Large Deviations Estimates for Dynamical Systems without the Specification Property. Application to the β-shifts.

C.-E. Pfister*
Institut d'analyse et calcul scientifique, EPF-L
CH-1015 Lausanne, Switzerland

W.G. Sullivan†
Department of Mathematics, UCD,
Belfield, Dublin 4, Ireland

August, 2004

Abstract

We consider dynamical systems whose sets of orbits verify an approximate product property. This allows us to obtain large deviations results, which were previously proven under the condition of specification property. We illustrate our results by considering the β-shifts. While the specification property holds for a set of $\beta > 1$ of Lebesgue measure zero, our approximate product property holds for any $\beta > 1$. For any β-shift the empirical measures verify a large deviations principle with respect to the probability measure of maximal entropy. We extend the dimension results of [PfS1] to any β-shift, obtaining a variational principle for the topological entropy of sets involving ergodic averages.

1 Introduction

The basic structure, where X is a metric space with metric d and $T\colon X \to X$ is a self-map, provides the foundation of the theory of dynamical systems. It is frequently assumed that X is compact. We use the term dynamical system in this article to mean *dynamical system* (X, d, T) *with* (X, d) *compact and* T *a continuous action of* L *on* X*, with* $L = \mathbf{Z}^d$ *or* $\mathbf{Z}_+^d$*,* $d \geq 1$. An important branch of this theory is symbolic dynamics, where X is a closed subspace of $\mathtt{A}^{\mathbf{Z}_+}$ and $\mathtt{A}$ is a finite set and the action of $L = \mathbf{Z}_+$ is given by the shift operator T. The structure (X, d, T) is also significant in various aspects of probability theory and in statistical mechanics of lattice systems on $\mathbf{Z}^d$ with finite state space $\mathtt{A}$. In this case T denotes the natural action of $L = \mathbf{Z}^d$ on $\mathtt{A}^{\mathbf{Z}^d}$. Problems involving scaled limits, occurring in dimension theory, multifractal analysis and large deviations estimates,

*E-mail: charles.pfister@epfl.ch
†E-mail: Wayne.Sullivan@ucd.ie

play an important role in these domains. The basic reference for dimension theory and multifractal formalism is [Pe]; see also [Te1] and [TV1]. The important references for large deviations results in the context of dynamical systems are [Y], [K1], [K2], [EKW], the last reference covering the case of statistical mechanics of lattice systems (see also [Pf]). These problems have been extensively studied, but mainly under the condition that the dynamical system has the specification property introduced by Bowen [Bo1]. An important consequence of specification, Theorem B in [EKW], asserts that any T-invariant probability measure ν is the weak limit of a sequence of ergodic measures $\{\nu_n\}$, such that *the entropy of* ν *is the limit of the entropies of the* ν_n. This is a central point in large deviations theory, which was first emphasized in [FO] (see section 3). This is also a key point in our recent approach for computing Billingsley dimension ([Bi]) on shift spaces [PfS1]. We refer to this property by saying that the ergodic measures are entropy-dense in the set of T-invariant measures on X. While specification property usually holds in statistical mechanics, often in a trivial way, it is a strong hypothesis for dynamical systems. Remarkably, if T is a continuous self-map of $[0,1]$, then specification holds if and only if T is topologically mixing [Bl]. For the class of β-shifts specification is exceptional: it is verified only on a set of β of Lebesgue measure zero [Bu], [B-M].
Our main purpose is to introduce a much weaker form of specification, which we call approximate product property, which is strong enough to imply that the ergodic measures are entropy-dense. Weaker form of specification of course appear already in the literature, for example in [M] and [H2]. In particular, in [H2] Hofbauer expressed ideas similar to ours, but did not developed them and considered only a special case, which was sufficient for his purpose (see [H2] p.303). The notion of approximate product property says that "in the large" the space of orbits of the dynamical system has almost a product structure. It is inspired by the notion of asymptotically decoupled probability measure introduced in [Pf] in the context of statistical mechanics. It expresses the basic fact, in this context, that the thermodynamic behaviour of the system is a consequence of the property that at a large scale the system is essentially the product of weakly interacting large subsystems. After fixing the notations and recalling basic definitions, the notion of approximate product property is defined in Definitions 2.9 and 2.10. The main result of that section is Theorem 2.1. In section 3 we give a general setting for large deviations of the empirical measures, giving a unified treatment for dynamical systems and statistical mechanics. In the short section 4 we specialize our results to the important case of shift spaces. In the last section we illustrate our results by considering the example of the β-shifts. As already mentioned, specification holds only for a set of Lebesgue measure zero, while the approximate product property holds for any β. We then adapt to these examples our results about Billingsley dimension for shift spaces [PfS1]. For most β, the Billingsley dimension of a subset A, computed with the measure of maximal entropy, is equal to the topological entropy of A divided by $\log \beta$. As as result we obtain a variational principle for the topological entropy of sets of generic points. Using a contraction principle we compute the topological entropy of sets involving ergodic averages.

2 Approximate product property

We use $\mathbf{R}$ to denote the set of real numbers; $\mathbf{Z}$, the set of integers; $\mathbf{N} := \{1, 2, \ldots\}$, $\mathbf{Z}_+ := \mathbf{N} \cup \{0\}$. We consider a compact metric space (X, d) and a continuous action T of the integer lattice L on X denoted by $T^i \colon X \to X$ with $i \in L$. We are interested in the cases $L = \mathbf{Z}_+^d$, particularly for $d = 1$, and $L = \mathbf{Z}^d$. The triple (X, d, T) is a **dynamical system**. An important special case is that in which X is a closed T-invariant subspace of $\mathtt{A}^L$, where $\mathtt{A} = \{0, \ldots, \mathtt{b} - 1\}$ is a finite alphabet with $\mathtt{b} > 1$ characters: $\mathtt{A}$ is given the discrete topology; $\mathtt{A}^L$, the resulting product topology. Any metric generating this topology may be used. The action of T is that induced by addition in L, $(T^j \omega)_k := \omega_{j+k}$ for all $k \in L$. We use the designation **shift space** to refer to this special case. We use the symbol $\mathcal{C}(X)$ to denote the continuous, real valued functions on X; $\mathcal{M}_1$ denotes the Borel probability measures on X; $\mathcal{M}_1^T$ denotes measures in $\mathcal{M}_1$ which are invariant under T^i, for all $i \in L$. The action of T on X extends to $\mathcal{C}(X)$ by defining $(T^i f)(x) := f(T^i(x))$. For the Borel measurable $f : X \to \mathbf{R}$, $\nu \in \mathcal{M}_1$, we use the notation

$$\langle f, \nu \rangle := \int f \, d\nu. \tag{2.1}$$

The action of T on $\mathcal{M}_1$ is defined so that $\langle f, T^i \nu \rangle = \langle T^i f, \nu \rangle$ for all $f \in \mathcal{C}(X)$. For $\Lambda \subset L$ and $i \in L$, $i + \Lambda := \{j \in L : \, j = i + k \, , \, k \in \Lambda\}$. We denote by $|B|$ the cardinality of a finite set B.

Definition 2.1 *By Λ_n we mean $[-n, n]^d$ when $L = \mathbf{Z}^d$ and $\Lambda_n := [0, \ldots, n]^d$ when $L = \mathbf{Z}_+^d$. We use D_n to denote the width of Λ_n: $D_n = 2n + 1$ for $L = \mathbf{Z}^d$ and $D_n = n + 1$ for $L = \mathbf{Z}_+^d$. We write $D_n L$ for the set $\{i \, D_n : i \in L\}$ and define $V_n := |\Lambda_n|$. For $n < 0$ we set $V_n := 0$.*

$$\mathcal{A}_n := \frac{1}{V_n} \sum_{i \in \Lambda_n} T^i. \tag{2.2}$$

For $m, n \in \mathbf{Z}_+$ we define

$$m * n := \begin{cases} mn + m + n & \text{if } L = \mathbf{Z}_+^d; \\ 2mn + m + n & \text{if } L = \mathbf{Z}^d. \end{cases} \tag{2.3}$$

Note that Λ_{m*n} can be decomposed into $V_{m*n}/V_m = V_n$ disjoint translates of Λ_m and

$$\mathcal{A}_{m*n} = \frac{1}{V_n} \sum_{j \in D_m \Lambda_n} T^j \mathcal{A}_m = \frac{1}{V_n} \sum_{j \in \Lambda_n} T^{j D_m} \mathcal{A}_m. \tag{2.4}$$

Notice also that $m * n = n * m$ and $k * (m * n) = (k * m) * n$.
The standard references for what follows is [W], in particular chapters 4 and 8, and [KH]. We fix the notations, and recall basic concepts and estimates.

Definition 2.2 *For each $x \in X$ and finite $\Lambda \subset L$ the **empirical measure** $\mathcal{E}_\Lambda(x)$ is*

$$\mathcal{E}_\Lambda(x) := \frac{1}{|\Lambda|} \sum_{i \in \Lambda} \delta_{T^i x} \, , \tag{2.5}$$

where δ_x denotes the unit measure concentrated on the point x. We use the notation $\mathcal{E}_n(x) \equiv \mathcal{A}_n \delta_x$ for $\mathcal{E}_{\Lambda_n}(x)$. For $B \subset \mathcal{M}_1$ we use the notation $X_{n,B}$ to specify

$$X_{n,B} := \{x \in X : \mathcal{E}_n(x) \in B\}. \tag{2.6}$$

*A point $x \in X$ is a **generic point** for $\nu \in \mathcal{M}_1$ if for each $f \in \mathcal{C}(X)$*

$$\lim_n \langle f, \mathcal{E}_n(x) \rangle = \langle f, \nu \rangle. \tag{2.7}$$

In order for $\nu \in \mathcal{M}_1$ to have a generic point in X, we must have $\nu \in \mathcal{M}_1^T$. We are interested in conditions which imply that each $\nu \in \mathcal{M}_1^T$ has a generic point. The Birkhoff Ergodic Theorem implies that each ergodic $\nu \in \mathcal{M}_1^T$ has a generic point.

