_f(X, \phi, \alpha)$: the set of all $f$-invariant Borel probability measures, such that $\int \phi d\mu = \alpha$.

We consider the weak topology on $M(X)$ and also on its subsets $M_f(X)$, $M^e_f(X)$, etc.; as is well known, $M(X)$ is compact metrizable space in the weak topology.

**Lemma 4.1.** For any $\alpha \in L_{\phi}$ the set $M_f(X, \phi, \alpha)$ is a non-empty, convex and closed (in the weak topology) subset of $M_f(X)$.

This result allows us to define the following quantity: for any $\alpha \in L_{\phi}$ put

\[
H_{\phi}(\alpha) = \sup \{ h_\mu(f) : \mu \in M_f(X, \phi, \alpha) \}. \tag{5}
\]

**Lemma 4.2.** For any $\varphi \in C(X, \mathbb{R})$, $H_{\phi}(\alpha)$ is a concave function on the interval of $(\inf L_{\phi}, \sup L_{\phi})$.

4.2. **Definition of $\Lambda_{\phi}(\alpha)$**. Here, following the approach of [6, 11], we introduce another dimension-like characteristic $\Lambda_{\phi}(\alpha)$ of the set $K_{\alpha}$. We use the word ‘dimension’ in association with $\Lambda_{\phi}(\alpha)$, because $\Lambda_{\phi}(\alpha)$ is defined in terms similar to the definition of Hausdorff or box-counting dimensions.

For $\alpha \in L_{\phi}$, any $\delta > 0$ and $n \in \mathbb{N}$ put

\[
P(\alpha, \delta, n) = \left\{ x \in X : \left| \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)) - \alpha \right| < \delta \right\}.
\]

Clearly, for $\alpha \in L_{\phi}$ and any $\delta > 0$ the set $P(\alpha, \delta, n)$ is not empty for sufficiently large $n$.

Fix some $\varepsilon > 0$ and let $N(\alpha, \delta, n, \varepsilon)$ be the minimal number of balls $B_{\varepsilon}(x, \varepsilon)$, which is necessary for covering the set $P(\alpha, \delta, n)$. (If $P(\alpha, \delta, n)$ is empty we let $N(\alpha, \delta, n, \varepsilon) = 1$.)

Obviously, $N(\alpha, \delta, n, \varepsilon)$ does not increase if $\delta$ decreases, and $N(\alpha, \delta, n, \varepsilon)$ does not decrease if $\varepsilon$ decreases. This observation guarantees that the following limit exists:

\[
\Lambda_{\phi}(\alpha) = \lim_{\varepsilon \to 0} \lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{n} \log N(\alpha, \delta, n, \varepsilon). \tag{6}
\]

One can give another equivalent definition of $\Lambda_{\phi}(\alpha)$. The equivalence of these definitions will be useful for subsequent arguments. Let us recall a notion of $(n, \varepsilon)$-separated sets: a set $E$ is called $(n, \varepsilon)$-separated if for any $x, y \in E, x \neq y$, $d_n(x, y) > \varepsilon$. 


By definition, we let \( M(\alpha, \delta, n, \varepsilon) \) be the cardinality of a maximal \((n, \varepsilon)\)-separated set in \( P(\alpha, \delta, n) \). Again, we put \( M(\alpha, \delta, n, \varepsilon) = 1 \) if \( P(\alpha, \delta, n) \) is empty. A standard argument shows that

\[
N(\alpha, \delta, n, \varepsilon) \leq M(\alpha, \delta, n, \varepsilon) \leq N(\alpha, \delta, n, \varepsilon/2)
\]

for every \( n \in \mathbb{N} \) and all \( \varepsilon, \delta > 0 \).

Moreover, if \( f \) satisfies specification, then taking an upper limit instead of the lower limit with respect to \( n \) in the definition of \( \varLambda(\alpha) \) will give the same number.

**Lemma 4.3.** If \( f \) satisfies specification, then

\[
\varLambda(\alpha) = \lim_{\varepsilon \to 0} \lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{n} \log N(\alpha, \delta, n, \varepsilon).
\]

We will not use this result, and therefore will not give a proof, which is based on establishing some sort of subadditivity of \( N(\alpha, \delta, n, \varepsilon) \):

\[
(N(\alpha, \delta, n, 4\varepsilon))^k \leq N(\alpha, 4\delta, nk + km(\varepsilon), \varepsilon)
\]

for all integers \( k \geq 1 \) and all sufficiently large \( n \), where \( m \) is taken from the definition of the specification property.

4.3. **Upper estimate for \( E_\varphi(\alpha) \) in terms of \( H_\varphi(\alpha) \) via \( \varLambda(\alpha) \).**

**Theorem 4.1.** For any \( \alpha \in \mathcal{L}_\varphi \) one has

\[
E_\varphi(\alpha) \leq \varLambda(\alpha) \leq H_\varphi(\alpha).
\]

**Proof.** The first inequality \( E_\varphi(\alpha) \leq \varLambda(\alpha) \) is quite easy. Its proof is based on a standard ‘box-counting’ argument. Following [6], for \( \alpha \in \mathcal{L}_\varphi \), \( \delta > 0 \) and \( k \in \mathbb{N} \) consider sets

\[
G(\alpha, \delta, k) = \bigcap_{n=k}^{\infty} P(\alpha, \delta, n) = \bigcap_{n=k}^{\infty} \left\{ x \in X : \left| \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)) - \alpha \right| < \delta \right\}.
\]

It is clear that for any \( \delta > 0 \)

\[
K_\alpha = \left\{ x \in X : \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)) = \alpha \right\} \subseteq \bigcup_{k=1}^{\infty} G(\alpha, \delta, k).
\]

We are going to show that \( h_{\text{top}}(f, G(\alpha, \delta, k), \varepsilon) \leq \varLambda(\alpha) \) holds for any \( k \geq 1 \), implying \( h_{\text{top}}(f, K_\alpha, \varepsilon) \leq \varLambda(\alpha) \) as well.

Fix arbitrary \( k \geq 1 \), then \( G(\alpha, \delta, k) \) (as a subset of \( P(\alpha, \delta, n) \) for \( n \geq k \)) can be covered by \( N(\alpha, \delta, n, \varepsilon) \) balls \( B(x, \varepsilon) \) for all \( n \geq k \). Therefore, for every \( s > 0 \) and all \( n \geq k \) we have

\[
m(G(\alpha, \delta, k), s, \varepsilon) \leq N(\alpha, \delta, n, \varepsilon) \exp(-ns).\]

Suppose now that \( s > \varLambda(\alpha) \) and put \( \gamma = (s - \varLambda(\alpha))/2 > 0 \). Since

\[
\varLambda(\alpha) = \lim_{\varepsilon \to 0} \lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{n} \log N(\alpha, \delta, n, \varepsilon),
\]
for all sufficiently small $\varepsilon > 0$ and $\delta > 0$, there exists a monotonic sequence of integers $n_l \to \infty$ such that

$$N(\alpha, \delta, n_l, \varepsilon) \leq \exp(n_l(\Lambda_\varphi(\alpha) + \gamma))$$

for all $l \geq 1$. Without loss of generality, we may assume that $n_1 \geq k$. Then, from (9) we obtain

$$m(G(\alpha, \delta, k), s, \varepsilon) \leq \exp(-n_l \gamma),$$

and hence $m(G(\alpha, \delta, k), s, \varepsilon) = 0$. Therefore, $h_{\text{top}}(f, G(\alpha, \delta, k), \varepsilon) \leq s$ and

$$h_{\text{top}}(f, K_\alpha, \varepsilon) \leq \sup_k h_{\text{top}}(f, G(\alpha, \delta, k), \varepsilon) \leq s$$

due to (8). Therefore, $h_{\text{top}}(f, K_\alpha) = \lim_{\varepsilon \to 0} h_{\text{top}}(f, K_\alpha, \varepsilon) \leq s$ as well. Finally, since $s > \Lambda_\varphi(\alpha)$ was chosen arbitrary, we conclude that $E_\varphi(\alpha) := h_{\text{top}}(f, K_\alpha) \leq \gamma \Lambda_\sigma(\varphi(\alpha))$.

The second inequality $\Lambda_\varphi(\alpha) \leq H_\varphi(\alpha)$ is closely related to the second statement of Theorem 1 by Young in [25], and is in fact a large deviation result. In the last stage of our proof, similar to [25], we will rely on one fact, which is established in a standard proof of the Variational Principle for the classical topological entropy [23].

In order to show the inequality $\Lambda_\varphi(\alpha) \leq H_\varphi(\alpha)$, it is sufficient, for any $\gamma > 0$, to present a measure $\mu \in \mathcal{M}_f(X, \varphi, \alpha)$ (i.e. an invariant measure with $\int \varphi d\mu = \alpha$) such that

$$h_\mu(f) \geq \Lambda_\varphi(\alpha) - \gamma.$$

Fix arbitrary $\gamma > 0$. By the definition of $\Lambda_\varphi(\alpha)$, there exists a sufficiently small $\varepsilon_0 > 0$ such that for all $\varepsilon \in (0, \varepsilon_0)$ one has

$$\Lambda_\varphi(\alpha, \varepsilon) = \lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{n} \log N(\alpha, \delta, n, \varepsilon) > \Lambda_\varphi(\alpha) - \frac{1}{3} \gamma.$$

Put $\varepsilon_k = \varepsilon_0/2^k$, $k \geq 1$. For any $k \geq 1$ one can find a sufficiently small $\delta_3, \delta_k \to 0$, such that

$$\lim_{n \to \infty} \frac{1}{n} \log N(\alpha, \delta_k, n, \delta_k) > \Lambda_\varphi(\alpha) - \frac{2}{3} \gamma.$$

Also, for any $k \geq 1$ we choose some $n_k \in \mathbb{N}, n_k \to \infty$, such that

$$N_k := N(\alpha, \delta_k, n_k, \delta_k) > \exp(n_k(\Lambda_\varphi(\alpha) - \gamma)).$$

Let $C_k$ be the centers of some minimal covering of $P(\alpha, \delta_k, n_k)$ by balls $B_{n_k}(x, \delta_k)$. Note, that $\#(C_k) = N_k$ and $B_{n_k}(x, \delta_k) \cap P(\alpha, \delta_k, n_k) \neq \emptyset$ for every $x \in C_k$. Otherwise, the covering would not be minimal. For any $k \geq 1$ define a probability measure

$$\sigma_k = \frac{1}{N_k} \sum_{x \in C_k} \delta_x,$$

and let

$$\mu_k = \frac{1}{n_k} \sum_{i=0}^{n_k-1} (f^{-i})^* \sigma_k = \frac{1}{N_k} \sum_{x \in C_k} \frac{1}{n_k} \sum_{i=0}^{n_k-1} \delta_{f^i(x)}.$$
On the variational principle for topological entropy 327

Let $\mu$ be some limit point for the sequence $\mu_k$. By [23, Theorem 6.9], $\mu$ is an invariant measure, and we claim that

$$\int \varphi \, d\mu = \alpha,$$

i.e. $\mu \in M_f(X, \varphi, \alpha)$. Indeed, for every $k \geq 1$, one has

$$\left| \int \varphi \, d\mu_k - \alpha \right| \leq \frac{1}{N_k} \sum_{x \in C_k} \frac{1}{n_k} \sum_{i=0}^{n_k-1} \varphi(f^i(x)) - \alpha \right|.