Definition 2.3 *An* ***f-neighbourhood*** *of $\nu \in \mathcal{M}_1$ is a set of the form*

$$F^{(\alpha)} := \{\rho \in \mathcal{M}_1 : |\langle f_i, \rho \rangle - \langle f_i, \nu \rangle| \leq \alpha\, \varepsilon_i\}, \tag{2.8}$$

where $\alpha > 0$, $f_i \in \mathcal{C}$, $\varepsilon_i > 0$, $i = 1, \ldots, k$ are a given finite collection of continuous functions and positive constants and $\|f_i\| \leq 1$ for each i, where $\|f\| = \sup_{x \in x} |f(x)|$. If $\alpha = 1$, we simply write F. The ***minimal-ε*** *of F is defined by*

$$\varepsilon_{\min}(F) := \min\{\varepsilon_1, \ldots, \varepsilon_k\}. \tag{2.9}$$

The notation $X_{n,F^{(\alpha)}}$ is defined by (2.6) with $B = F^{(\alpha)}$.

The f-neighbourhoods form a neighbourhood base for the weak* topology on $\mathcal{M}_1$, which is the topology we use.

Definition 2.4 *Let ξ, η be collections of subsets of X. The* **join** *of ξ, η, $\xi \vee \eta$ is the collection of all sets of the form $A \cap B$ for $A \in \xi$, $B \in \eta$. We write*

$$T^{\vee\Lambda}\xi := \bigvee_{k \in \Lambda} T^{-k}\xi. \tag{2.10}$$

An **atom** *of $T^{\vee\Lambda}\xi$ is a nonempty subset of X of the form $\bigcap_{k\in\Lambda} T^{-k}A_k$, where each $A_k \in \xi$. We write $T^{\vee n}\xi$ for $T^{\vee\Lambda_n}\xi$.*

Definition 2.5 *A* **finite partition** ξ **of** X *is a finite collection of Borel measurable sets $\{B_1, \ldots, B_k\}$ such that $B_i \cap B_j = \emptyset$ for $i \neq j$ and $\bigcup_{B_i \in \xi} B_i = X$. The* **entropy of the measure** $\nu \in \mathcal{M}_1$ **with respect to the finite partition** ξ **of** X *is given by*

$$H(\nu, \xi) := -\sum_{B_i \in \xi} \nu(B_i) \log \nu(B_i). \tag{2.11}$$

The T**-entropy of** $\nu \in \mathcal{M}_1^T$ **with respect to** ξ *is defined by*

$$h(T, \nu, \xi) := \limsup_{n \to \infty} \frac{1}{V_n} H(\nu, T^{\vee n}\xi). \tag{2.12}$$

The T**-entropy of** ν *is defined by*

$$h(T, \nu) := \sup_{\xi} h(T, \nu, \xi), \tag{2.13}$$

where the supremum is over all finite partitions of X.

Let $\mathtt{A}$ be a finite set. The elements of $\mathtt{A}^{\Lambda_n}$ are called **words**.

Definition 2.6 *The* ***Hamming distance*** *on the words of $\mathtt{A}^{\Lambda_n}$ is*

$$d_n^H(\mathrm{v}, \mathrm{w}) := |\{i \in \Lambda_n : \mathrm{v}_i \neq \mathrm{w}_i\}|. \tag{2.14}$$

We recall two combinatorial estimates, which may be deduced as Chernoff bounds for the binomial distribution.

Lemma 2.1 *If* $\binom{n}{k}$ *denotes the number of combinations of* n *objects taken* k *at a time and* $0 \le \delta \le 1/2$ *then*

$$\sum_{k \le \delta n} \binom{n}{k} \le 2^{n\eta(\delta)}, \tag{2.15}$$

where

$$\eta(\delta) := -\delta \log_2 \delta - (1-\delta)\log_2(1-\delta). \tag{2.16}$$

Proof: This appears as Lemma 1.5.4 in [S]. For completeness we give a proof. For $k/n \le \delta \le 1/2$, $\delta/(1-\delta) \le 1$ and

$$2^{-n\eta(\delta)} = \delta^{n\delta}(1-\delta)^{n-n\delta} \le \delta^k (1-\delta)^{n-k} \quad \text{for } k \le n\delta. \tag{2.17}$$

Multiplying by $\binom{n}{k}$ and summing gives the result,

$$2^{-n\eta(\delta)} \sum_{k \le \delta n} \binom{n}{k} \le \sum_{k \le \delta n} \binom{n}{k} \delta^k (1-\delta)^{n-k} \le \sum_{k \le n} \binom{n}{k} \delta^k (1-\delta)^{n-k} = 1. \tag{2.18}$$

□

Lemma 2.2 *With* η *defined by (2.16), for* $\mathrm{w} \in \mathtt{A}^{\Lambda_n}$ *and* $0 \le \delta \le (|\mathtt{A}|-1)/|\mathtt{A}|$,

$$|\{\mathrm{v} \in \mathtt{A}^{\Lambda_n} : \, d_n^H(\mathrm{v},\mathrm{w}) \le \delta V_n\}| \le 2^{V_n \eta(\delta)} (|\mathtt{A}|-1)^{\delta V_n}. \tag{2.19}$$

Proof: For $\delta \le 1/2$ this follows from the previous lemma by noting that

$$|\{\mathrm{v} \in \mathtt{A}^{\Lambda_n} : \, d_n^H(\mathrm{v},\mathrm{w}) = k\}| = \binom{V_n}{k} (|\mathtt{A}|-1)^k. \tag{2.20}$$

□

We now consider one of our main results, Theorem 2.1. We first prove Proposition 2.1, which plays a central role. The terminology *entropy-approachable by ergodic measures*, *entropy-dense* and (δ, n, ε)*-separated sets* is not found in [W].

Definition 2.7 *The measure* $\nu \in \mathcal{M}_1^T$ *is* ***entropy-approachable by ergodic measures*** *if for each neighbourhood* G *of* ν *and each* $h^* < h(T,\nu)$*, there exists an ergodic measure* ρ *such that* $\rho \in G \cap \mathcal{M}_1^T$ *and* $h(T,\rho) > h^*$*. The ergodic measures of the dynamical system* (X,d,T) *are* ***entropy-dense*** *in* X *if each* $\nu \in \mathcal{M}_1^T$ *is entropy-approachable by ergodic measures.*

In [EKW] it is shown that, if the dynamical system (X,d,T) satisfies specification, then the ergodic measures are entropy-dense.

Definition 2.8 *For finite* $\Lambda \subset L$ *we define* d_Λ *by* $d_\Lambda \equiv 0$ *if* $\Lambda = \emptyset$*; otherwise*

$$d_\Lambda(x,y) := \max\left\{ d(T^i x, T^i y) : i \in \Lambda \right\}. \tag{2.21}$$

A set $\Gamma \subset X$ *is* (d_Λ, ε)-**separated** *if whenever* $x \ne y \in \Gamma$, $d_\Lambda(x,y) > \varepsilon$*. We write* (n,ε)*-separated in short for* $(d_{\Lambda_n}, \varepsilon)$*-separated. A set* $\Gamma \subset X$ *is* (δ, n, ε)*-**separated** if for each* $\Lambda \subset \Lambda_n$ *satisfying* $|\Lambda_n \setminus \Lambda| \le \delta V_n$, Γ *is* (d_Λ, ε)*-separated.*

Note that if $|\Gamma| = 1$, then Γ is (δ, n, ε)-separated for all $\delta, \varepsilon > 0$ and all $n \in \mathbf{Z}_+$. Two points x, y are (δ, n, ε)-separated if, and only if, $|\{j \in \Lambda_n : d(T^j x, T^j y) > \varepsilon\}| > \delta V_n$.

Proposition 2.1 *Let $\nu \in \mathcal{M}_1^T$ be ergodic and $h^* < h(T,\nu)$. Then there exists $\delta^* > 0$ and $\varepsilon^* > 0$ so that for each neighbourhood F of ν in $\mathcal{M}_1$, there exists $n_F^* \in \mathbf{N}$ such that for any $n \geq n_F^*$, there exists $\Gamma_n \subset X_{n,F}$ which is $(\delta^*,\, n_F^*,\, \varepsilon^*)$-separated and satisfies $\log|\Gamma_n| \geq V_n h^*$.*

Proof: If $h^* \leq 0$, we choose a generic point $x^* \in X$ for ν, and set $\Gamma_n := \{x^*\}$. This suffices for the case $h(T,\nu) = 0$.
Let $h(T,\nu) > 0$ and $0 < h^* < h(T,\nu)$. We choose h', h'' satisfying $h^* < h'' < h' < h(T,\nu)$. There exists a finite partition $\xi = \{A_1, \ldots, A_k\}$ of X with $h(T,\nu,\xi) > h'$. Define $\phi_n \colon X \to \{1,\ldots,k\}^{\Lambda_n}$ so that $\mathrm{w} := \phi_n(x)$ satisfies for all $i \in \Lambda_n$, $\mathrm{w}_i = j \Leftrightarrow T^i x \in A_j$; define $\phi \colon X \to \{1,\ldots,k\}^L$ analogously (cf. [S], §1.2). Let Y_n denote the image of X under ϕ_n; Y, the image of X under ϕ. Define $\overline{\nu} := \nu \circ \phi^{-1}$. Note that for $\mathrm{w} \in Y_n$,

$$\overline{\nu}(\mathrm{w}) := \overline{\nu}\{\omega \in Y : \omega_i = \mathrm{w}_i \,\forall\, i \in \Lambda_n\} = \nu\Big(\bigcap_{i \in \Lambda_n} T^{-i} A_{\mathrm{w}_i}\Big). \tag{2.22}$$

Then $h(T,\nu,\xi) > h'$ implies that there exists n_ξ so that whenever $n \geq n_\xi$,

$$-\sum_{\mathrm{w} \in Y_n} \overline{\nu}(\mathrm{w}) \log \overline{\nu}(\mathrm{w}) > V_n h'. \tag{2.23}$$

Since ν is regular, for $\delta > 0$ there exist compact sets, $B_j^\delta \subset A_j$, $j = 1, \ldots, k$, satisfying

$$B^\delta := \bigcup_{j=1}^k B_j^\delta, \quad \nu(B^\delta) > 1 - \frac{\delta}{2}. \tag{2.24}$$