However, for every $x \in C_k$ there exists $y = y(x) \in P(\alpha, \delta_k, n_k)$ such that $d_{n_k}(x, y) < \varepsilon_k$. Therefore,

$$\left| \frac{1}{n_k} \sum_{i=0}^{n_k-1} \varphi(f^i(x)) - \alpha \right| \leq \frac{1}{n_k} \sum_{i=0}^{n_k-1} \varphi(f^i(x)) - \varphi(f^i(y))| + \delta_k \leq \text{Var}(\varphi, \varepsilon_k) + \delta_k,

where $\text{Var}(\varphi, \varepsilon_k) = \sup(|\varphi(x) - \varphi(y)| : d(x, y) < \varepsilon_k) \to 0$ as $k \to \infty$, since $\varphi$ is continuous. Hence, we conclude that

$$\int \varphi \, d\mu_k \to \alpha, \quad k \to \infty.$$

The above invariant measure $\mu$ is a limit point for the sequence $\mu_k$. Hence, there exists a sequence $k_j \to \infty$ such that $\mu_{k_j} \to \mu$ weakly. This in particular means that

$$\int \varphi \, d\mu_{k_j} \to \int \varphi \, d\mu.$$

Therefore, we obtain (10). Finally, repeating the second half of the proof of the Classical Variational Principle [23, Theorem 8.6, pp. 189–190] we conclude that

$$h_\mu(f) \geq \lim_{k \to \infty} \frac{1}{n_k} \log N_k \geq \lim_{k \to \infty} \frac{1}{n_k} \log N_k \geq \Lambda_\varphi(\alpha) - \gamma.

This finishes the proof.

5. Lower estimate on $E_\varphi(\alpha)$

The main result of this section is the following theorem.

THEOREM 5.1. Let $f : X \to X$ be a continuous transformation with the specification property and $\varphi \in C(X, \mathbb{R})$. Then for any $\alpha \in \mathcal{L}_\varphi$ one has

$$E_\varphi(\alpha) = \Lambda_\varphi(\alpha) = H_\varphi(\alpha).$$

Proof. In Theorem 4.1 we proved that for any continuous transformation $f$ one has $E_\varphi(\alpha) \leq \Lambda_\varphi(\alpha) \leq H_\varphi(\alpha)$ for all $\alpha \in \mathcal{L}_\varphi$. Hence, it is sufficient for the proof of (11) to show the opposite inequalities $E_\varphi(\alpha) \geq \Lambda_\varphi(\alpha) \geq H_\varphi(\alpha)$. We start with the inequality $\Lambda_\varphi(\alpha) \geq H_\varphi(\alpha)$. Our proof relies on the proof of statement 3 of Theorem 1 in [25], but let us first recall one result of Katok [10].
THEOREM 5.2. Let \( f : X \to X \) be a continuous transformation on a compact metric space and let \( \nu \) be an ergodic invariant measure. For \( \epsilon > 0, \delta > 0 \), denote by \( N^\nu_\epsilon(\delta, \epsilon, n) \) the minimal number of \( \epsilon \)-balls in the \( d_\nu \)-metric which cover a set of measure at least \( 1 - \delta \).

Then, for each \( \delta \in (0, 1) \), we have

\[
\lim_{\epsilon \to 0} \frac{1}{n} \log N^\nu_\epsilon(\delta, \epsilon, n) = h^\nu(\delta, \epsilon, n).
\]

Remark 5.1. Suppose \( \nu \) is ergodic and \( Y \subseteq X \) is such, that \( \nu(Y) \geq 1 - \delta \). Denote by \( S(Y, \epsilon, n) \) the maximal cardinality of an \((n, \epsilon)\)-separated set in \( Y \). Similar to (7) we conclude that \( S(Y, \epsilon, n) \geq N^\nu_\epsilon(\delta, \epsilon, n) \).

To prove the inequality \( \Lambda_\psi(\alpha) \geq H_\psi(\alpha) \), it is sufficient to show that for any \( \gamma > 0 \) and every \( \mu \in M_f(X, \psi, \alpha) \) one has

\[
\Lambda_\psi(\alpha) \geq h_\mu(f) - 4\gamma.
\]

Choose arbitrary \( \gamma > 0 \), and let \( \epsilon > 0, \delta > 0 \) be so small that the following hold:

1. \( \gamma > \delta \);
2. \( d(x, y) < \epsilon \Rightarrow |\psi(x) - \psi(y)| < \delta \);
3. \( \lim_{n \to \infty} \frac{1}{n} \log N(\alpha, 3\delta, n, \epsilon) < \Lambda_\psi(\alpha) + \gamma \).

We can fulfill these requirements because \( N(\alpha, \delta, n, \epsilon) \) is increasing as \( \epsilon \) decreases, and therefore for every \( \epsilon > 0 \)

\[
\lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{n} \log N(\alpha, 3\delta, n, \epsilon) \leq \Lambda_\psi(\alpha).
\]

Now take \( \delta > 0 \) so small that conditions (1) and (3) are true, and then choose a sufficiently small \( \epsilon > 0 \) such that (2) holds as well.

We can approximate \( \mu \) by an invariant measure \( \nu \) with the following properties (see [25, p. 535]):

1. \( \nu = \sum_{i=1}^{k} \lambda_i v_i \), where \( \lambda_i > 0, \sum_i \lambda_i = 1 \) and \( v_i \) is an ergodic invariant measure for every \( i = 1, \ldots, k \);
2. \( h^\nu(f) \geq h^\mu(f) - \gamma \);
3. \( \left| \int \psi \, d\nu - \int \psi \, d\mu \right| < \delta \).

Since \( v_i \) is ergodic for every \( i \), there exists a sufficiently large \( N \) such that the set of points

\[
Y_i(N) = \left\{ x \in X : \left| \frac{1}{n} \sum_{j=0}^{n-1} \psi(f^j(x)) - \int \psi \, d\nu \right| < \gamma \text{ for all } n > N \right\}
\]

has a \( v_i \)-measure of at least \( 1 - \gamma \) for every \( i = 1, \ldots, k \).

Therefore, according to Theorem 5.2, there exist integers \( N_i \) such that for all \( n_i > N_i \) the minimal number of \( 4\epsilon \)-balls in the \( d_{\nu_i} \)-metric which is necessary to cover \( Y_i(N) \) is greater than or equal to \( \exp(n_i(h_{\nu_i}(f) - \gamma)) \). This implies, according to Remark 5.1, that the cardinality of a maximal \((n_i, 4\epsilon)\)-separated set in \( Y_i(N) \) is greater than or equal to \( \exp(n_i(h_{\nu_i}(f) - \gamma)) \). Finally, choose a sufficiently large integer \( N_0 \) such that for every \( n > N_0 \) one has

\[
n_i := \lfloor \lambda_i n \rfloor > \max(N_i, N)
\]
for all \( i = 1, \ldots, k \), also denote by \( C(n_i, 4\varepsilon) \) some maximal \((n_i, 4\varepsilon)\)-separated set in \( Y_i(N) \). For every \( k \)-tuple \((x_1, \ldots, x_k)\), where \( x_i \in C(n_i, 4\varepsilon) \), find a point \( y = y(x_1, \ldots, x_k) \in X \) such that it shadows pieces of orbits \( \{x_i, \ldots, f^{m-1}x_i : i = 1, \ldots, k\} \) within the distance \( \varepsilon \) and the gap \( m = m(\varepsilon) \). Put \( \hat{n} = m(k-1)+\sum n_i \). Firstly, we observe that different points \( y = y(x_1, \ldots, x_k) \) correspond to different \((x_1, \ldots, x_k) \in C(n_1) \times \cdots \times C(n_k) \). This is indeed the case, because for \( y = y(x_1, \ldots, x_k) \) and \( y' = y(x_1', \ldots, x_k') \) one has
\[
d_{\hat{d}}(y, y') > 2\varepsilon.
\] Secondly, for every \( y = y(x_1, \ldots, x_k) \) one has
\[
\left| \frac{1}{\hat{n}} \sum_{p=0}^{\hat{n}-1} \varphi(f^p(y)) - \alpha \right| < 2\delta + \frac{km}{\hat{n}} \|\varphi\|_{C^0}.
\]
Hence, for sufficiently large \( \hat{n} \) (i.e. large \( n \)) every point \( y = y(x_1, \ldots, x_k) \) is in \( P(\alpha, 3\delta, \hat{n}) \).

On the other hand, due to (12), one would need at least
\[
\#(C_n) \times \cdots \times #(C_n) \geq \exp([\lambda_1 n][h_{\mu}(f) - \gamma] + \cdots + [\lambda_k n][h_{\mu}(f) - \gamma])
\]
\[
\geq \exp(n(h_{\mu}(f) - 2\gamma)) \geq \exp(n(h_{\mu}(f) - 3\gamma))
\]
\( \varepsilon \)-balls in the \( d_{\hat{d}} \)-metric to cover \( P(\alpha, 3\delta, \hat{n}) \). Therefore,
\[
\lim_{n \to \infty} \frac{1}{n} \log N(\alpha, 3\delta, \hat{n}, \varepsilon) \geq h_{\mu}(f) - 3\gamma.
\]
Hence, due to the choice of \( \varepsilon, \delta > 0 \), we have \( \Lambda_{\varphi}(\alpha) + \gamma > h_{\mu}(f) - 3\gamma \). This finishes the proof of our first inequality \( \Lambda_{\varphi}(\alpha) \geq H_{\varphi}(\alpha) \).

A much more difficult inequality to prove is the remaining one: \( E_{\varphi}(\alpha) \geq \Lambda_{\varphi}(\alpha) \). In order to show it we will construct a Moran fractal, suitable for the purposes of computation of topological entropy. Roughly speaking a Moran fractal is a limit set with the following geometric construction. Consider a monotonic sequence of compact sets \([F_k], F_{k+1} \subseteq F_k\), such that \( F_k \) is a union of \( N_k \) closed sets \( \Delta_i^{(k)} \), \( i = 1, \ldots, N_k \), of approximately the same size. Moreover, the sets \( \Delta_i^{(k+1)} \) forming the \((k+1)\)th level of the construction are somewhat similar to the sets \( \Delta_i^{(k)} \) of the \( k \)th level. The Moran fractal associated to this construction is the set \( F \),
\[
F = \bigcap_k F_k.
\]
One could think of a Moran fractal as a generalization of a standard middle-third Cantor set. A particular choice of \( F_k \) will ensure that the limit set \( F \) will be a closed subset of \( K_\alpha \), but it will also allow us to construct a probability measure \( \mu \) on \( F \), satisfying the conditions of the Entropy Distribution Principle with \( s = \Lambda_{\varphi}(\alpha) - \gamma \) for any \( \gamma > 0 \). Thus the topological entropy of \( F \) will be larger than or equal to \( s \). Since \( F \subseteq K_\alpha \), the same will be true for the topological entropy of \( K_\alpha \).

Fix some \( \gamma > 0 \), and choose a sufficiently small \( \varepsilon > 0 \) such that
\[
\lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{n} \log M(\alpha, \delta, n, 8\varepsilon) \geq \Lambda_{\varphi}(\alpha) - \gamma/2.
\]
Assume that \( f \) satisfies specification, let \( m = m(\epsilon) \) be as in the definition of the specification property and let
\[
m_k = m(\epsilon/2^k), \quad k \geq 1.
\]
Choose also some sequence \( \delta_k \downarrow 0 \) and a sequence \( n_k \uparrow +\infty \) such that
\[
M_k := M(\alpha, \delta_k, n_k, 8\epsilon) > \exp(n_k(\Lambda_{\psi}(\alpha) - \gamma)) \quad \text{and} \quad n_k \geq 2^m_k.
\]
To shorten the notation we put \( s = \Lambda_{\psi}(\alpha) - \gamma \).

By definition, \( M_k \) is the cardinality of a maximal \((n_k, 8\epsilon)\)-separated set in \( P(\alpha, \delta_k, n_k) \).

Denote by \( C_k = \{x_i^k : i = 1, \ldots, M_k\} \) one of these maximal \((n_k, 8\epsilon)\)-separated sets.

**Step 1. Construction of intermediate sets \( D_k \).** We start by choosing some sequence of integers \( \{N_k\} \) such that \( N_1 = 1 \) and the two following conditions are satisfied:

1. \( N_k \geq 2^{n_k+1+m_k+1} \) for \( k \geq 2 \);
2. \( N_{k+1} \geq 2^{N_1 n_1 + \cdots + N_k(n_k + m_k)} \) for \( k \geq 1 \).

Then this sequence \( N_k \) grows very fast, and in particular
\[
\lim_{k \to \infty} \frac{n_k+1 + m_k+1}{N_k} = 0 \quad \text{and} \quad \lim_{k \to \infty} \frac{N_1 n_1 + \cdots + N_k(n_k + m_k)}{N_{k+1}} = 0. \quad (13)
\]

For any \( N_k \)-tuple \((i_1, \ldots, i_{N_k}) \in \{1, \ldots, M_k\}^{N_k}\) let \( y(i_1, \ldots, i_{N_k}) \) be some point which shadows pieces of orbits \( \{x^k_{i_1}, f x^k_{i_1}, \ldots, f^{n_k-1} x^k_{i_1}\}, j = 1, \ldots, N_k\), with a gap \( m_k\), i.e.
\[
d_{n_k}(x_{i_1}, f^{a_j} y(i_1, \ldots, i_{N_k})) < \frac{\epsilon}{2^k},
\]
where \( a_j = (n_k + m_k)(j - 1), j = 1, \ldots, N_k\). Such a point \( y(i_1, \ldots, i_{N_k}) \) exists, because \( f \) satisfies specification. Collect all such points into the set
\[
D_k = \{y(i_1, \ldots, i_{N_k}) : i_1, \ldots, i_{N_k} \in \{1, \ldots, M_k\}\}. \quad (14)
\]

We claim that different tuples \((i_1, \ldots, i_{N_k})\) produce different points \( y(i_1, \ldots, i_{N_k}) \), and that these points are sufficiently separated in the metric \( d_{n_k} \), where
\[
t_k = N_k n_k + (N_k - 1)m_k.
\]

This is the content of the following lemma.

**Lemma 5.1.** If \((i_1, \ldots, i_{N_k}) \neq (j_1, \ldots, j_{N_k})\), then
\[
d_{n_k}(y(i_1, \ldots, i_{N_k}), y(j_1, \ldots, j_{N_k})) > 6\epsilon. \quad (15)
\]

Hence, \( \#(D_k) = M_k^{N_k} \).

Since \( N_1 = 1 \), without loss of generality we may assume that \( D_1 = C_1 \).
Step 2. Construction of $L_k$. Here we construct inductively a sequence of finite sets $L_k$. Points of $L_k$ will be the centers of balls forming the $k$th level of our Moran construction.

Let $L_1 = D_1$ and put $l_1 = n_1$. Suppose we have already defined a set $L_k$, now we present a construction of $L_{k+1}$. We let

$$l_{k+1} = l_k + m_{k+1} + t_{k+1} = N_1 n_1 + N_2 (n_2 + m_2) + \cdots + N_{k+1} (n_{k+1} + m_{k+1}). \quad (16)$$

For every $x \in L_k$ and $y \in D_{k+1}$ let $z = z(x, y)$ be some point such that

$$d_{l_k}(x, z) < \frac{\varepsilon}{2^{k+1}} \quad \text{and} \quad d_{l_{k+1}}(y, f^{l_k+m_k+1}z) < \frac{\varepsilon}{2^{k+1}}. \quad (17)$$

Such a point exists due to the specification property of $f$. Collect all these points into the set

$$L_{k+1} = \{ z = z(x, y) : x \in L_k, y \in D_{k+1} \}. \quad (18)$$

Similar to the proof of Lemma 5.1 we can show that different pairs $(x, y), x \in L_k, y \in D_{k+1}$, produce different points $z = z(x, y)$. Hence, $\#(L_{k+1}) = \#(L_k) \#(D_{k+1})$.

Therefore, by induction

$$\#(L_k) = \#(D_1) \cdots \#(D_k) = M_1^{N_k} \cdots M_k^{N_k}.$$

It immediately follows from (15) and (17) that, for every $x \in L_k$ and any $y, y' \in D_{k+1}, y \neq y'$, one has

$$d_{l_k}(z(x, y), z(x, y')) < \frac{\varepsilon}{2^k} \quad \text{and} \quad d_{l_{k+1}}(z(x, y), z(x, y')) > 5\varepsilon. \quad (19)$$

There is an obvious tree structure in the construction of the sets $L_k$. We will say that a point $z \in L_{k+1}$ descends from $x \in L_k$ if there exists $y \in D_{k+1}$ such that $z = z(x, y)$. We also say that a point $z \in L_{k+p}$ descends from $x \in L_k$ if there exists a sequence of points $(z_k, \ldots, z_{k+p}), z_k = x, z_{k+p} = z$, and $z_l \in L_l$, such that $z_{l+1}$ descends from $z_l$ in the above sense for every $l = k, \ldots, k + p - 1$.

Step 3. The Moran fractal $F$. For every $k$ put

$$F_k = \bigcup_{x \in L_k} \overline{B}_{l_k} \left( x, \frac{\varepsilon}{2^{k-1}} \right),$$

where $\overline{B}_{l}(x, \delta)$ is the closed ball around $x$ of radius $\delta$ in the metric $d_l$, i.e.

$$\overline{B}_{l}(x, \delta) = \{ y \in X : d_{l}(x, y) \leq \delta \}.$$

**Lemma 5.2.** For every $k$ the following is satisfied:

1. for any $x, x' \in L_k, x \neq x'$, the sets $\overline{B}_{l_k}(x, \varepsilon/2^{k-1}), \overline{B}_{l_k}(x', \varepsilon/2^{k-1})$ are disjoint;
2. if $z \in L_{k+1}$ descends from $x \in L_k$, then

$$\overline{B}_{l_{k+1}} \left( z, \frac{\varepsilon}{2^k} \right) \subseteq \overline{B}_{l_k} \left( x, \frac{\varepsilon}{2^{k-1}} \right).$$

Hence, $F_{k+1} \subseteq F_k$. 

Finally, we put
\[ F = \bigcap_{k \geq 1} F_k. \]
It is clear that \( F \) is a non-empty closed subset of \( X \).

**Lemma 5.3.** For every \( x \in F \) one has
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \phi(f^i(x)) = \alpha. \]
Therefore, \( F \subseteq K_\alpha \).

**Step 4. A special probability measure \( \mu \).** For every \( k \geq 1 \) define an atomic probability measure \( \mu_k \) as follows:
\[ \mu_k(\{z\}) = \frac{1}{\#(L_k)} \quad \text{for every } z \in L_k. \]
Obviously, \( \mu_k(F_k) = 1 \).

**Lemma 5.4.** A sequence of probability measures \( \{\mu_k\} \) converges in the weak topology. Denote the limiting measure by \( \mu \), then \( \mu(F) = 1 \).

An important property of the limiting measure \( \mu \) is formulated in the next lemma.

**Lemma 5.5.** For every sufficiently large \( n \) and every point \( x \in X \) such that
\[ B_n(x, \varepsilon/2) \cap F \neq \emptyset \]
one has
\[ \mu(B_n(x, \varepsilon/2)) \leq e^{-n(s-\gamma)}. \]
(20)

Summarizing the above we see that, for every positive \( \gamma \) and every sufficiently small \( \varepsilon > 0 \), we have constructed a compact set \( F, F \subseteq K_\alpha \), and a measure \( \mu \) such that (20) holds.

From the Entropy Distribution Principle and the fact that \( F \subseteq K_\alpha \), we conclude that
\[ \Lambda_\psi(\alpha) - 2\gamma = s - \gamma \leq h_{\text{top}}(f, F, \varepsilon/2) \leq h_{\text{top}}(f, K_\alpha, \varepsilon/2), \]
and hence
\[ \mathcal{E}_\psi(\alpha) = h_{\text{top}}(f, K_\alpha) = \lim_{\varepsilon \to 0} h_{\text{top}}(f, K_\alpha, \varepsilon) \geq \Lambda_\psi(\alpha) - 2\gamma. \]
Since \( \gamma > 0 \) is arbitrary, we finally conclude that \( \mathcal{E}_\psi(\alpha) \geq \Lambda_\psi(\alpha) \), which finishes the proof of Theorem 5.1. 

6. **Manneville–Pomeau map**
Before we start we the detailed discussion of the multifractal spectrum for Lyapunov exponents of the Manneville–Pomeau maps, let us establish a general relation between the multifractal spectra in general and the Legendre transform of the pressure function.
For a continuous function $\varphi : X \to \mathbb{R}$ and $q \in \mathbb{R}$, let $P_\varphi(q) = P(q\varphi)$, where $P(\cdot)$ is the topological pressure. By the Classical Variational Principle one has

$$P(\psi) = \sup \left\{ h_\mu(f) + \int \psi \, d\mu : \mu \in \mathcal{M}_f(X) \right\}.$$  

Since we have assumed that the topological entropy of $f$ is finite, $P(\psi)$ is finite for every continuous $\psi$. Moreover, $P(\cdot)$ is convex, Lipschitz continuous, increasing and $P(c + \psi + \xi - \xi \circ f) = c + P(\psi)$, whenever $c \in \mathbb{R}$ and $\psi, \xi \in C(X, \mathbb{R})$.

For any $\alpha \in \mathbb{R}$ define the Legendre transform $P^*_\varphi(\alpha)$ by

$$P^*_\varphi(\alpha) = \inf_{q \in \mathbb{R}} (P(q\varphi) - q\alpha).$$

Note that $P^*_\varphi(\alpha) < +\infty$ for all $\alpha \in \mathbb{R}$; however, it is possible that $P^*_\varphi(\alpha) = -\infty$.