Let F be a neighbourhood of ν. Then by the ergodic theorem

$$X_{n,F}^{B^\delta} := X_{n,F} \cap \{x \in X : \langle\, I_{B^\delta}, T_n(x)\,\rangle > 1 - \delta\}\,, \tag{2.25}$$

where I_{B^δ} denotes the indicator function of B^δ, satisfies

$$\lim_n \nu(X_{n,F}^{B^\delta}) = 1. \tag{2.26}$$

Define $\widehat{\nu}_{n,\delta}$ for $\mathrm{w} \in Y_n$ by

$$\widehat{\nu}_{n,\delta}(\mathrm{w}) := \nu\Big(\bigcap_{i \in \Lambda_n} T^{-i} A_{\omega_i} | X_{n,F}^{B^\delta}\Big) := \nu\Big(\bigcap_{i \in \Lambda_n} T^{-i} A_{\omega_i} \cap X_{n,F}^{B^\delta}\Big) / \nu(X_{n,F}^{B^\delta}). \tag{2.27}$$

Note that for any Borel set $A \subset X$,

$$H(\nu,\xi) \leq \log 2 + \nu(A) H(\nu(\,\cdot\,|A),\xi) + \nu(X \setminus A) H(\nu(\,\cdot\,|X \setminus A),\xi)\,. \tag{2.28}$$

Furthermore $H(\rho, T^{\vee n}\xi) \leq V_n \log k$ for any $\rho \in \mathcal{M}_1$. It then follows from (2.26) that there exists $n_{F,\delta} \geq n_\xi$ so that $n \geq n_{F,\delta}$ implies

$$\log|\{\mathrm{w} \in Y_n : \widehat{\nu}_{n,\delta}(\mathrm{w}) > 0\}| \geq -\sum_{\mathrm{w} \in Y_n} \widehat{\nu}_{n,\delta}(\mathrm{w}) \log \widehat{\nu}_{n,\delta}(\mathrm{w}) > V_n h''\,. \tag{2.29}$$

We define the set $\Xi_n \subset X$ by selecting for each $\mathrm{w} \in Y_n$ with $\widehat{\nu}_{n,\delta}(\mathrm{w}) > 0$ a point $x_{n,\mathrm{w}} \in X_{n,F}^{B^\delta}$ with $\phi_n(x_{n,\mathrm{w}}) = \mathrm{w}$, so that

$$\phi_n(\Xi_n) = \{\mathrm{w} \in Y_n : \widehat{\nu}_{n,\delta}(\mathrm{w}) > 0\},\ |\Xi_n| = |\phi_n(\Xi_n)|. \tag{2.30}$$

Define Γ_n to be a subset of Ξ_n of maximal cardinality satisfying

$$x \neq x' \in \Gamma_n \Longrightarrow d_n^H(\phi_n(x), \phi_n(x')) > 3\delta V_n, \tag{2.31}$$

where d_n^H denotes the Hamming distance defined by (2.14). Since B_j^δ, $j = 1, \dots, k$, are compact, there exists $\varepsilon_\delta > 0$ so that $x \in B_j^\delta$, $x' \in B_{j'}^\delta$, $j \neq j'$, implies $d(x, x') > \varepsilon_\delta$. If $x \neq x' \in \Xi_n$, (2.25) implies that the number of $i \in \Lambda_n$ for which $T^i x \notin B^\delta$ or $T^i x' \notin B^\delta$ is not more than $2\delta V_n$. Thus for $x \neq x' \in \Gamma_n$ the number of $i \in \Lambda_n$ so that $T^i x \in B^\delta$, $T^i x' \in B^\delta$ and $d(T^i x, T^i x') > \varepsilon_\delta$ is greater than $3\delta V_n - 2\delta V_n = \delta V_n$. This implies that Γ_n is $(\delta, n, \varepsilon_\delta)$- separated. The maximality of Γ_n implies that for each $x \in \Xi_n$, there exists $x' \in \Gamma_n$ so that $d_n^H(\phi_n(x), \phi_n(x')) \leq 3\delta V_n$; from (2.19) we have for $n \geq n_{F,\delta}$,

$$|\Gamma_n| \geq \exp(V_n h'')/(2^{\eta(3\delta)V_n}(k-1)^{3\delta V_n}). \tag{2.32}$$

Since $\lim_{\delta \downarrow 0} \eta(\delta) = 0$, we may choose δ^* so that

$$\eta(3\delta^*) \log 2 + 3\delta^* \log(k-1) < h'' - h^*. \tag{2.33}$$

The conclusion of the proposition holds for $n_F^* := n_{F,\delta^*}$ and $\varepsilon^* := \varepsilon_{\delta^*}$. □

The main property for the dynamical system (X, d, T), which is introduced in this paper, is the *approximate product property*: "in the large" the space of orbits of the dynamical system has almost a product structure. We give two distinct definitions which capture this basic idea. For $L = \mathbf{Z}^d$ we define the approximate L-product property; later we define the approximate product property, which is restricted to $L = \mathbf{Z}_+$.

Definition 2.9 *Let $L = \mathbf{Z}^d$ or $L = \mathbf{Z}_+^d$. The dynamical system (X, d, T) has the* ***approximate L-product property*** *if the following holds. Given any $\Delta > 0$ and $\delta > 0$, there exists $N(\Delta, \delta)$ such that for any $n \geq N(\Delta, \delta)$ and any collection $\{x^i\}_{i \in L} \subset X$, there exist a collection $\{\Lambda^i\}_{i \in L}$, a collection $\{j_i\}_{i \in L} \subset L$ and $x \in X$ verifying 1) and 2):*

1) for all $i \in L$, $\Lambda^i \subset \Lambda_n$, $\Lambda^i + j_i \subset \Lambda_n$ and $|\Lambda_n \backslash \Lambda^i| \leq \delta V_n$;

2) for all $i \in L$, $d_{\Lambda^i}(T^{iD_n + j_i} x, x^i) \leq \Delta$.

Remarks 1) Suppose that (X, d, T) has the *specification property* as defined for example in [TV2]: for any $\Delta > 0$ there exists $m(\Delta)$ such that for any finite collection of k intervals $I_j = [a_j, b_j] \subset \mathbf{Z}_+$, $j = 0, \dots, k-1$, such that $a_j - b_{j-1} \geq m(\Delta)$, $j = 1, \dots, k-1$, and any collection of k points $\{x^0, \dots, x^{k-1}\}$ there exists $x \in X$ such that

$$d(T^{a_j + p} x, T^p x^j) \leq \Delta \quad \forall\, p = 0, \dots b_j - a_j \quad \text{and} \quad \forall\, j = 0, \dots, k-1\,. \tag{2.34}$$

Then (X, d, T) has the approximate $\mathbf{Z}_+$-product property. Indeed, let $\Delta > 0$, $\delta > 0$ and N be given, and let $n > N + m(\Delta)$ so that $|\Lambda_n \setminus \Lambda_{n - m(\Delta)}| \leq \delta V_n$. Let $a_j = jD_n$ and $b_j = jD_n + n - m(\Delta)$, $j = 0, \dots, k-1$. Let $\{x^i\}_{i \in \mathbf{Z}_+}$ be given. Let $\Lambda^i = \Lambda_{n - m(\Delta)}$ and $j_i = 0$. The specification property implies that for any $k \geq 1$ there exists $x_k \in X$ so that

$$d_{\Lambda_{n-m(\Delta)}}(T^{jD_n} x_k, x^j) \leq \Delta \quad \forall\, j = 0, \dots k-1\,. \tag{2.35}$$

Since X is compact we can choose for x any accumulation point of $\{x_k\}$.
2) The case $L = \mathbf{Z}^d$ is important for statistical mechanics, where a typical situation is that X is a subshift of $\mathtt{A}^{\mathbf{Z}^d}$, the full shift over a finite alphabet $\mathtt{A}$. Often $j_i \equiv 0$ and $\Lambda^i = \Lambda_{n(1-\delta)}$ for all i. If Δ is small enough, then the condition $d_{\Lambda_{n(1-\delta)}}(T^{iD_n}x, x^i) \leq \Delta$ implies that the restriction of the configuration x to $iD_n + \Lambda_{n(1-\delta)}$ is equal to the restriction of x^i to $\Lambda_{n(1-\delta)}$ (see section 4).
3) The case $L = \mathbf{Z}_+$ is important for dynamical systems. It is convenient to define the following variant of Definition 2.9. Instead of placing partial orbits somewhere inside regularly spaced segments, one places the partial orbits with variable gaps between.

Definition 2.10 *Let $L = \mathbf{Z}_+$. The dynamical system (X, d, T) has the **approximate product property** if the following holds. Given any $\Delta > 0$, $\delta_1 > 0$ and $\delta_2 > 0$, there exists $N(\Delta, \delta_1, \delta_2)$ such that for any $n \geq N(\Delta, \delta_1, \delta_2)$ and for any collection $\{x^i\}_{i \in L} \subset X$, there exist a collection $\{\Lambda^i\}_{i \in L}$, a collection $\{j_i\}_{i \in L} \subset L$ and $x \in X$ verifying 1) and 2):*

1) for all $i \in L$, $\Lambda^i \subset \Lambda_n$ and $|\Lambda_n \backslash \Lambda^i| \leq \delta_1 V_n$;

2) $j_0 = 0$, $n \leq j_{i+1} - j_i \leq n(1 + \delta_2)$ and for all $i \in L$, $d_{\Lambda^i}(T^{j_i}x, x^i) \leq \Delta$.

It is convenient to introduce δ_1 and δ_2 instead of a single δ, because they play different roles. The sizes of gaps between the specified orbit-segments are controlled by δ_2; the orbit-segments are only Δ-shadowed on subsets of Λ_n whose sizes are controlled by δ_1. Borrowing a terminology from [S] we say that the orbit $\{T^k x : k \in L\}$ is $(1 - \delta_2)$**- built-up from partial orbits** $\{T^{j_i + k}x : k \in \Lambda_n\}$, $i \in L$, which approximately Δ-shadow the given finite orbits $\{T^k x^i : k \in \Lambda_n\}$, $i \in L$, of length D_n. The case $L = \mathbf{Z}$ can be treated similarly.