**Theorem 6.1.** Let $f : X \to X$ be a continuous transformation with the specification property, and let $\varphi : X \to \mathbb{R}$ be a continuous function. Then:

(i) for any $\alpha \in \mathcal{L}_\varphi$, one has

$$H_\varphi(\alpha) \leq P^*_\varphi(\alpha);$$

(ii) if, moreover, $f$ is such that the entropy map $\mu \to h_\mu(f)$ is upper semi-continuous, then for any $\alpha$ from the interior of $\mathcal{L}_\varphi$ one has

$$H_\varphi(\alpha) = P^*_\varphi(\alpha).$$

**Remark 6.1.** Transformations $f : X \to X$ with an upper semi-continuous entropy map

$$H(\mu) : \mathcal{M}_f(X) \to [0, +\infty) : \mu \to h_\mu(f)$$

play a special role in the theory of equilibrium states. This class of transformations includes, for example, all expansive maps [23]. A useful property of such transformations is that every continuous function $\psi$ has at least one equilibrium state.

**Proof of Theorem 6.1.** (i) For any $\alpha \in \mathcal{L}_\varphi$ and any $q \in \mathbb{R}$ one has

$$H_\varphi(\alpha) = \sup \left\{ h_\mu(f) : \mu \in \mathcal{M}_f(X), \int \psi \, d\mu = \alpha \right\}$$

$$= \sup \left\{ h_\mu(f) + q \int \psi \, d\mu : \mu \in \mathcal{M}_f(X), \int \psi \, d\mu = \alpha \right\} - q\alpha$$

$$\leq \sup \left\{ h_\mu(f) + q \int \psi \, d\mu : \mu \in \mathcal{M}_f(X) \right\} - q\alpha = P(q\varphi) - q\alpha,$$  

where in the last equality we used the Variational Principle for topological pressure applied to $q\varphi$:

$$P(q\varphi) = \sup \left\{ h_\mu(f) + q \int \psi \, d\mu : \mu \in \mathcal{M}_f(X) \right\}.$$  

From (21) we conclude that $H_\varphi(\alpha) \leq \inf_{q \in \mathbb{R}} (P(q\varphi) - q\alpha) = P^*_\varphi(\alpha)$.

(ii) It was shown by Jenkinson [9] that, if the entropy map is upper semi-continuous, then for any $\alpha$ from the interior of $\mathcal{L}_\varphi$ there exists $q^* \in \mathbb{R}$ and an invariant measure $\nu$, which is an equilibrium state for $q^* \varphi$, such that

$$\int \varphi \, d\nu = \alpha.$$
Hence
\[ H_\psi(\alpha) = \sup \left\{ h_\mu(f) : \mu \in \mathcal{M}_f(X), \int \psi \, d\mu = \alpha \right\} \geq h_\nu(f) = P(q^*\psi) - q^*\alpha. \]

Therefore, \( H_\psi(\alpha) \geq P(q^*\psi(\alpha)) \) and the result follows.

The following theorem is an immediate corollary of Theorems 5.1 and 6.1.

**Theorem 6.2.** Suppose \( f : X \to X \) is a continuous transformation with the specification property such that the entropy map is upper semi-continuous. Then for any \( \alpha \in (\inf \mathcal{L}_\psi, \sup \mathcal{L}_\psi) \) one has
\[ \mathcal{E}_\psi(\alpha) = P(q^*\psi(\alpha)). \]

**Remark 6.2.** Note that, for transformations with the specification property, \( \mathcal{L}_\psi \) is an interval.

Let us consider in greater detail an application of the above theorem to the multifractal analysis of the Manneville–Pomeau maps.

For a given number \( s, 0 < s < 1 \), a corresponding Manneville–Pomeau map is given by
\[ f : [0, 1] \to [0, 1] : x \mapsto x + x^{1+s} \mod 1. \]

The map \( f \) is topologically conjugate to a full one-sided shift on two symbols, and thus satisfies the specification property. Moreover, \( f \) is expansive, and hence the entropy map is upper semi-continuous. Let \( \psi(x) = \log f'(x) \). With such a choice the level sets \( K_\alpha \) are precisely the level sets of pointwise Lyapunov exponents, which are defined (provided the limit exists, of course) as
\[ \lambda(x) = \lim_{n \to \infty} \frac{1}{n} \log |(f^n)'(x)| \quad \text{and} \quad K_\alpha = \{ x : \lambda(x) = \alpha \}. \]

Due to the fact that \( x = 0 \) is an indifferent fixed point for the Manneville–Pomeau map, there exist points \( x \) with \( \lambda(x) \) arbitrary close to zero, and hence \( \inf \mathcal{L}_\psi = 0 \).

Let us discuss some thermodynamic properties of the Manneville–Pomeau maps. First of all, there exists a unique absolutely continuous \( f \)-invariant measure \( \mu \). Moreover, \( \mu \) is an equilibrium state for the potential \( -\psi \) and \( \mu \) is ergodic. However, there exists another equilibrium state for \( -\psi \), namely, the Dirac measure at zero, \( \delta_0 \). The coexistence of two equilibrium states results in a non-analytic behavior of the pressure function \( P_\psi(q) : = P(q\psi) \). Namely, it was shown in [19, 22] that \( P_\psi(q) \) is positive and strictly convex for \( q > -1 \), and \( P_\psi(q) \equiv 0 \) for \( q \leq -1 \), see Figure 1. Moreover, the equilibrium states for the potentials \( q\psi, q \neq 1 \), have been identified as well. For \( q < -1 \) the unique equilibrium state is the Dirac measure at zero. For every \( q > -1 \) there is also a unique equilibrium state, but it has a positive entropy.

Since \( f \) satisfies the specification and is expansive, Theorem 6.2 is applicable and hence \( \mathcal{E}_\psi(\alpha) = P^*_\psi(\alpha) \). The graph of \( P^*_\psi(\alpha) \) is shown in Figure 1.

The entropy spectrum \( \mathcal{E}_\psi(\alpha) \) is concave, but not strictly concave. The graph of \( \mathcal{E}_\psi(\alpha) \) contains a piece of a straight line.
We represent the interval \([\inf L_\varphi, \sup L_\varphi] = [0, \bar{\alpha}]\) as the union of two intervals \([0, \alpha_0]\) and \((\alpha_0, \bar{\alpha}]\), where \(\alpha_0\) is the largest \(\alpha\) such that \(P^*(\alpha) = \alpha\), i.e. \(P^*(\cdot)\) is linear on \([0, \alpha_0]\).

In fact
\[
\alpha_0 = h_\mu(f) = \int \log f' \, d\mu,
\]
where \(\mu\) is an absolutely continuous invariant measure.

We can identify measures \(\mu_\alpha \in \mathcal{M}_f([0, 1], \varphi, \alpha)\) which maximize entropy. In fact, Theorem 6.1 gives us a method of identifying the maximizing measure: this measure must be the equilibrium state for \(q_\varphi\) for an appropriately chosen \(q = q(\alpha)\).

For each \(\alpha \in (\alpha_0, \bar{\alpha})\) there exists a unique invariant measure \(\mu_\alpha \in \mathcal{M}_f([0, 1], \varphi, \alpha)\) such that
\[
h_{\mu_\alpha}(f) = H_\varphi(\alpha) = \sup \left\{ h_v(f) : v \text{ is invariant and } \int \varphi \, dv = \alpha \right\},
\]
i.e. \(\mu_\alpha\) is a measure of maximal entropy in \(\mathcal{M}_f([0, 1], \varphi, \alpha)\), and hence
\[
h_{\text{top}}(f, K_\alpha) = h_{\mu_\alpha}(f).
\]
This measure can be obtained in the following way. For \(\alpha > \alpha_0\), there exists a unique \(q^* > -1\) such that \(P^*_\varphi(q^*) = \alpha\). Uniqueness of such a \(q^*\) follows from the strict convexity of the pressure function \(P(q_\varphi)\) for \(q > -1\) (see above). Existence follows from a result of Urbański [22], who showed that
\[
\lim_{q \searrow -1} P^*_\varphi(q) = \alpha_0 = h_\mu(f),
\]
where \(\mu\) is the absolutely continuous invariant measure of the Manneville–Pomeau map.

Let \(\mu_\alpha\) be the equilibrium state for \(q^*\varphi\), then \(\mu_\alpha\) is the measure of maximal entropy in \(\mathcal{M}_f([0, 1], \varphi, \alpha)\). From a general result of Walters [24], we know that for expansive systems, the differentiability of the pressure function \(P(q_\varphi)\) at \(q\) is equivalent to the uniqueness of the equilibrium state for \(q_\varphi\), and the derivative is given by
\[
P'(q_\varphi) = \int \psi \, dv_\varphi,
\]
where \( \nu_q \) is the unique equilibrium state for \( q\psi \). For \( q > -1 \) we know that the pressure function is differentiable, and \( q^* \) has been chosen such that \( P'(q^*\phi) = \alpha \). Therefore,

\[
\int \phi \, d\mu_\alpha = \alpha,
\]

and hence \( \mu_\alpha \) is indeed in \( \mathcal{M}_f([0, 1], \phi, \alpha) \). Now suppose that there exists another measure \( \nu \in \mathcal{M}_f([0, 1], \phi, \alpha) \) such that \( h_\nu(f) > h_{\mu_\alpha}(f) \). But then

\[
h_\nu(f) + q^* \int \phi \, d\nu = h_\nu(f) + q \alpha > h_{\mu_\alpha}(f) + q^* \int \phi \, d\mu_\alpha = P(q^*\psi),
\]

which gives a contradiction with the Variational Principle. Similarly one obtains that since \( \mu_\alpha \) is the unique equilibrium state for \( q^*\phi \), \( \mu_\alpha \) is the unique measure of maximal entropy in \( \mathcal{M}_f([0, 1], \phi, \alpha) \).

Moreover, since \( \mu_\alpha \) is the unique equilibrium state, it must be ergodic, and since \( \int \phi \, d\mu_\alpha = \alpha \), by the Ergodic Theorem, \( \mu_\alpha(K_{\alpha}) = 1 \).

For any \( \alpha \in (0, \alpha_0) \) take

\[
\mu_\alpha = \frac{\alpha}{\alpha_0} \mu + \frac{\alpha_0 - \alpha}{\alpha_0} \delta_0,
\]

where \( \mu \) is the absolutely continuous invariant measure mentioned above. Since both measures \( \mu \) and \( \delta_0 \) are equilibrium states for \( -\phi \), so is \( \mu_\alpha \). Since \( \int \phi \, d\mu = \alpha_0 \) (see (22)), we have \( \int \phi \, d\mu_\alpha = \alpha \). Again, since \( \mu_\alpha \) is an equilibrium state, there does not exist an invariant measure \( \nu \) with \( \int \phi \, d\nu = \alpha \) with \( h_\nu(f) > h_{\mu_\alpha}(f) \).

Since \( \mu, \delta_0 \) are ergodic, and \( K_{\alpha} \) are invariant sets, we conclude that

\[
\mu_\alpha(K_{\alpha}) = 0
\]

for all \( \alpha \in (0, \alpha_0) \). This is a new phenomenon, because a typical situation in multifractal analysis would be \( \mu_\alpha(K_{\alpha}) = 1 \) for the ‘maximal’ measure \( \mu_\alpha \). The explanation of this phenomenon lies in the fact that the pressure function has a phase transition of the first order at \( q = -1 \).

7. Multi-dimensional spectra and the Contraction Principle

A straightforward modification of the proof of Theorem 5.1 shows that a similar result is also valid in a multi-dimensional setting.

**Theorem 7.1.** Suppose \( f : X \to X \) is a continuous transformation of a compact metric space \( (X, d) \) satisfying the specification property and \( \phi = (\phi_1, \ldots, \phi_d) : X \to \mathbb{R}^d \) is a continuous function. For \( \alpha \in \mathbb{R}^d \) consider the set

\[
K^\phi_{\alpha} = \left\{ x \in X : \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \phi_j(f^i(x)) = \alpha_j, \ j = 1, \ldots, d \right\}.
\]

Then for any \( \alpha \in \mathbb{R}^d \) with \( K^\phi_{\alpha} \neq \emptyset \) one has

\[
\mathcal{E}_\phi(\alpha) = h_{\text{top}}(f, K^\phi_{\alpha}) = \sup \left\{ h_\mu(f) : \mu \text{ is invariant and } \int \phi \, d\mu = \alpha \right\}. \tag{23}
\]
Suppose now that we are also given a continuous map
\[ \Psi : U \to \mathbb{R}^m \]
where \( U \subseteq \mathbb{R}^d \) is such that \( \text{Im}(\varphi) = \{ \varphi(x) : x \in X \} \subseteq U \). For any \( \beta \in \mathbb{R}^m \) define a set
\[ K_{\beta}^\Psi = \left\{ x \in X : \lim_{n \to \infty} \Psi \left( \frac{1}{n} \sum_{k=0}^{n-1} \varphi(f^k(x)) \right) = \beta \right\}. \]

We are interested in the entropy spectrum of \( \Psi \circ \varphi \), i.e. the function
\[ E_{\Psi \circ \varphi}(\beta) = h_{\text{top}}(f, K_{\beta}^\Psi \circ \varphi), \]
defined on a set \( L_{\Psi \circ \varphi} = \{ \beta : K_{\beta}^\Psi \neq \emptyset \} \). Our claim is the following.

**Theorem 7.2.** Let \( f \) be a continuous transformation satisfying the specification property and such that the entropy map is upper semi-continuous. Suppose \( \varphi : X \to \mathbb{R}^d, \Psi : \mathbb{R}^d \to \mathbb{R}^m \) are continuous maps such that \( \Psi \circ \varphi \) is well defined. Then for every \( \beta \in L_{\Psi \circ \varphi} \) one has
\[ E_{\Psi \circ \varphi}(\beta) = \sup_{\alpha : \Psi(\alpha) = \beta} \{ h_\mu(f) : \mu \text{ is invariant and } \Psi \left( \int \varphi \, d\mu \right) = \alpha \}. \]

(24)

An important corollary of this theorem is the **Contraction Principle for Entropy Spectra**:

**Theorem 7.3.** Under the conditions of Theorem 7.2, for any \( \beta \in L_{\Psi \circ \varphi} \) one has
\[ E_{\Psi \circ \varphi}(\beta) = \sup_{\alpha : \Psi(\alpha) = \beta} E_{\varphi}(\alpha). \]

(25)

**Proof.** The proof is contained in the proof of Theorem 7.2. \(\square\)

In our opinion, it is an interesting question whether the Contraction Principle (25) is valid for systems without specification.

For transformations \( f \) with the specification property, the domain \( L_{\varphi} \) is a convex set and \( E_{\varphi}(\alpha) \) is a concave function. Theorems 7.2 and 7.3 can be used to produce multifractal spectra \( E_{\Psi \circ \varphi} \) which are not concave or defined on a non-convex domain \( L_{\Psi \circ \varphi} \). For another setup which also leads to a non-concave multifractal spectra see [1, Proposition 10].

Now let us give the proof of Theorem 7.2.

**Proof of Theorem 7.2.** Let us start with a simple observation. For any \( \beta \in \mathbb{R}^m \) one has
\[ H_{\Psi \circ \varphi}(\beta) = \sup \left\{ h_\mu(f) : \mu \in \mathcal{M}_f(X) \text{ and } \Psi \left( \int \varphi \, d\mu \right) = \beta \right\} \]
\[ = \sup_{\alpha : \Psi(\alpha) = \beta} \sup \left\{ h_\mu(f) : \mu \in \mathcal{M}_f(X) \text{ and } \int \varphi \, d\mu = \alpha \right\} \]
\[ = \sup_{\alpha : \Psi(\alpha) = \beta} H_\varphi(\alpha). \]

(26)

Let us start the proof of (26) by establishing first the inequality \( H_{\Psi \circ \varphi}(\beta) \geq \sup_{\alpha : \Psi(\alpha) = \beta} H_\varphi(\alpha) \). Since \( K_{\varphi,\alpha} = K_{\beta}^\Psi \subseteq K_{\beta}^\Psi \subseteq K_{\beta}^\Psi \) for any \( \alpha \) such that \( \Psi(\alpha) = \beta \), one immediately concludes that
\[ E_{\Psi \circ \varphi}(\beta) \geq \sup_{\alpha : \Psi(\alpha) = \beta} E_{\varphi}(\alpha). \]
Hence, using Theorem 7.1 and (26) one has
\[ E_{\psi \circ \phi}(\beta) \geq \sup \left\{ h_\mu(f) : \mu \in \mathcal{M}_f(X) \text{ and } \Psi \left( \int \varphi \, d\mu \right) = \beta \right\} = H_{\psi \circ \phi}(\beta). \]

To prove the opposite inequality we will have to modify the proof of Theorem 4.1. For this we introduce the following analogue of \( \Lambda_\psi(\alpha) \).

Let \( E \) be a closed subset of \( \mathbb{R}^d \). For any \( \delta > 0 \) and \( n \in \mathbb{N} \) define
\[ P(E, \delta, n) = \left\{ x \in X : \rho \left( \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)), E \right) < \delta \right\}, \]
where \( \rho \) is a standard metric on \( \mathbb{R}^d \). For any \( \varepsilon > 0 \), denote by \( N(E, \delta, n, \varepsilon) \) the minimal number of balls \( B_n(x, \varepsilon) \), which is necessary for covering the set \( P(E, \delta, n) \). (If \( P(E, \delta, n) \) is empty we let \( N(E, \delta, n, \varepsilon) = 1 \).)

Obviously, \( N(E, \delta, n, \varepsilon) \) does not increase if \( \delta \) decreases and \( N(E, \delta, n, \varepsilon) \) does not decrease if \( \varepsilon \) decreases. Hence the following limit exists:
\[ \Lambda_\psi(E) = \lim_{\varepsilon \to 0} \lim_{\delta \to 0} \lim_{n \to \infty} \frac{1}{n} \log N(E, \delta, n, \varepsilon). \quad (27) \]

Now, by setting \( E = \Phi^{-1}(\beta) = \{ \alpha : \Psi(\alpha) = \beta \} \subseteq \mathbb{R}^d \) and repeating the first part of the proof of Theorem 4.1, we conclude that
\[ E_{\psi \circ \phi}(\beta) \leq \Lambda_\psi(\Psi^{-1}(\beta)). \]

To complete the proof of our claim we have to show the inequality \( \Lambda_\psi(\Psi^{-1}(\beta)) \leq H_{\psi \circ \phi}(\beta) \). If we proceed now in a similar way to the second part of the proof of Theorem 5.1, we will encounter the following problem: equality (10) is no longer true. The reason for this is the fact that the set \( E = \Psi^{-1}(\beta) \) is not necessarily convex, and hence if we choose some \( N \) points \( x_1, \ldots, x_N \) with ergodic averages \( \sum_{i=0}^{n-1} \varphi(f^i(x_j))/n \) lying within a distance \( \varepsilon \) from \( E \), then the average value
\[ \frac{1}{N} \sum_{j=1}^{N} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x_j)) \]
does not have to be close to \( E \) at all. Note, however, that if \( E \) were a convex subset of \( \mathbb{R}^d \) then we would not have such a problem. Now let us show how we use this observation for the proof of our claim.

Suppose \( F, G \) are closed subsets of \( \mathbb{R}^d \), then obviously
\[ \Lambda_\psi(F \cup G) \leq \max\{ \Lambda_\psi(F), \Lambda_\psi(G) \}. \]

Hence, for \( E = \Psi^{-1}(\beta) \) and for any \( r > 0 \) we have
\[ \Lambda_\psi(E) \leq \sup_{\alpha \in E} \Lambda_\psi(\overline{B}_r(\alpha)), \quad (28) \]
where \( \overline{B}_r(\alpha) \subseteq \mathbb{R}^d \) is the closed ball of radius \( r \) with the center \( \alpha \).
Since the balls $B_r(\alpha)$ are convex, we can proceed as in the second part of Theorem 5.1 and conclude that
\[
\Lambda_\psi(B_r(\alpha)) \leq \sup \left\{ h_\mu(f) : \mu \in \mathcal{M}_f(X) \text{ and } \int \psi \, d\mu \in B_r(\alpha) \right\}.
\]
Since (28) holds for any $r > 0$ we obtain that
\[
\Lambda_\psi(E) \leq \inf_{r > 0} \sup_{\alpha \in E} \left\{ h_\mu(f) : \mu \in \mathcal{M}_f(X) \text{ and } \int \psi \, d\mu \in B_r(\alpha) \right\}.
\]
Now we are going to show that for systems with an upper semi-continuous entropy map, the right-hand side of the previous inequality does not exceed $\sup_{\alpha \in E} H_\psi(\alpha)$. Assume the contrary. Then, for an arbitrary sequence $r_n \to 0$, there exists $\gamma > 0$ and a sequence $\alpha_n \in E$ such that
\[
\sup_{\alpha \in E} \left\{ h_\mu(f) : \mu \in \mathcal{M}_f(X) \text{ and } \int \psi \, d\mu \in B_{r_n}(\alpha_n) \right\} \geq \sup_{\alpha \in E} H_\psi(\alpha) + \gamma.
\]
Hence there exists a sequence of invariant measures $\{\mu_n\}$ such that $\int \psi \, d\mu_n \in B_{r_n}(\alpha_n)$ and
\[
h_{\mu_n}(f) \geq \sup_{\alpha \in E} H_\psi(\alpha) + \frac{\gamma}{2}.
\]
(29)
The sequence $\{\alpha_n\}$ is drawn from a compact set $E$, and hence has a convergent subsequence $\alpha_{n_k} \to \alpha \in E$. Without loss of generality, we may assume that a sequence of measures $\mu_{n_k}$ converges in the weak topology to an invariant measure $\mu$. (Otherwise choose a convergent subsequence.) It is clear that
\[
\int \psi \, d\mu_{n_k} = \alpha_{n_k} + O(r_{n_k}) \to \alpha \quad \text{as } k \to \infty,
\]
and $\mu$ is such that $\int \psi \, d\mu = \alpha$. Since the entropy map is upper semi-continuous, we have
\[
h_{\mu}(f) \geq \lim_{k \to \infty} h_{\mu_{n_k}}(f).
\]
(30)
Here we have arrived at a contradiction. Indeed, on the one hand $h_{\mu}(f) \leq H_\psi(\alpha)$ since $\mu \in \mathcal{M}_f(X, \psi, \alpha)$, but on the other hand from (29) and (30) we have
\[
h_{\mu}(f) \geq \sup_{\alpha \in E} H_\psi(\alpha) + \frac{\gamma}{2}.
\]
Therefore,
\[
E_{\psi|\phi}(\beta) \leq \sup_{\alpha \in E} H_\psi(\alpha),
\]
which finishes the proof.