Proposition 2.2 *Let (X, d, T) and (Y, d', S) be dynamical systems. Let $\phi\colon X \to Y$ be a continuous surjection satisfying $\phi \circ T = S \circ \phi$. If (X, d, T) has the approximate product property, then so does (Y, d', S). In particular the approximate product property does not depend on the choice of metric on X as long as the topology of X remains the same. The same applies for the approximate L-product property.*

Proof: Assume that X satisfies the approximate L-product property. Let $\Delta' > 0$, $\delta > 0$ and N be given. The function $d'(\phi(x), \phi(x^*))$ from $X \times X$ to $\mathbf{R}$ is continuous; by compactness it is uniformly continuous. It follows that there exists $\Delta > 0$ so that $d(x, x^*) < \Delta$ implies $d'(\phi(x), \phi(x^*)) < \Delta'$. Take $n \geq N$ corresponding to Δ and δ. For any collection $\{y^i\}_{i \in L}$ with each $y^i \in Y$ choose $\{x^i\}_{i \in L}$ so that $\phi(x^i) = y^i$ for each $i \in L$. There exist $\{\Lambda^i\}_{i \in L}$, $\{j_i\}_{i \in L}$ and $x \in X$ so that $d_{\Lambda_i}(T^{iD_N + j_i}x, x^i) < \Delta$ for all $i \in L$. Then with $y := \phi(x)$, $d'_{\Lambda_i}(T^{iD_N + j_i}y, y^i) < \Delta'$ for all $i \in L$. □
We next state useful elementary lemmas.

Lemma 2.3 *Let $f : X \to \mathbf{R}$ satisfy $d(x, y) \leq \Delta \Rightarrow |f(x) - f(y)| \leq \varepsilon$. For $\Lambda \subset \Lambda_n$, if $d_\Lambda(x, y) \leq \Delta$, then*

$$|\langle\, f, \mathcal{E}_n(x) - \mathcal{E}_n(y)\,\rangle| \leq \varepsilon + 2\|f\| \frac{V_n - |\Lambda|}{V_n}. \tag{2.36}$$

Proof: We write $\langle\, f, \mathcal{E}_n(x) - \mathcal{E}_n(y)\,\rangle$ as

$$\sum_{j \in \Lambda} \frac{1}{V_n}\Big(f(T^j x) - f(T^j y)\Big) + \sum_{j \in \Lambda_n \backslash \Lambda} \frac{1}{V_n}\Big(f(T^j x) - f(T^j y)\Big). \tag{2.37}$$

The first summation is bounded by ε; the second by $2\|f\|(|\Lambda_n| - |\Lambda|)/V_n$. □

Lemma 2.4 *For $f : X \to \mathbf{R}$, $\nu \in \mathcal{M}_1$ and $m \geq n \in \mathbf{N}$, if $|\langle f, \mathcal{E}_n(x) - \nu \rangle| \leq \varepsilon$, then*

$$|\langle f, \mathcal{E}_m(x) - \nu \rangle| \leq \varepsilon + 2\|f\| \frac{V_m - V_n}{V_m}. \tag{2.38}$$

Proof: We write $\langle f, \mathcal{E}_m(x) - \nu \rangle$ as

$$\frac{V_n}{V_m} \langle f, \mathcal{E}_n(x) - \nu \rangle + \sum_{j \in \Lambda_m \setminus \Lambda_n} \frac{1}{V_m} \langle f, \delta_{T^j x} - \nu \rangle. \tag{2.39}$$

The first summation is bounded by ε; the second by $2\|f\|(|\Lambda_m| - |\Lambda_n|)/V_m$. □

Lemma 2.5 *Let $f : X \to \mathbf{R}$ and let $g, m \in \mathbf{Z}_+$. Then for any $j \in \Lambda_g$ and any $x \in X$,*

$$|\langle f, \mathcal{E}_m(x) - \mathcal{E}_m(T^j x) \rangle| \leq 2\|f\| \frac{V_m - V_{m-g}}{V_m}. \tag{2.40}$$

Proof: Recall $V_{m-g} := 0$ for $g > m$. One can match identical terms of $\mathcal{E}_m(x)$ with ones of $\mathcal{E}_m(T^j x)$ on $\Lambda_m \cap (j + \Lambda_m)$. For $j \in \Lambda_g$ and $k \in \Lambda_{m-g}$ we have $j + k \in \Lambda_m$. The bound (2.40) results from terms indexed by points of $(\Lambda_m \setminus (j + \Lambda_m)) \cup ((j + \Lambda_m) \setminus \Lambda_m)$. □

Corollary 2.1 *Let F be an f-neighbourhood of $\nu \in \mathcal{M}_1$. For $\alpha' > \alpha > 0$ and $j \in \Lambda_g$, whenever $x \in X_{m,F(\alpha)}$, then $T^j x \in X_{m,F(\alpha')}$ provided m is sufficiently large that*

$$2\frac{V_m - V_{m-g}}{V_m} \leq (\alpha' - \alpha)\varepsilon_{\min}(F). \tag{2.41}$$

Proof: When $x \in X_{m,F(\alpha)}$, for each f_i of the functions in the definition of F we have

$$|\langle f_i, \mathcal{E}_m(x) - \nu \rangle| \leq \alpha\, \varepsilon_i. \tag{2.42}$$

From (2.40) and (2.41) we have

$$|\langle f_i, \mathcal{E}_m(T^j x) - \nu \rangle| \leq \alpha\varepsilon_i + (\alpha' - \alpha)\varepsilon_{\min}(F) \leq \alpha'\varepsilon_i. \tag{2.43}$$

□

We use $h_{\mathrm{top}}(X, T)$ to denote the topological entropy of (X, d, T). See [W] for details. The next proposition is inspired by the proof of Theorem B in [EKW].

Proposition 2.3 *Suppose that (X, d, T) has the approximate L-product property or approximate product property and that $\nu \in \mathcal{M}_1^T$ verifies the conclusions of Proposition 2.1. Let $0 < h' < h(T, \nu)$. Then, there exists $\varepsilon' > 0$ such that for any neighbourhood G of ν, there exists a closed T-invariant subspace $Y \subset X$ with the following properties.*

1) *There exists $n'_G \in \mathbf{N}$ such that $\mathcal{E}_n(y) \in G$ whenever $n \geq n'_G$ and $y \in Y$.*

2) *There exists $n''_G \in \mathbf{N}$ such that for all $\ell \in \mathbf{N}$ there exists a subset of Y, which is $(\ell * n''_G, \varepsilon')$-separated and has cardinality greater than $\exp(V_{\ell * n''_G} h')$.*

In particular $h_{\mathrm{top}}(Y, T) \geq h'$.

Proof: We treat the approximate L-product case; the approximate product case uses similar ideas. Take $h' < h^* < h(T,\nu)$; given the neighbourhood G of $\nu \in \mathcal{M}_1^T$, take an f-neighbourhood of ν, $F \subset G$, with fixed $\{(f_i,\varepsilon_i) \colon i = 1,\ldots,k\}$. Let δ^*, $n^* := n^*_{F^{(1/5)}}$ and ε^*, be the parameters in Proposition 2.1 for the neighbourhood $F^{(1/5)}$. Whenever $n \geq n^*$, one can find a subset Γ_n, which is $(\delta^*, n, \varepsilon^*)$-separated and satisfies

$$\Gamma_n \subset X_{n,F^{(1/5)}} \;\;,\;\; \log|\Gamma_n| \geq h^* V_n\,. \tag{2.44}$$

Because the f_i which define F are uniformly continuous on X, there exists $\Delta > 0$ so that

$$\Delta < \frac{\varepsilon^*}{3}, \quad d(x,y) \leq \Delta \Longrightarrow |f_i(x) - f_i(y)| \leq \frac{\varepsilon_i}{5} \tag{2.45}$$

for each $(f_i,\,\varepsilon_i)$ associated with F. With η defined by (2.16), we choose $\delta \leq \delta^*$ so that $0 < \delta < 1/2$ and

$$1)\;4\delta \leq \frac{\varepsilon_{\min(F)}}{5}, \quad 2)\;\eta(\delta) < \frac{h^* - h'}{2\log 2}\,. \tag{2.46}$$

There exists $n^{\#}$ satisfying

$$n^{\#} \geq \max\{N(\Delta,\delta),\, n^*\}, \;\; n \geq n^{\#} \Longrightarrow \frac{\log V_n}{V_n} < \frac{h^* - h'}{2}\,. \tag{2.47}$$

Note that for $n \geq n^{\#}$ and the chosen Δ and δ, the consequences of Definition 2.9 obtain. Consider $Z^{\#}_{F,n} \subset X$ defined by the requirement that $x \in Z^{\#}_{F,n}$ if and only if for all $i \in L$, there exist

$$x^i \in X_{n,F^{(1/5)}},\; j_i \text{ and } \Lambda^i \subset \Lambda_n \text{ with } \Lambda^i + j_i \subset \Lambda_n,\; |\Lambda_n \setminus \Lambda^i| \leq \delta V_n, \tag{2.48}$$

such that for all $i \in L$,

$$d_{\Lambda^i}(T^{iD_n + j_i}x, x^i) \leq \Delta. \tag{2.49}$$

Since $X_{n,F^{(1/5)}} \neq \emptyset$, the approximate L-product property implies that $Z^{\#}_{F,n}$ is nonempty and

$$x \in Z^{\#}_{F,n},\, k \in L \Longrightarrow T^{kD_n}x \in Z^{\#}_{F,n}. \tag{2.50}$$

Let $x \in Z^{\#}_{F,n}$. We now show that for each $i \in L$, $T^{iD_n}x \in X_{n,F^{(3/5)}}$. First using Lemma 2.3 and (2.49), we go from $x^i \in X_{n,F^{(1/5)}}$, to $T^{iD_n+j_i}x \in X_{n,F^{(\alpha_1)}}$. Then the method of Lemma 2.5, noting that $|\Lambda_n \cap (j_i + \Lambda_n)| \geq (1-\delta)V_n$, is used to show $T^{iD_n}x \in X_{n,F^{(\alpha_2)}}$. From (2.45) and (2.46) it follows that we can have $\alpha_2 = 3/5$. It follows from (2.4) that if $x \in Z^{\#}_{F,n}$, $k \in \mathbf{Z}_+$ and $i \in L$, then $T^{iD_n}x \in X_{k*n,F^{(3/5)}}$. We now show that there exists k_n so that $k \geq k_n$ and $x \in Z^{\#}_{F,n}$ imply $T^j x \in X_{k*n,F^{(4/5)}}$ for all $j \in L$. Only the case of $T^j x$ for $j \in \Lambda_n$ remains. Corollary 2.1 implies the existence of k_n so that $k \geq k_n$, $j \in \Lambda_n$ and $x \in X_{k*n,F^{(3/5)}}$ implies $T^j x \in X_{k*n,F^{(4/5)}}$.
Let $k \geq k_n$; define

$$Y_{k,n} := \Big\{x \in X : T^j x \in X_{k*n,F^{(4/5)}}\, \forall\, j \in L\Big\}\,. \tag{2.51}$$