8. Proofs

Proof of Lemma 2.1. Any continuous transformation of a compact metric space admits an invariant probability measure. Moreover, there exist ergodic invariant measures. Suppose $\mu$ is ergodic, then by the Ergodic Theorem
\[
\frac{1}{n} \sum_{i=0}^{n-1} \psi(f^i(x)) \to \int \psi \, d\mu, \quad \text{as } n \to \infty
\]
for $\mu$-a.e. $x \in X$. Hence, $\mathcal{L}_\psi \neq \emptyset$. Clearly, $\mathcal{L}_\psi \subseteq [-\|\psi\|_{C^0}, \|\psi\|_{C^0}]$, where $\|\psi\|_{C^0} = \max_x |\psi(x)| < \infty$. 

Proof of Lemma 2.2. Suppose \( K_{\alpha_i} \neq \emptyset \), \( i = 1, 2 \). Let \( t \in (0, 1) \) and put \( \alpha = t\alpha_1 + (1 - t)\alpha_2 \). Choose some \( x_i \in K_{\alpha_i} \) and take any \( \mu_i \in V(x_i), i = 1, 2 \), where \( V(x) \) is the set of limit points for the sequence of probability measure

\[
\delta_{x,n} = \frac{1}{n} \sum_{k=0}^{n-1} \delta_{f^k(x)}.
\]

Then \( \mu_i \) is an invariant measure with \( \int \varphi \, d\mu_i = \alpha_i, i = 1, 2 \) (see the proof of Lemma 4.1 below). Put \( \alpha = t\mu_1 + (1 - \alpha)\mu_2 \). Obviously, \( \int \varphi \, d\mu = \alpha \). Now, we apply [4, Proposition 21.14], which says that for a transformation with the specification property every invariant measure (not necessarily ergodic!) has a generic point, i.e. there exists a point \( x \in X \) such that \( \delta_{x,n} \to \mu \) as \( n \to \infty \). Hence, for the same point \( x \) one has

\[
\int \varphi \, d\delta_{x,n} = \frac{1}{n} \sum_{i=0}^{n-1} \varphi(f^i(x)) \to \int \varphi \, d\mu, \quad k \to \infty.
\]

Since \( x \in K_{\alpha} \), we obtain that \( \int \varphi \, d\mu = \alpha \), and hence, \( \mu \in \mathcal{M}_f(X, \varphi, \alpha) \). Convexity and closedness of \( \mathcal{M}_f(X, \varphi, \alpha) \) are trivial. 

Proof of Lemma 4.1. We start by showing that \( \mathcal{M}_f(X, \varphi, \alpha) \) is not empty for any \( \alpha \in \mathcal{L}_\varphi \). Take any \( x \in K_{\alpha} \), and denote by \( V(x) \) the set of all limit points of the sequence \( \{\delta_{x,n}\} \). Due to compactness of \( \mathcal{M}(X) \) the set \( V(x) \) is not empty. Moreover, \( V(x) \subseteq \mathcal{M}_f(X) \) [23, Theorem 6.9]. Consider an arbitrary measure \( \mu \in V(x) \). By the construction of \( V(x) \), there exists a sequence \( n_k \to \infty \) such that \( \delta_{x,n_k} \to \mu \) weakly. Hence

\[
\frac{1}{n_k} \sum_{i=0}^{n_k-1} \varphi(f^i(x)) \to \int \varphi \, d\mu, \quad k \to \infty.
\]

Since \( x \in K_{\alpha} \), we obtain that \( \int \varphi \, d\mu = \alpha \), and hence, \( \mu \in \mathcal{M}_f(X, \varphi, \alpha) \). Convexity and closedness of \( \mathcal{M}_f(X, \varphi, \alpha) \) are trivial. 

Proof of Lemma 4.2. For every \( \alpha \) in the interval \( (\inf \mathcal{L}_\varphi, \sup \mathcal{L}_\varphi) \) we have an invariant measure \( \mu \) with \( \int \varphi \, d\mu = \alpha \). Indeed, for every \( \alpha \in \mathcal{L}_\varphi \) this was shown in the previous Lemma 4.1. Any other \( \alpha \) in the interval \( (\inf \mathcal{L}_\varphi, \sup \mathcal{L}_\varphi) \) can be represented as a linear combination

\[
\alpha = \lambda \alpha_1 + (1 - \lambda)\alpha_2,
\]

where \( \lambda \in (0, 1) \) and \( \alpha_1, \alpha_2 \in \mathcal{L}_\varphi \). Then take \( \mu = \lambda \mu_1 + (1 - \lambda)\mu_2 \), where \( \mu_i \) is an invariant measure with \( \int \varphi \, d\mu_i = \alpha_i, i = 1, 2 \). The fact that \( H_\varphi(\alpha) \) is concave follows from the affinity of the entropy map \( h_\mu(f) : \mathcal{M}_f(X) \to [0, +\infty) \) [4]. Indeed, let \( \alpha, \alpha_1, \alpha_2 \) be some points from \( (\inf \mathcal{L}_\varphi, \sup \mathcal{L}_\varphi) \) such that \( \alpha = \lambda \alpha_1 + (1 - \lambda)\alpha_2 \), where \( \lambda \in [0, 1] \). Then

\[
\lambda H_\varphi(\alpha_1) + (1 - \lambda) H_\varphi(\alpha_2)
\]

\[
= \lambda \sup \left\{ h_{\mu_1}(f) : \mu_1 \in \mathcal{M}_f(X), \int \varphi \, d\mu_1 = \alpha_1 \right\}
\]

\[
+ (1 - \lambda) \sup \left\{ h_{\mu_2}(f) : \mu_2 \in \mathcal{M}_f(X), \int \varphi \, d\mu_2 = \alpha_2 \right\}
\]
\[ h_{\lambda_1 + (1-\lambda_2)}(f) = \sup \left\{ h_{\mu_1}(f) : \mu_1 \in M_f(X), \int \phi \, d\mu_1 = \alpha_1, \int \phi \, d\mu_2 = \alpha_2 \right\} \]

\[ \leq \sup \left\{ h_{\mu}(f) : \mu \in M_f(X), \int \phi \, d\mu = \alpha \right\} = H_\alpha(\phi). \]

**Proof of Lemma 5.1.** If \((i_1, \ldots, i_{N_k}) \neq (j_1, \ldots, j_{N_k})\), there exists \(l\) such that \(i_l \neq j_l\). By the construction of \(y(i_1, \ldots, i_{N_k})\) and \(y(j_1, \ldots, j_{N_k})\) we have
\[
d_{n_k}(x^k_i, f^{a_l}y(i_1, \ldots, i_{N_k})) < \epsilon \quad \text{and} \quad d_{n_k}(x^k_j, f^{a_l}y(j_1, \ldots, j_{N_k})) < \epsilon.
\]
Since \(x^k_i, x^k_j\) are different points in the \((n_k, 8\varepsilon)\)-separated set, one has
\[
d_{n_k}(f^{a_l}y(i_1, \ldots, i_{N_k}), f^{a_l}y(j_1, \ldots, j_{N_k})) \geq d_{n_k}(x^k_i, x^k_j) - d_{n_k}(x^k_i, f^{a_l}y(i_1, \ldots, i_{N_k})) - d_{n_k}(x^k_j, f^{a_l}y(j_1, \ldots, j_{N_k})) > 8\varepsilon - \varepsilon - \varepsilon = 6\varepsilon.
\]
Since \(d_{n_k}(y(i_1, \ldots, i_{N_k}), y(j_1, \ldots, j_{N_k})) \geq d_{n_k}(f^{a_l}y(i_1, \ldots, i_{N_k}), f^{a_l}y(j_1, \ldots, j_{N_k}))\), the proof is finished.

**Proof of Lemma 5.2.** (1) By (19) for \(x, x' \in L_k\), \(x \neq x'\), one has \(d_{n_k}(x, x') > 5\varepsilon\). Hence
\[
\overline{B}_{n_k}(x, \frac{\varepsilon}{2^{k-1}}) \cap \overline{B}_{n_k}(x', \frac{\varepsilon}{2^{k-1}}) = \emptyset.
\]
(2) For \(x \in L_k\) and \(z \in L_{k+1}\) such that \(z\) descends from \(x\), by (19) one has \(d_{n_k}(x, z) < \varepsilon/2^k\). Hence, \(\overline{B}_{n_k}(z, \varepsilon/2^k) \subseteq \overline{B}_{n_k}(x, \varepsilon/2^{k-1})\). Finally, since \(l_{k+1} > l_k\), one has
\[
\overline{B}_{n_{k+1}}(z, \varepsilon/2^k) \subseteq \overline{B}_{n_k}(z, \varepsilon/2^k).
\]

**Proof of Lemma 5.3.**

**Estimate on \(D_k\).** Let us introduce some notation: for any \(c > 0\) put
\[
\var{\phi, c} = \sup \{|\phi(x) - \phi(y)| : d(x, y) < c\}.
\]
Note that, due to compactness of \(X\), \(\var{\phi, c} \to 0\) as \(c \to 0\) for any continuous function \(\phi\). Also, if \(d_n(x, y) < c\), then
\[
\left| \sum_{i=0}^{n-1} \phi(f^i(x)) - \sum_{i=0}^{n-1} \phi(f^i(y)) \right| \leq \sum_{i=0}^{n-1} |\phi(f^i(x)) - \phi(f^i(y))| \leq n \var{\phi, c}.
\]
Suppose now that \(y \in D_k\); let us estimate \(\left| \sum_{p=0}^{n_k-1} \phi(f^p(y)) - t_k\alpha \right|\). By the definition of \(D_k\), there exist a \(N_k\)-tuple \((i_1, \ldots, i_{N_k})\) and points \(x^k_{ij} \in C_k\) for \(j = 1, \ldots, N_k\), such that
\[
d_{n_k}(x^k_{ij}, f^{a_j}y) < \frac{\varepsilon}{2^k}
\]
where \(a_j = (n_k + m_k)(j - 1)\). Hence,
\[
\left| \sum_{p=0}^{n_k-1} \phi(f^p x^k_j) - \sum_{p=0}^{n_k-1} \phi(f^{a_j+p}y) \right| \leq n_k \var{\phi, \frac{\varepsilon}{2^k}}.
\]
Since $s^k_{ij} \in C_k \subseteq P(\alpha, \delta_k, n_k)$ we have
\[
\left| \sum_{p=0}^{n_k-1} \psi(f^{a_j+p}y) - n_k \alpha \right| \leq n_k \left( \text{Var} \left( \varphi, \frac{\varepsilon}{2^k} \right) + \delta_k \right). \tag{31}
\]

To estimate $\left| \sum_{p=0}^{n_k-1} \psi(f^P(y)) - n_k \alpha \right|$ we represent the interval $[0, t_k - 1]$ as the union
\[
\bigcup_{j=0}^{N_k-1} [a_j, a_j + n_k - 1] \cup \bigcup_{j=0}^{N_k-2} [a_j + n_k, a_j + n_k + m_k - 1].
\]

On the intervals $[a_j, a_j + n_k - 1]$ we will use the estimate (31), and on the intervals $[a_j + n_k, a_j + n_k + m_k - 1]$ we use that
\[
\left| \sum_{p=0}^{m_k-1} \psi(f^{a_j+n_k+p}y) - m_k \alpha \right| \leq m_k(\|\psi\|_{C^0} + |\alpha|) \leq 2m_k\|\psi\|_{C^0},
\]

since $\alpha \in L_\psi \subseteq [-\|\psi\|_{C^0}, \|\psi\|_{C^0}]$. Therefore,
\[
\left| \sum_{p=0}^{n_k-1} \psi(f^P(y)) - t_k \alpha \right| \leq N_k n_k \left( \text{Var} \left( \varphi, \frac{\varepsilon}{2^k} \right) + \delta_k \right) + 2(N_k - 1)m_k\|\psi\|_{C^0}. \tag{32}
\]

\textit{Estimate on } $L_k$. Introduce
\[
R_k = \max_{z \in L_k} \left| \sum_{p=0}^{l_k-1} \psi(f^P(z)) - l_k \alpha \right|.
\]

Let us obtain by induction an upper estimate on $R_k$.

If $k = 1$, then $L_1 = D_1 = C_1 \subseteq P(\alpha, \delta_1, n_1)$ (note that $l_1 = n_1$), therefore we have
\[
R_1 \leq l_1 \delta_1.
\]

By the definition of $L_{k+1}$ every $z \in L_{k+1}$ is obtained by the shadowing of some points $x \in L_k$ and $y \in D_{k+1}$:
\[
d_{L_k}(x, z) \leq \frac{\varepsilon}{2^{k+1}}, \quad d_{D_{k+1}}(y, f^{l_k+m_k+1}z) \leq \frac{\varepsilon}{2^{k+1}}.
\]

Hence,
\[
\left| \sum_{p=0}^{l_{k+1}-1} \psi(f^P(z)) - l_{k+1} \alpha \right| \leq \left| \sum_{p=0}^{l_k-1} \psi(f^P(z)) - \sum_{p=0}^{l_k-1} \psi(f^P(x)) \right| + \left| \sum_{p=0}^{l_k-1} \psi(f^P(x)) - l_k \alpha \right| + \left| \sum_{p=l_k}^{l_{k+1}-1} \psi(f^P(z)) - m_{k+1} \alpha \right|
\]
\[
+ \left| \sum_{p=0}^{l_{k+1}-1} \psi(f^{l_k+m_k+1+p}(z)) - \sum_{p=0}^{l_{k+1}-1} \psi(f^P(y)) \right| + \left| \sum_{p=0}^{l_{k+1}-1} \psi(f^P(y)) - t_{k+1} \alpha \right|
\]
\[
\leq l_k \text{Var} \left( \varphi, \frac{\varepsilon}{2^{k+1}} \right) + R_k + 2m_{k+1}\|\psi\|_{C^0} + t_{k+1} \text{Var} \left( \varphi, \frac{\varepsilon}{2^{k+1}} \right)
\]
\[
+ N_{k+1} n_{k+1} \left( \text{Var} \left( \varphi, \frac{\varepsilon}{2^{k+1}} \right) + \delta_{k+1} \right) + 2(N_{k+1} - 1)m_{k+1}\|\psi\|_{C^0}.
\]
On the variational principle for topological entropy

where we have used the estimate (32) for $|\sum_{p=0}^{l_{k+1}-1} \varphi(f^p(y)) - t_{k+1} \alpha|$. Hence

$$R_{k+1} \leq R_k + 2l_{k+1} \text{Var} \left( \varphi, \frac{\varepsilon}{2^{k+1}} \right) + l_{k+1} \delta_{k+1} + 2N_{k+1}m_{k+1} \| \varphi \|_{C^0},$$

and by induction

$$R_k \leq 2 \sum_{p=1}^{k} l_p \left( \text{Var} \left( \varphi, \frac{\varepsilon}{2^p} \right) + \delta_p + \frac{N_p m_p \| \varphi \|_{C^0}}{l_p} \right). \quad (33)$$

Let us analyze the expression obtained for $R_k$. We claim that $R_k / l_k \to 0$ as $k \to \infty$. We start by observing that $\text{Var}(\varphi, \varepsilon/2^k) \to 0$ since $\varphi$ is continuous. By the choice of the sequence $\{\delta_k\}$ one also has $\delta_k \to 0$. Moreover, since $l_k \geq N_k(n_k + m_k)$ and the sequence $\{n_k\}$ is such that $n_k \to \infty$ as $k \to \infty$, and $n_k \geq 2^{m_1}$, we also conclude that $m_k / n_k \to 0$. Therefore, we can rewrite (33) as

$$R_k \leq \sum_{p=1}^{k} l_p c_p,$$

where $c_k \to 0$ as $k \to \infty$. By the choice of $N_k$ (13), we have $l_k \geq 2^{l_{k-1}}$, hence for sufficiently large $k$ one has

$$\frac{R_k}{l_k} \leq c_k + \frac{1}{k} \sum_{p=1}^{k-1} c_p,$$

and hence $R_k / l_k \to 0$ as $k \to \infty$.  

Estimate on $F$. Now, suppose $x \in F, n \in \mathbb{N}$ and $n > l_1$. Then there exists a unique $k \geq 1$ such that

$$l_k < n \leq l_{k+1}.$$

Also, there exists a unique $j, 0 \leq j \leq N_k - 1$, such that

$$l_k + j(n_{k+1} + m_{k+1}) < n \leq l_k + (j + 1)(n_{k+1} + m_{k+1}).$$

Since $x \in F$ there exists $z \in L_{k+1}$ such that

$$d_{l_{k+1}}(x, z) < \frac{\varepsilon}{2^{k+1}}.$$

On the other hand, since $z \in L_{k+1}$ there exist $\bar{x} \in L_k$ and $y \in D_{k+1}$ such that

$$d_{l_k}(\bar{x}, z) < \frac{\varepsilon}{2^{k+1}}, \quad d_{l_{k+1}}(y, f^{l_k+m_{k+1}}z) < \frac{\varepsilon}{2^{k+1}}.$$

Therefore,

$$d_{l_k}(x, \bar{x}) < \frac{\varepsilon}{2^{k+1}}, \quad d_{l_{k+1}}(f^{l_k+m_{k+1}}x, y) < \frac{\varepsilon}{2^{k+1}}.$$

Moreover, if $j > 0$, then by the definition of $D_{k+1}$ there exist points $x_i^{k+1}, \ldots, x_j^{k+1} \in C_{k+1}$ such that

$$d_{m_{k+1}}(x_i^{k+1}, f^{a_t}y) < \frac{\varepsilon}{2^{k+1}},$$

where $a_t = (n_{k+1} + m_{k+1})(t - 1), t = 1, \ldots, j$, and hence

$$d_{m_{k+1}}(x_i^{k+1}, f^{l_k+m_{k+1}+a_t}x) < \frac{\varepsilon}{2^{k+1}}. \quad (34)$$
We represent \([0, n - 1]\) as the union
\[ [0, l_k - 1] \cup \bigcup_{t=1}^{j} [l_k + (t - 1)(m_{k+1} + n_{k+1}), l_k + t(m_{k+1} + n_{k+1}) - 1] \]
\[ \cup [l_k + j(m_{k+1} + n_{k+1}), n - 1]. \]

One has
\[ \left| \sum_{p=0}^{l_k-1} \varphi(f^p x) - l_k \alpha \right| \leq \left| \sum_{p=0}^{l_k-1} \varphi(f^p x) - \sum_{p=0}^{l_k-1} \varphi(f^p \bar{x}) \right| + \left| \sum_{p=0}^{l_k-1} \varphi(f^p \bar{x}) - l_k \alpha \right| \]
\[ \leq l_k \text{Var}(\varphi, \frac{\varepsilon}{2^{k-1}}) + R_k. \]

On each of the intervals \([a_i, a_i + (m_{k+1} + n_{k+1}) - 1]\), where \(a_i = l_k + (t - 1)(m_{k+1} + n_{k+1})\), we estimate that
\[ \sum_{p=a_i}^{a_i+m_{k+1}+n_{k+1}-1} \varphi(f^p x) - (m_{k+1} + n_{k+1})\alpha \]
\[ \leq 2m_{k+1}\|\varphi\|_{C^0} + n_{k+1}\delta_{k+1} + n_{k+1}\text{Var}(\varphi, \varepsilon/2^{k-2}), \]

because of (34) and the fact that \(A_{ij}^{k+1} \in C_{k+1} \subseteq P(\alpha, \delta_{k+1}, n_{k+1})\).

Finally, on \([l_k + j(m_{k+1} + n_{k+1}), n - 1]\) we have
\[ \left| \sum_{p=l_k+j(m_{k+1}+n_{k+1})}^{n-1} \varphi(f^p x) - (n - l_k - j(m_{k+1} + n_{k+1}))\alpha \right| \]
\[ \leq 2(n - l_k - j(m_{k+1} + n_{k+1}))\|\varphi\|_{C^0} \leq 2(n_{k+1} + m_{k+1})\|\varphi\|_{C^0}. \]

Collecting all estimates together we have
\[ \left| \sum_{p=0}^{n-1} \varphi(f^p x) - n\alpha \right| \leq R_k + (l_k + jn_{k+1}) \text{Var}(\varphi, \frac{\varepsilon}{2^{k-2}}) \]
\[ + 2(n_{k+1} + (j + 1)m_{k+1})\|\varphi\|_{C^0} + jn_{k+1}\delta_{k+1}. \]

Now, since \(n > l_k + j(n_{k+1} + m_{k+1})\) and \(l_k > N_k\), we obtain
\[ \left| \frac{1}{n} \sum_{p=0}^{n-1} \varphi(f^p x) - \alpha \right| < \frac{R_k}{l_k} \text{Var}(\varphi, \frac{\varepsilon}{2^{k-2}}) + 2 \left( \frac{n_{k+1} + m_{k+1}}{N_k} + \frac{m_{k+1} + n_{k+1}}{n_{k+1}} \right)\|\varphi\|_{C^0} + \delta_{k+1}. \]

Since the right-hand side tends to zero as \(k \to \infty\) and \(k \to \infty\) for \(n \to \infty\), we finally conclude that
\[ \lim_{n \to \infty} \frac{1}{n} \sum_{p=0}^{n-1} \varphi(f^p x) = \alpha \]

for all \(x \in F\), and hence \(F \subseteq K_a\). \(\Box\)
Proof of Lemma 5.4. We are going to show that for every continuous function $\psi$ there exists a limit
\[
I(\psi) = \lim_{k \to \infty} \int \psi \, d\mu_k.
\] (35)
Obviously, if $I(\psi)$ is well defined, then $I$ is a positive linear functional on $C(X, \mathbb{R})$. Hence by the Riesz Theorem there exists a unique probability measure $\mu$ on $X$ such that
\[
I(\psi) = \int \psi \, d\mu \quad \text{for every } \psi \in C(X, \mathbb{R}),
\]
and thus $\mu_k \to \mu$ weakly.