We have shown that $Z^{\#}_{F,n} \subset Y_{k,n}$. Each such $Y_{k,n}$ is a closed T-invariant subspace of X. From (2.4) we deduce that $x \in Y_{k,n}$ implies $x \in Y_{j*k,n}$ for all $j \in \mathbf{Z}_+$. Now

$$\lim_{j\to\infty} \frac{V_{(j+1)*m} - V_{j*m}}{V_{(j+1)*m}} = 0, \tag{2.52}$$

so Lemma 2.4 implies the existence of n'_G so that $m \geq n'_G$ and $x \in Y_{k,n}$ imply $\mathcal{E}_m(x) \in F$. This proves statement 1).
We prove statement 2). We set $n := n''_G := n^{\#}$ from (2.47) and $\varepsilon' := \varepsilon^*/3$. We consider the following subset $Z^{\flat}_{F,n} \subset Z^{\#}_{F,n}$: for each $\{x^i\}_{i\in L}$, with $x^i \in \Gamma_n$ for all $i \in L$, we construct a point x using the approximate L-product property; the set $Z^{\flat}_{F,n}$ is the set of these constructed points. Let $\ell \in \mathbf{N}$ and $\Gamma_n^{\Lambda_\ell}$ be the cartesian product of V_ℓ copies of Γ_n. For each element $\{x^i\} \in \Gamma_n^{\Lambda_\ell}$ we select one point x of $Z^{\flat}_{F,n}$ and one collection $\mathcal{C} := \{(\Lambda^i, j_i) : i \in \Lambda_\ell\}$ so that

$$d_{\Lambda^i}(T^{iD_n+j_i}x, x^i) \leq \Delta \ \forall\, i \in \Lambda_\ell\,. \tag{2.53}$$

We define

$$\phi_\ell(\{x^i\}) := x,\ \widehat{\phi}_\ell(\{x^i\}) := \mathcal{C}, \tag{2.54}$$

and

$$E_\ell(\mathcal{C}) := \{\{x^i\} \in \Gamma_n^{\Lambda_\ell} : \widehat{\phi}_\ell(\{x^i\}) := \mathcal{C}\}. \tag{2.55}$$

For some $\mathcal{C}$, $E_\ell(\mathcal{C})$ may be empty, but

$$\bigcup_{\mathcal{C}} E_\ell(\mathcal{C}) = \Gamma_n^{\Lambda_\ell}, \tag{2.56}$$

where the union is over all admissible collections $\mathcal{C}$. Assume for a specific $\mathcal{C}$, $E_\ell(\mathcal{C}) \neq \emptyset$. Since Γ_n is $(\delta^*, n, \varepsilon^*)$-separated and $\delta \leq \delta^*$, if $\{x^i\}, \{y^i\} \in E_\ell(\mathcal{C})$ and $x^{i'} \neq y^{i'}$, $d_{\Lambda^{i'}}(x^{i'}, y^{i'}) > \varepsilon^*$. Because $\{\Lambda^i, j_i\}$ is the same for $\{x_i\}$ and $\{y_i\}$, (2.45) implies that

$$d_{\Lambda_{\ell*n}}(\phi_\ell(\{x^i\}), \phi_\ell(\{y^i\}) > \varepsilon^*/3. \tag{2.57}$$

Thus whenever $E_\ell(\mathcal{C}) \neq \emptyset$, $\phi_\ell(E_\ell(\mathcal{C}))$ is an $(\ell * n, \varepsilon^*/3)$ separated subset of $Z^{\#}_{F,n}$ with cardinality equal to that of $E_\ell(\mathcal{C})$.
We can bound the number of admissible $\mathcal{C}$ as follows. The number of $\Lambda \subset \Lambda_n$ with $|\Lambda_n \setminus \Lambda| \leq \delta V_n$ is not more than $2^{V_n\eta(\delta)}$. The number of possible values for j_i for fixed i is bounded by V_n. We deduce that the number of admissible collections $\mathcal{C}$ is bounded above by $2^{V_\ell V_n \eta(\delta)} V_n^{V_\ell}$. Because of (2.56), if $\mathcal{C}^*$ maximizes the cardinality of $E_\ell(\mathcal{C})$,

$$|E_\ell(\mathcal{C}^*)| \geq |\Gamma_n|^{V_\ell} / \left(2^{V_\ell V_n \eta(\delta)} V_n^{V_\ell}\right). \tag{2.58}$$

Then $\phi_\ell(E_\ell(\mathcal{C}^*))$ is $(\ell * n, \varepsilon^*/3)$-separated with

$$\frac{1}{V_{\ell*n}} \log|\phi_\ell(E_\ell(\mathcal{C}^*))| \geq h^* - \eta(\delta)\log 2 - \frac{\log V_n}{V_n} \geq h' \tag{2.59}$$

because of (2.46) and (2.47). □

Corollary 2.2 *Under the hypothesis of Proposition 2.3, ν is entropy-approachable by ergodic measures.*

Proof: For any neighbourhood G of ν and $h^* < h(T,\nu)$, let $F \subset G$ be an f-neighbourhood of ν. Let Y be a closed T-invariant subspace of X such that $h_{\text{top}}(Y,T) > h^*$ and $\mathcal{E}_n(y) \in F$ if $y \in Y$ and $n \geq n'_G$. Then by the argument of Corollary 8.6.1 of [W], there exists an ergodic ρ with $\rho(Y) = 1$ and $h(T,\rho) > h^*$. Let $y \in Y$ be a generic point for ρ. Since $\{\mathcal{E}_n(y)\}$ converges to ρ and for $n \geq n'_G$, $\mathcal{E}_n(y) \in F$, we have $\rho \in F \subset G$. □

Lemma 2.6 *If X, $\widehat{\nu} \in \mathcal{M}_1^T$ and $\widetilde{\nu} \in \mathcal{M}_1^T$ satisfy the hypothesis of Proposition 2.3, then $\nu := 0.5(\widehat{\nu} + \widetilde{\nu})$ is entropy-approachable by ergodic measures.*

Proof: Note $h(T,\nu) = 0.5(h(T,\widehat{\nu}) + h(T,\widetilde{\nu}))$. Given the f-neighbourhood F of ν, consider the corresponding f-neighbourhoods, $\widehat{F}$ and $\widetilde{F}$ of $\widehat{\nu}$ and $\widetilde{\nu}$ with the same (f_i, ε_i). For $h^* < h(T,\nu)$, we select $\widehat{h}^* < h(T,\widehat{\nu})$ and $\widetilde{h}^* < h(T,\widetilde{\nu})$ so that $h^* = 0.5(\widehat{h}^* + \widetilde{h}^*)$. We can find $(\widehat{\delta}^*, \widehat{n}^*_{\widehat{F}(1/5)}, \widehat{\varepsilon}^*)$ for $\widehat{\nu}$ and $(\widetilde{\delta}^*, \widetilde{n}^*_{\widetilde{F}(1/5)}, \widetilde{\varepsilon}^*)$ for $\widetilde{\nu}$ from Proposition 2.1. We choose $(\delta^*, n^*, \varepsilon^*)$, so that n^* is the larger of the n's; for δ^* and ε^* we take the smaller. We say that $i \in L \subset \mathbf{Z}^d$ is **even** if the sum of the integers in the vector i is even. Next we consider $Z^{\#}_{F,n}$ defined as for (2.48) except that $x^i \in X_{n,\widehat{F}(1/5)}$ for i even, but $x^i \in X_{n,\widetilde{F}(1/5)}$ otherwise. The proof then follows that of Proposition 2.3 and the Corollary with straightforward modifications. □

Corollary 2.3 *If (X, d, T) has the approximate product property, then the set of measures in $\mathcal{M}_1^T$ which are entropy-approachable by ergodic measures is closed under finite convex combinations.*

Theorem 2.1 *If (X, d, T) has the approximate product property, then the ergodic measures are entropy-dense.*

Proof: The above shows that finite convex combinations of ergodic measures are entropy-approachable by ergodic measures. One then uses the ergodic decomposition of each $\nu \in \mathcal{M}_1^T$ to show that it can be approximated by convex combinations of ergodic measures with entropy close to that of ν. See e.g., [EKW]. □

3 Large deviations

Definition 3.1 *For $x \in X$, $\Lambda \subset L$ and $\varepsilon > 0$ we define the open and closed Λ-balls in X by*

$$\begin{aligned} B_\Lambda(x,\varepsilon) &:= \{y \in X : d_\Lambda(x,y) < \varepsilon\}; \\ \overline{B}_\Lambda(x,\varepsilon) &:= \{y \in X : d_\Lambda(x,y) \leq \varepsilon\}. \end{aligned} \tag{3.1}$$

We use the notations $B_n(x,\varepsilon)$ and $\overline{B}_n(x,\varepsilon)$ when $\Lambda = \Lambda_n$.

Counting arguments combined with estimates concerning $\nu(B_n(x,\varepsilon))/\nu(B_n(y,\varepsilon))$ yield large deviation consequences. The following extends ideas in [Y] and [PfS2].