Let us prove (35). It is sufficient to show that for every $\delta > 0$ there exists $K = K(\delta) > 0$ such that for all $k_1, k_2 > K$ one has
\[
\left| \int \psi \, d\mu_{k_1} - \int \psi \, d\mu_{k_2} \right| = \left| \frac{1}{\#(L_{k_1})} \sum_{x \in L_{k_1}} \psi(x) - \frac{1}{\#(L_{k_2})} \sum_{y \in L_{k_2}} \psi(y) \right| < \delta.
\]
Without loss of generality, we may assume that $k_1 > k_2$. Then
\[
\left| \frac{1}{\#(L_{k_1})} \sum_{x \in L_{k_1}} \psi(x) - \frac{1}{\#(L_{k_2})} \sum_{y \in L_{k_2}} \psi(y) \right| \leq \frac{1}{\#(L_{k_1})} \sum_{x \in L_{k_1}} |\psi(x) - \psi(y(x))|,
\]
where $y(x) \in L_{k_2}$ is a unique point in $L_{k_2}$ such that $x$ descends from $y(x)$. Taking into account the way the sets $L_k$ were constructed, we conclude that
\[
d(x, y(x)) \leq \frac{\varepsilon}{2^{k_1}}.
\]
Hence, for $k_1, k_2 > K$ one has
\[
\left| \int \psi \, d\mu_{k_1} - \int \psi \, d\mu_{k_2} \right| \leq \sup \left( |\psi(x) - \psi(y)| : d(x, y) < \frac{\varepsilon}{2^{k_1}} \right) \to 0 \quad \text{as } K \to \infty.
\]

Now, we have to show that $\mu(F) = 1$. Note that $\mu_{k+p}(F_k) = 1$ for all $p \geq 0$, since $F_{k+p} \subseteq F_k$ and $\mu_{k+p}(F_{k+p}) = 1$ by construction. Since $\mu$ is the weak limit of $\{\mu_k\}$ and $F_k$ are closed, using the properties of weak convergence of measures we obtain
\[
\mu(F_k) \geq \lim_{p \to \infty} \mu_{k+p}(F_k) = 1,
\]
and hence $\mu(F_k) = 1$. Finally, since $F = \bigcap_k F_k$, one has $\mu(F) = 1$. □

Proof of Lemma 5.5. By definition $B_n(x, \varepsilon)$ is an open set, thus, since $\mu_k \to \mu$, we have
\[
\mu(B_n(x, \varepsilon)) \leq \lim_{k \to \infty} \mu_k(B_n(x, \varepsilon)) = \lim_{k \to \infty} \frac{1}{\#(L_k)} \#(\{z \in L_k : z \in B_n(x, \varepsilon)\}).
\]
Suppose $n \geq l_1 = n_1$, then there exists $k \geq 1$ such that
\[
l_k < n \leq l_{k+1}.
\]
As in the proof of Lemma 5.3, let $j \in \{0, \ldots, N_{k+1} - 1\}$ be such that
\[
l_k + (n_{k+1} + m_{k+1}) j < n \leq l_k + (n_{k+1} + m_{k+1})(j + 1).
\]
We start by showing that \( \#(B_n(x, \varepsilon) \cap L_k) \leq 1 \), and thus \( \mu_k(B_n(x, \varepsilon)) \leq \#(L_k)^{-1} \). Indeed, suppose there are two points \( z_1, z_2 \in L_k \) such that \( z_1, z_2 \in B_n(x, \varepsilon) \). This means that \( d_n(z_1, z_2) < 2\varepsilon \). However, from (19) we know that \( d_l(z_1, z_2) > 5\varepsilon \). Hence, we have arrived at a contradiction, since \( n > l_k \) and thus \( d_l(z_1, z_2) \geq d_l(z_1, z_2) \).

We continue by showing that \( \mu_{k+1}(B_n(x, \varepsilon)) \) does not exceed \( \#(L_k) \times M_{k+1}^{-1} \). Suppose, two points \( z_1, z_2 \in L_{k+1} \) are in \( B_n(x, \varepsilon) \) as well. Therefore, there exist points \( x_1, x_2 \in L_k \) and \( y_1, y_2 \in D_{k+1} \) such that

\[
\begin{align*}
z_1 &= z(x_1, y_1), \\
z_2 &= z(x_2, y_2).
\end{align*}
\]

All the points in \( D_{k+1} \) are obtained by shadowing certain combinations of points from \( C_{k+1} \) (see (14)), i.e.

\[
y_1 = y(i_1, \ldots, i_{N_{k+1}}), \\
y_2 = y(i'_1, \ldots, i'_{N_{k+1}}),
\]

where \( (i_1, \ldots, i_{N_{k+1}}), (i'_1, \ldots, i'_{N_{k+1}}) \in [1, \ldots, M_{k+1}]^{N_{k+1}} \).

We claim that \( x_1 = x_2 \) and \( (i_1, \ldots, i_j) = (i'_1, \ldots, i'_j) \). Indeed, if \( x_1 \neq x_2 \) then

\[
\begin{align*}
d_l(x_1, x_2) &\leq d_l(x_1, z_1) + d_l(z_1, x) + d_l(x, z_2) + d_l(z_2, x_2) \\
&\leq \frac{\varepsilon}{2k} + \varepsilon + \frac{\varepsilon}{2} \leq 5\varepsilon,
\end{align*}
\]

and thus we have a contradiction with (19). Similarly we proceed with our second claim.

If \( j = 0 \) there is nothing to prove. Suppose \( j > 0 \) and there exists \( t, 1 \leq t \leq j \), such that \( i_t \neq i_t' \). Since \( y_1 = y(i_1, \ldots, i_{N_{k+1}}) \) and \( y_2 = y(i'_1, \ldots, i'_{N_{k+1}}) \), one has

\[
d_{n+1}(x_{i_t}^{k+1}, f^m y_1) < \frac{\varepsilon}{2k+1}, \quad d_{n+1}(x_{i'_t}^{k+1}, f^m y_2) < \frac{\varepsilon}{2k+1}.
\]

Moreover,

\[
d_{n+1}(z_1, y_1) < \frac{\varepsilon}{2k+1}, \quad d_{n+1}(z_2, y_2) < \frac{\varepsilon}{2k+1},
\]

and hence

\[
\begin{align*}
d_{n+1}(x_{i_t}^{k+1}, x_{i'_t}^{k+1}) &\leq d_{n+1}(x_{i_t}^{k+1}, f^m y_1) + d_{n+1}(y_1, f^{l_t+m_{k+1}+1} z_1) \\
&+ d_n(z_1, z_2) + d_{n+1}(f^{l_t+m_{k+1}+1} z_2, y_2) + d_{n+1}(y_2, x_{i'_t}^{k+1}) \\
&\leq \frac{\varepsilon}{2k+1} + \frac{\varepsilon}{2k+1} + \varepsilon + \frac{\varepsilon}{2k+1} + \frac{\varepsilon}{2k+1} < 6\varepsilon,
\end{align*}
\]

which contradicts the fact that \( d_{n+1}(x_{i_t}^{k+1}, x_{i'_t}^{k+1}) > 8\varepsilon \), since \( x_{i_t}^{k+1}, x_{i'_t}^{k+1} \) are different points in a \( (n_{k+1}, 8\varepsilon) \)-separated set \( C_{k+1} \).

Since \( (i_1, \ldots, i_j) \) is the same for all points \( z = z(x, y(i_1, \ldots, i_j, \ldots, i_{N_{k+1}})) \) which can lie in \( B_n(x, \varepsilon) \), we easily conclude that there are at most \( M_{k+1}^{N_{k+1}+j} \) such points. Hence

\[
\mu_{k+1}(B_n(x, \varepsilon)) \leq \frac{1}{\#(L_k) M_{k+1}^{N_{k+1}+j}} = \frac{1}{\#(L_k) M_{k+1}^{j}}.
\]

For any \( p > 1 \) one also has

\[
\mu_{k+p}(B_n(x, \varepsilon/2)) \leq \frac{1}{\#(L_k) M_{k+1}^{j}}.
\]
This is indeed the case, because the points of $L_{k+p}$ which lie in $B_n(x, \varepsilon/2)$ can only descend from the points of $L_{k+1}$ which are in $B_n(x, \varepsilon)$. We prove this finally by contradiction. Suppose we can find points $z_1 \in L_{k+1}$ and $z_2 \in L_{k+p}$, where $z_2$ descends from $z_1$, such that

$$d_n(z_2, x) < \varepsilon/2 \quad \text{and} \quad d_n(z_1, x) > \varepsilon.$$

This implies that $d_n(z_1, z_2) \geq d_n(z_1, x) - d_n(x, z_2) > \varepsilon/2$. The latter, however, is not possible, since

$$d_n(z_1, z_2) \leq d_{k+1}(z_1, z_2) \leq \frac{\varepsilon}{2^{k+2}} + \frac{\varepsilon}{2^{k+3}} + \cdots = \frac{\varepsilon}{2^{k+1}}.$$ 

Hence there are exactly $\#(D_{k+2}) \ldots \#(D_{k+p})$ points in $L_{k+p}$, $p \geq 2$, which descend from a given point in $L_{k+1}$. Hence

$$\mu_{k+p}(B_n(x, \varepsilon/2)) \leq \frac{\#(D_{k+2}) \ldots \#(D_{k+p})}{\#(L_k)^{N_{k+1}} \ldots \#(D_{k+p})} \leq \frac{1}{\#(L_k)^{N_{k+1}} \ldots \#(D_{k+p})}.$$ 

Therefore,

$$\mu(B_n(x, \varepsilon/2)) \leq \lim_{p \to \infty} \mu_{k+p}(B_n(x, \varepsilon/2)) \leq \frac{1}{\#(L_k)^{N_{k+1}} \ldots \#(D_{k+p})}.$$ 

Now, by the choice of $k$ and $j$ we have

$$n - l_k - j(n_{k+1} + m_{k+1}) \leq n_{k+1} + m_{k+1},$$

where $l_k = N_1 n_1 + N_2 n_2 + \cdots + N_k(n_k + m_k)$. Therefore,

$$\frac{n - l_k - j(n_{k+1} + m_{k+1})}{l_k + j(n_{k+1} + m_{k+1})} \leq \frac{n_{k+1} + m_{k+1}}{N_k} \to 0 \quad \text{as} \quad k \to \infty$$

because of the choice of $N_k$. Since $M_k$ has been chosen in a such way that $M_k \geq \exp(sn_k)$, and $m_k$ are much smaller than $n_k$, for large $k$ we obtain

$$\#(L_k)^{N_{k+1}} \ldots \#(D_{k+p}) \geq \exp((s - \gamma/2)(N_1 n_1 + \cdots + N_k(n_k + m_k) + j(n_{k+1} + m_{k+1}))) \geq \exp((s - \gamma)n).$$

Therefore, since $k \to \infty$ as $n \to \infty$, for all sufficiently large $n$ one has

$$\mu(B_n(x, \varepsilon/2)) \leq \exp(-n(s - \gamma))$$

for every $x$ such that $B_n(x, \varepsilon/2) \cap F \neq \emptyset$. \square
Acknowledgements. We are grateful to Ai Hua Fan for making available to us a copy of [6], and to Jörg Schmeling for valuable discussions. The second author acknowledges the support of The Netherlands Organization for Scientific Research (NWO), grant 613-06-551, and The Netherlands Ministry of Economic Affairs, EET grant K99124.

REFERENCES