Definition 3.2 *The function $f : X \to \mathbf{R}$ is* ***upper semicontinuous*** *if*

$$\{x \in X : f(x) < a\} \tag{3.2}$$

is open for each $a \in \mathbf{R}$. Let $\nu \in \mathcal{M}_1$. We say that $e_\nu : X \to \mathbf{R}^+$ is a lower-energy function for ν if e_ν is upper semicontinuous and

$$\lim_{\varepsilon \downarrow 0} \liminf_n \inf_{x \in X} \left(\frac{1}{V_n} \log \nu(B_n(x,\varepsilon)\,) + \langle\, e_\nu, \mathcal{E}_n(x)\,\rangle \right) \geq 0. \tag{3.3}$$

Proposition 3.1 *Let G be an open subset of $\mathcal{M}_1$ and let $e_\nu : X \to \mathbf{R}^+$ be a lower-energy function for ν. Then for any ergodic $\rho \in G$,*

$$\liminf_n \frac{1}{V_n} \log \nu(\{x : \mathcal{E}_n(x) \in G\}) \geq h(T,\rho) - \langle\, e_\nu, \rho \,\rangle \tag{3.4}$$

Proof: Since e_ν is upper semicontinuous, the function $\mu \mapsto \langle\, e_\nu, \mu \,\rangle$ is upper semicontinuous on $\mathcal{M}_1$ (see the appendix of [Schw]). Then for $\delta > 0$ the set

$$G^{\nu,\delta} := G \cap \{\mu \in \mathcal{M}_1 :\, \langle\, e_\nu, \mu \,\rangle < \langle\, e_\nu, \rho \,\rangle + \delta\} \tag{3.5}$$

is open and contains an f-neighbourhood of ρ which we denote by F. By Proposition 2.1, for any $h^* < h(T,\rho)$ there exist constants n^* and ε^* so that for any $n \geq n^*$ there is an (n,ε^*)-separated subset $\Gamma_n \subset X_{n,F^{(1/2)}}$ with $\log|\Gamma_n| \geq V_n\, h^*$. Since the functions f_i which belong to F are uniformly continuous, there exists $\delta^* > 0$ so that $d(x,y) \leq \delta^*$ implies $|f_i(x) - f_i(y)| \leq \varepsilon_i/2$ for each i associated with F. Define $\varepsilon' := \min\{\varepsilon^*/2, \delta^*\}$. Then for $n \geq n^*$ and $0 < \varepsilon \leq \varepsilon'$, the set $\Gamma_n^\varepsilon := \Gamma_n$ is $(n, 2\varepsilon)$-separated, hence $x \neq y \in \Gamma_n^\varepsilon \Rightarrow B_n(x,\varepsilon) \cap B_n(y,\varepsilon) = \emptyset$ and

$$\log|\Gamma_n^\varepsilon| \geq V_n\, h^*, \qquad \bigcup_{x\in\Gamma_n^\varepsilon} B_n(x,\varepsilon) \subset X_{n,F}. \tag{3.6}$$

Then for $n \geq n^*$, $0 < \varepsilon \leq \varepsilon'$, since $F \subset G$,

$$\log \nu(\{x : \mathcal{E}_n(x) \in G\}) \geq V_n h^* + \inf_{x\in X_{n,F}} \log \nu(B_n(x,\varepsilon)\,); \tag{3.7}$$

$$\begin{aligned}\frac{1}{V_n}\log \quad & \nu(\{x : \mathcal{E}_n(x) \in G\}) \geq h^* \\ & + \inf\nolimits_{x\in X_{n,F}} \Big(\tfrac{1}{V_n}\log \nu(B_n(x,\varepsilon)) + \langle\, e_\nu, \mathcal{E}_n(x) - \rho \,\rangle - \delta\Big).\end{aligned} \tag{3.8}$$

We deduce from (3.3) that

$$\liminf_n \frac{1}{V_n} \log \nu(\{x : \mathcal{E}_n(x) \in G\}) \geq h^* - \langle\, e_\nu, \rho \,\rangle - \delta. \tag{3.9}$$

Since $\delta > 0$ and $h^* < h(T,\rho)$ are arbitrary, we have (3.4). □

Definition 3.3 *Let $\nu \in \mathcal{M}_1$. We say that $e_\nu : X \to \mathbf{R}^+$ is a* ***weak lower-energy function for*** *ν if e_ν is upper semicontinuous and for each ergodic $\rho \in \mathcal{M}_1^T$, each $h^* < h(T,\rho)$, each $\delta > 0$ and each neighbourhood G of ρ there exists $\varepsilon > 0$ and n_ε so that for $n \geq n_\varepsilon$ the set*

$$\Big\{x \in X : \mathcal{E}_n(x) \in G, \frac{1}{V_n}\log \nu(B_n(x,\varepsilon)\,) + \langle\, e_\nu, \mathcal{E}_n(x) \,\rangle > -\delta\Big\} \tag{3.10}$$

has an (n,ε)-separated subset Γ_n with $\log|\Gamma_n| \geq V_n h^*$.

Note that if e_ν is lower-energy function, then

$$\log \nu(B_n(x,\varepsilon)\,) + V_n \langle\, e_\nu, \mathcal{E}_n(x) \,\rangle > -V_n\delta \tag{3.11}$$

for all $x \in X$ when n is sufficiently large. For the weak lower-energy case we only need the existence of separated sets of sufficiently large cardinality.

We observe that the proof of Proposition 3.1 with minor alterations yields the following.

Proposition 3.2 *Let G be an open subset of $\mathcal{M}_1$ and let $e_\nu : X \to \mathbf{R}^+$ be a weak lower-energy function for ν. Then for any ergodic $\rho \in G$,*

$$\liminf_n \frac{1}{V_n} \log \nu(\{x : \mathcal{E}_n(x) \in G\}) \geq h(T,\rho) - \langle\, e_\nu, \rho\,\rangle. \tag{3.12}$$

Theorem 3.1 *Let G be an open subset of $\mathcal{M}_1$ and let $e_\nu : X \to \mathbf{R}^+$ be a weak lower-energy function for ν. Then for any $\rho \in G \cap \mathcal{M}_1^T$ which is entropy-approachable by ergodic measures,*

$$\liminf_n \frac{1}{V_n} \log \nu(\{x : \mathcal{E}_n(x) \in G\}) \geq h(T,\rho) - \langle\, e_\nu, \rho\,\rangle. \tag{3.13}$$

If the ergodic measures of (X,d,T) are entropy-dense, then

$$\liminf_n \frac{1}{V_n} \log \nu(\{x : \mathcal{E}_n(x) \in G\}) \geq \sup_{\rho \in G \cap \mathcal{M}_1^T} \Big(h(T,\rho) - \langle\, e_\nu, \rho\,\rangle\Big). \tag{3.14}$$

Proof: When $\rho \in \mathcal{M}_1^T$ is entropy-approachable by ergodic measures, for each neighbourhood G of ρ, each $h^* < h(T,\rho)$ and each $\delta > 0$, $G^{\rho,\delta}$, defined in the same way as (3.5), contains in its interior an ergodic $\mu \in \mathcal{M}_1$ satisfying $h(T,\mu) > h^*$, so we may apply the previous proposition and take limits. □

Next we examine the large deviation upper bound. We do not consider a weak variant in this case.

Definition 3.4 *Let $\nu \in \mathcal{M}_1$. We say that $e_\nu : X \to \mathbf{R}^+$ is an **upper-energy function for** ν if $-e_\nu$ is upper semicontinuous, bounded and*

$$\lim_{\varepsilon \downarrow 0} \limsup_n \sup_{x \in X} \left(\frac{1}{V_n} \log \nu(B_n(x,\varepsilon)\,) + \langle\, e_\nu, \mathcal{E}_n(x)\,\rangle\right) \leq 0. \tag{3.15}$$

Definition 3.5 *For $D \subset \mathcal{M}_1$, $\overline{D}$ denotes the closure of D; $\overline{\overline{D}}$ denotes the smallest closed convex set containing D.*

Definition 3.6 *For $D \subset \mathcal{M}_1$ we use the notation $s_n(\varepsilon, D)$ to denote the maximum cardinality of an (n,ε)-separated subset of $\{x \in X : \mathcal{E}_n(x) \in D\}$; then we define*

$$s(\varepsilon, D) := \limsup_n \frac{1}{V_n} \log s_n(\varepsilon, D). \tag{3.16}$$

Lemma 3.1 *Let D be a nonempty subset of $\mathcal{M}_1$. Then for $\varepsilon > 0$,*

$$s(\varepsilon, D) \leq \sup_{\rho \in \overline{\overline{D}} \cap \mathcal{M}_1^T} h(T,\rho). \tag{3.17}$$

If $h(T,\cdot\,)$ is upper semicontinuous on $\mathcal{M}_1^T$, then

$$s(\varepsilon, D) \leq \sup_{\rho \in \overline{D} \cap \mathcal{M}_1^T} h(T,\rho). \tag{3.18}$$

Proof: The argument uses techniques from [Mi]. Let $\Gamma_n \subset \{x \in X : \mathcal{E}_n(x) \in D\}$ be a sequence of (ε, n)-separated subsets with $|\Gamma_n| = s_n(\varepsilon, D)$. Define

$$\sigma_n := \frac{1}{s_n(\varepsilon, D)} \sum_{x \in \Gamma_n} \delta_x, \quad \mu_n := \frac{1}{s_n(\varepsilon, D)} \sum_{x \in \Gamma_n} \mathcal{E}_n(x). \tag{3.19}$$

Since $\mathcal{M}_1$ is compact, there exists a subsequence $\{n_g\}$ so that $\{\log s_{n_g}(\varepsilon, D)/V_{n_g}\}$ converges to $s(\varepsilon, D)$ and $\{\mu_{n_g}\}$ converges to ρ. It follows that $\rho \in \mathcal{M}_1^T$. Since each μ_n is a convex combination of measures in D, it follows that $\rho \in \overline{\overline{D}}$. Select a partition $\xi := \{B_1, \ldots, B_k\}$ such that $x, y \in B_i$ imply $d(x, y) < \varepsilon$ and $\rho(\partial B_i) = 0$ for $i = 1, \ldots, k$. Note that

$$H(\sigma_n, T^{\vee n}\xi) = \log s_n(\varepsilon, D) \tag{3.20}$$

because each atom of $T^{\vee n}\xi$ contains at most one point of Γ_n. For $0 \le m \le n$ and $j \in \Lambda_m$ we define $q_{n,m,j}$ to be the unique element of $\mathbf{Z}_+$ so that

$$j + \Lambda_{m * q_{n,m,j}} \subset \Lambda_n, \; j + \Lambda_{m*(q_{n,m,j}+1)} \not\subset \Lambda_n. \tag{3.21}$$

For each m, n, j we have

$$T^{\vee n}\xi = \bigvee_{i \in \Lambda_{q_{n,m,j}}} T^{-(iD_m + j)} T^{\vee m}\xi \quad \bigvee_{i \in \Lambda_n \setminus (j + \Lambda_{m*q_{n,m,j}})} T^{-i}\xi. \tag{3.22}$$

Then

$$H(\sigma_n, T^{\vee n}\xi) \le \sum_{i \in \Lambda_{q_{n,m,j}}} H(\sigma_n \circ T^{-(iD_m+j)}, T^{\vee m}\xi) + (V_n - V_{m*q_{n,m,j}}) \log k, \tag{3.23}$$

where k is the number of elements of the partition ξ. Summing over $j \in \Lambda_m$ we deduce

$$V_m H(\sigma_n, T^{\vee n}\xi) \le \sum_{j \in \Lambda_{n-m}} H(\sigma_n \circ T^{-j}, T^{\vee m}\xi) + V_m(V_n - V_{n-m}) \log k, \tag{3.24}$$

because $\sum_{j \in \Lambda_m} \sum_{i \in \Lambda_{q_{n,m,j}}} T^{-(iD_m+j)}$ equals $\sum_{j \in \Lambda_{n-m}} T^{-j}$. Now

$$H(\mu_n, T^{\vee m}\xi) = H(\mathcal{A}_n \sigma_n, T^{\vee m}\xi) \ge$$

$$\frac{1}{V_n} \sum_{j \in \Lambda_n} H(\sigma_n \circ T^{-j}, T^{\vee m}\xi) \ge \frac{1}{V_n} \sum_{j \in \Lambda_{n-m}} H(\sigma_n \circ T^{-j}, T^{\vee m}\xi). \tag{3.25}$$

Now for each atom B of $T^{\vee m}\xi$, we have $\lim_g \mu_{n_g}(B) = \rho(B)$ because $\rho(\partial T^{-j} B_i) = 0$ for each $j \in L$ and $B_i \in \xi$. Hence

$$\lim_{g \to \infty} H(\mu_{n_g}, T^{\vee m}\xi) = H(\rho, T^{\vee m}\xi). \tag{3.26}$$

We have $\lim_n (V_n - V_{n-m})/V_n = 0$. Using (3.20), (3.24), (3.25) and taking the limit over $\{n_g\}$, we deduce

$$H(\rho, T^{\vee m}\xi) \ge V_m s(\varepsilon, D). \tag{3.27}$$

Then dividing by V_m and taking the limit $m \to \infty$, we have

$$h(T, \rho, \xi) \ge s(\varepsilon, D), \tag{3.28}$$

from which $h(T,\rho) \geq s(\varepsilon, D)$ follows.
Next we consider the case in which $h(T,\,\cdot\,)$ is upper semicontinuous on $\mathcal{M}_1^T$. For each $\delta > 0$ there exists a finite cover of D by balls $\{B(x^i,\delta)\}$ of radius δ. For at least one i,

$$\limsup_n \frac{1}{V_n} \log s_n(\varepsilon, D \cap B(x^i,\delta)) = s(\varepsilon, D). \tag{3.29}$$

The above construction yields a $\rho_\delta \in \mathcal{M}_1^T \cap \overline{B}(x^i,\delta)$ with $h(T,\rho_\delta) \geq s(\varepsilon, D)$. Taking a sequence $\{\delta_j > 0\}$ with $\delta_j \to 0$, we find a limit point $\rho \in \overline{D}$. The upper semicontinuity of $h(T,\,\cdot\,)$ implies that $h(T,\rho) \geq s(\varepsilon, D)$. □

Theorem 3.2 *Let F be a closed convex subset of $\mathcal{M}_1$ and let $e_\nu : X \to \mathbf{R}^+$ be an upper-energy function for ν. Then*

$$\limsup_n \frac{1}{V_n} \log \nu(\{x : \mathcal{E}_n(x) \in F\}) \leq \sup_{\rho \in F \cap \mathcal{M}_1^T} \Big(h(T,\rho) - \langle\, e_\nu, \rho \,\rangle \Big). \tag{3.30}$$

If $h(T,\,\cdot\,)$ is upper semicontinuous, the above obtains even if F is not convex.

Proof: Since e_ν is bounded, given $\delta > 0$, we can select constants $0 = a_0 < a_1 < \ldots < a_k$ with $a_k := \sup_x e_\nu(x)$ so that $a_j - a_{j-1} < \delta$, $j = 1, \ldots, k$. For $j = 1, \ldots, k$ define

$$F_j := \{\rho \in F : a_{j-1} \leq \langle\, e_\nu, \rho \,\rangle \leq a_j\}, \tag{3.31}$$

which is convex when F is convex. Then in the notation (2.6)

$$\limsup_n \frac{1}{V_n} \log \nu(X_{n,F}) = \max_{j=1,\ldots,k} \limsup_n \frac{1}{V_n} \log \nu(X_{n,F_j}). \tag{3.32}$$

For sufficiently small $\varepsilon > 0$ there exists n_ε so that $n \geq n_\varepsilon$ implies that for all $x \in X$,

$$\frac{1}{V_n} \log \nu(B_n(x,\varepsilon)\,) + \langle\, e_\nu, \mathcal{E}_n(x) \,\rangle \leq \delta. \tag{3.33}$$

If Γ is a maximal (n,ε)-separated subset of Y, then $Y \subset \bigcup_{x\in\Gamma} B_n(x,\varepsilon)$, so for $n \geq n_\varepsilon$,

$$\frac{1}{V_n} \log \nu(X_{n,F_j}) \leq \frac{1}{V_n} \left(\log s_n(\varepsilon, F_j) + \sup_{x\,:\,\mathcal{E}_n(x)\in F_j} \log \nu(B_n(x,\varepsilon)) \right). \tag{3.34}$$

$$\frac{1}{V_n} \log \nu(X_{n,F_j}) \leq \frac{1}{V_n} \log s_n(\varepsilon, F_j) - \inf_{\rho\in F_j} \langle\, e_\nu, \rho \,\rangle + \delta. \tag{3.35}$$

Using $\overline{F}_j \subset F$ to denote the closure of F_j, from Lemma 3.1 we deduce

$$\limsup_n \frac{1}{V_n} \log \nu(X_{n,F_j}) \leq \sup_{\rho\in\overline{F}_j\cap\mathcal{M}_1^T} \Big(h(T,\rho) - a_{j-1} \Big) + \delta. \tag{3.36}$$

The argument showing that $G^{\nu,\delta}$ defined by (3.5) is open may be applied to $-e_\nu$ to show that $\{\rho \in \mathcal{M}_1 : \langle\, e_\nu, \rho \,\rangle \leq a_j\}$ is closed, so that $\rho \in \overline{F}_j$ implies $\langle\, \rho, e_\nu \,\rangle \leq a_j \leq a_{j-1} + \delta$, so that

$$\limsup_n \frac{1}{V_n} \log \nu(X_{n,F_j}) \leq \sup_{\rho\in\overline{F}_j\cap\mathcal{M}_1^T} \Big(h(T,\rho) - \langle\, e_\nu, \rho \,\rangle \Big) + 2\delta. \tag{3.37}$$

Since each $\overline{F}_j \subset F$ and $\delta > 0$ is arbitrary, we deduce

$$\limsup_n \frac{1}{V_n} \log \nu(X_{n,F}) \leq \sup_{\rho\in F\cap\mathcal{M}_1^T} \Big(h(T,\rho) - \langle\, e_\nu, \rho \,\rangle \Big). \tag{3.38}$$

□

4 Finite alphabet shift spaces

One important special case, both in statistical mechanics and dynamical systems, is when X is a **shift space**, that is X is a closed subset of $\mathtt{A}^L$ with $\mathtt{A} = \{0, \ldots, r-1\}$, which is invariant under the action of L on $\mathtt{A}^L$ induced by addition. The space $\mathtt{A}^L$ is compact with the product topology. There are many possible metrics on $\mathtt{A}^L$. Here we consider those defined by $\{\Lambda_n\}$ and a strictly decreasing sequence $\{\Delta_n\}$ of positive real numbers (see [E] and definition below). Points of $\mathtt{A}^L$ are denoted by $\omega = (\omega_k : k \in L)$ in place of the x we use in the context of an abstract dynamical system (X, d, T). In shift spaces $h(T, \cdot)$ is upper semicontinuous as a function on $\mathcal{M}_1$ (see [W], chapter 8). We briefly adapt the former results to this particular case.

Definition 4.1 *For $\emptyset \neq \Lambda \subset L$, $J_\Lambda : \mathtt{A}^L \to \mathtt{A}^\Lambda$ denotes the function which maps $\omega \in \mathtt{A}^L$ to its restriction to Λ : for $k \in \Lambda$, $(J_\Lambda(\omega))_k := \omega_k$. We use J_n to denote J_{Λ_n}. X_n denotes the image of the shift space X under J_n. For $\nu \in \mathcal{M}_1$, we use ν_n to denote the measure on X_n, $\nu_n := \nu \circ X_n^{-1}$. The metric d on $\mathtt{A}^L$ is defined so that for $\omega \neq \omega'$,*

$$d(\omega, \omega') := \Delta_{\underline{n}} \text{ with } \underline{n} := \min\{n \in \mathbf{Z}_+ : J_n(\omega) \neq J_n(\omega')\}, \tag{4.1}$$

where $\{\Delta_n : n \in \mathbf{Z}_+\}$ is a specified strictly decreasing sequence of real numbers with $\lim_n \Delta_n = 0$. We define for $0 < \Delta \leq \Delta_0$,

$$n_\Delta := \max\{n \in \mathbf{Z}_+ : \Delta \leq \Delta_n\}, \tag{4.2}$$

so that $\Delta_{n_\Delta} \geq \Delta > \Delta_{n_\Delta + 1}$.

An immediate consequence of the definitions is the following.

Lemma 4.1 *With $0 < \Delta < \Delta_0$ and ω, $\omega' \in \mathtt{A}^L$ we have*

$$d(\omega, \omega') < \Delta \iff J_{n_\Delta}(\omega) = J_{n_\Delta}(\omega'). \tag{4.3}$$

$$d_{\Lambda_m}(\omega, \omega') < \Delta \iff J_{m+n_\Delta}(\omega) = J_{m+n_\Delta}(\omega'). \tag{4.4}$$

In the case shift spaces it is convenient to reformulate the approximate product property. We do it for the case $L = \mathbf{Z}_+$, the case $L = \mathbf{Z}^d$ can be treated in a similar way.

Definition 4.2 *Let $L = \mathbf{Z}_+$. The dynamical system (X, d, T) has the **approximate product property** if the following holds. Given any $\delta_1 > 0$ and $\delta_2 > 0$, there exists $N(\delta_1, \delta_2)$, such that for any $n \geq N(\delta_1, \delta_2)$ and for any collection $\{\omega^i\}_{i \in L} \subset X$, one can find a collection $\{j_i\}_{i \in L} \subset L$ and $\omega \in X$ so that $j_0 = 0$, $n \leq j_{i+1} - j_i \leq n(1 + \delta_2)$ and*

$$d_n^H(T^{j_i}\omega, \omega^i) \leq \delta_1 |\Lambda_n| \quad \forall\, i \in L\,. \tag{4.5}$$

It is not difficult to see that this definition is equivalent to the former one.

Definition 4.3 *With $X_{n,F}$ is specified by (2.6) we define*

$$\widehat{X}_{n,F} := \left\{ \mathrm{w} \in X_n : J_n^{-1}\{\mathrm{w}\} \cap X_{n,F} \neq \emptyset \right\}. \tag{4.6}$$

Proposition 4.1 *Let X be a shift space and $\nu \in \mathcal{M}_1^T$. The following two conditions are equivalent.*
I. For each neighbourhood F of ν in $\mathcal{M}_1$ and each $h^* < h(T,\nu)$, there exists $n^* \in \mathbf{N}$ so that $n \geq n^*$ implies

$$\log |\widehat{X}_{n,F}| \geq V_n h^*. \tag{4.7}$$

II. There exists $\varepsilon^\# > 0$ so that for each neighbourhood F of ν in $\mathcal{M}_1$ and each $h' < h(T,\nu)$, there exist $\delta^\# > 0$ and $n^\# \in \mathbf{N}$ so that whenever $n \geq n^\#$ there exists $\Gamma_n \subset X_{n,F}$ which is $(\delta^\#, n, \varepsilon^\#)$-separated and satisfies $\log |\Gamma_n| \geq V_n h'$.

Proof: Given $\nu \in \mathcal{M}_1^T$ satisfying I, $h' < h(T,\nu)$ and a neighbourhood F of ν, take h^* satisfying $h' < h^* < h(T,\nu)$. Set $\varepsilon^\# := \Delta_1$ so that $d(T^j\omega, T^j\omega') \leq \varepsilon^\# \Leftrightarrow \omega_j = \omega'_j$. Choose $n^\#$ so that (4.7) obtains whenever $n \geq n^\#$. Choose $\overline{\Gamma}_n$ so that for each $\mathrm{w} \in \widehat{X}_{n,F}$, $\overline{\Gamma}_n$ contains exactly one point $\omega \in X_{n,F}$ with $J_n(\omega) = \mathrm{w}$. Choose $\delta^\# > 0$ so

$$\eta(\delta^\#) \log 2 + \delta^\# \log(|\mathtt{A}| - 1) < h^* - h'. \tag{4.8}$$

It follows from (2.19) that for each $\mathrm{w} \in \mathtt{A}^{\Lambda_n}$,

$$K_n := |\{\mathrm{v} \in \mathtt{A}^{\Lambda_n} : d_n^H(\mathrm{v}, \mathrm{w}) \leq \delta^\# V_n\}| \tag{4.9}$$

satisfies $\log K_n < V_n(h^* - h')$. Choose Γ_n to be a maximal subset of $\overline{\Gamma}_n$ satisfying

$$\omega \neq \omega' \in \Gamma_n \Longrightarrow d_n^H(J_n(\omega), J_n(\omega')) > \delta^\# V_n. \tag{4.10}$$

The maximality of Γ_n implies

$$|\Gamma_n| \geq |\overline{\Gamma}_n| / K_n \geq \exp(V_n h'). \tag{4.11}$$

Because of (4.10), if $\Lambda \subset \Lambda_n$, $|\Lambda_n \setminus \Lambda| \leq \delta^\# V_n$, $d_\Lambda(\omega, \omega') = \Delta_0 > \varepsilon^\#$. The converse is trivial. $\square$

Next we consider lower- and upper-energy functions in the context of shift spaces. The following is a direct consequence of Lemma 4.1.

Proposition 4.2 *Let the shift space $X \subset \mathtt{A}^L$ be considered as a dynamical system (X, d, T). Let $\nu \in \mathcal{M}_1$. Then the bounded, lower semicontinuous function $e_\nu \colon X \to \mathbf{R}^+$ is an upper-energy function for ν if, and only if,*

$$\limsup_n \sup_{\mathrm{w} \in X_n} \left(\frac{1}{V_n} \log \nu_n(\mathrm{w}) + \sup_{\omega \in X \,:\, J_n(\omega) = \mathrm{w}} \langle\, e_\nu, \mathcal{E}_n(\omega)\, \rangle \right) \leq 0. \tag{4.12}$$

Also the upper semicontinuous $e_\nu \colon X \to \mathbf{R}^+$ is a lower-energy function for ν if, and only if,

$$\liminf_n \inf_{\mathrm{w} \in X_n} \left(\frac{1}{V_n} \log \nu_n(\mathrm{w}) + \inf_{\omega \in X \,:\, J_n(\omega) = \mathrm{w}} \langle\, e_\nu, \mathcal{E}_n(\omega)\, \rangle \right) \geq 0. \tag{4.13}$$

For some β-shifts we use a weak lower-energy functions. The associated $\nu^\beta \in \mathcal{M}_1^T$ can have a very small $\nu_n(\mathrm{v})$ for some word $\mathrm{v} \in X_n$, but there is a word $\mathrm{w} \in X_n$ with $d_n^H(\mathrm{w}, \mathrm{v}) = 1$ and $\nu_n(\mathrm{w}) \gg \nu_n(\mathrm{v})$. This suggests the following.

Proposition 4.3 *Let $\nu \in \mathcal{M}_1^T$ and let $e_\nu : X \to \mathbf{R}_+$ be upper semicontinuous. If for each $\delta > 0$ there exists n_δ so that $n \geq n_\delta$ implies that for each $\mathrm{v} \in X_n$ there exists $\mathrm{w} \in X_n$ satisfying $d_n^H(\mathrm{v}, \mathrm{w}) \leq \delta|\Lambda_n|$ and*

$$\frac{1}{V_n} \log \nu_n(\mathrm{w}) + \inf_{\omega:\, J_n(\omega)=\mathrm{w}} \langle\, e_\nu, \mathcal{E}_n(\omega)\,\rangle \geq -\delta, \tag{4.14}$$

then e_ν is a weak lower-energy function for ν.

Proof: Let $\rho \in \mathcal{M}_1^T$ be ergodic and let F be an f-neighbourhood of ρ. By Propositions 2.1 and 4.1 for any $h' < h^* < h(T, \rho)$ there exists n^* such that $n \geq n^*$ implies

$$\frac{1}{V_n} \log |\widehat{X}_{n,F(1/2)}| > h^*. \tag{4.15}$$

Given $\mathrm{v} \in \widehat{X}_{n,F(1/2)}$, (2.19) is a bound for the number of $\mathrm{w} \in X_n$ satisfying $d_n^H(\mathrm{v}, \mathrm{w}) \leq \delta V_n$, and at least one such w satisfies (4.14). Take $\delta > 0$ so that (4.8) is satisfied for $\delta^\# = \delta$. Then by Corollary 2.1 there exists $n' \geq \max\{n^*, n_\delta\}$ so that $n \geq n'$ implies

$$\frac{1}{V_n} \log \Big|\Big\{\mathrm{w} \in \widehat{X}_{n,F} : \frac{1}{V_n} \log \nu_n(\mathrm{w}) + \inf_{\omega:\, J_n(\omega)=\mathrm{w}} \langle\, e_\nu, \mathcal{E}_n(\omega)\,\rangle > -\delta\Big\}\Big| > h'. \tag{4.16}$$

This shows that e_ν is a weak lower-energy function for ν. □

We end this section by a remark about topological entropy and Hausdorff dimension, which we shall use when discussing Billingsley dimension for β-shifts. Basic reference is section 11 of [Pe]. We recall the definition of the topological entropy [Bo2] for a dynamical system (X, d, T). We follow [Pe] p.74-75. We consider the case $L = \mathbf{Z}_+$ and cover means always finite or countable cover. Let $s \in \mathbf{R}^+$, $N \in \mathbf{N}$, $\Delta > 0$ and $A \subset X$. Set

$$\mathcal{T}(A; s, N, \Delta) := \inf\left\{\sum_j \exp(-sn_j) : n_j \geq N,\ A \subset \bigcup_j B_{n_j}(x; \Delta)\right\}, \tag{4.17}$$

$$\mathcal{T}(A; s, \Delta) = \lim_{N\to\infty} \mathcal{T}(A; s, N, \Delta)\,, \tag{4.18}$$

and

$$h_{\mathrm{top}}(A, T; \Delta) := \inf\{s : \mathcal{T}(A; s, \Delta) = 0\} = \sup\{s : \mathcal{T}(A; s, \Delta) = \infty\}\,. \tag{4.19}$$

The *topological entropy* of A is defined by

$$h_{\mathrm{top}}(A, T) := \lim_{\Delta\to 0} h_{\mathrm{top}}(A, T; \Delta)\,. \tag{4.20}$$

We now restrict the discussion to shift spaces. Let $\{\Delta_n : n \in \mathbf{Z}_+\}$ be a strictly decreasing sequence to 0 and d be the associated metric. We denote by $\mathcal{S}$ the family of cylinder sets, which are by definition the sets of the form

$$C_m(\omega) := \{\omega' \in X : J_m(\omega) = J_m(\omega')\}\,. \tag{4.21}$$

Hausdorff dimensions are computed with cylinders. Let $s \in \mathbf{R}^+$, $\delta > 0$ and $A \subset X$. Set

$$\mathcal{C}(A; s, \delta) := \inf\left\{\sum_j \mathrm{diam}(U_j)^s : A \subset \bigcup_j U_j\ ,\ \mathrm{diam}(U_j) \leq \delta\right\} \tag{4.22}$$

where each $U_j \in \mathcal{S}$ and

$$\mathcal{C}(A;s) := \lim_{\delta\downarrow 0} \mathcal{C}(A;s,\delta)\,. \tag{4.23}$$

The *Hausdorff dimension* $\dim_H(A)$ of A is

$$\dim_H(A) := \inf\{s : \mathcal{C}(A;s) = 0\} = \sup\{s : \mathcal{C}(A;s) = \infty\}\,. \tag{4.24}$$

Let $0 < \Delta \leq \Delta_0$. We have $\Delta_{n_\Delta} \geq \Delta > \Delta_{n_\Delta+1}$ and

$$B(\omega;\Delta) = \{\omega' \in X : d(\omega,\omega') < \Delta\} \tag{4.25}$$
$$= \{\omega' \in X : d(\omega,\omega') < \Delta_{n_\Delta}\} \tag{4.26}$$
$$= \{\omega' \in X : J_{n_\Delta}(\omega) = J_{n_\Delta}(\omega')\} = C_{n_\Delta}(\omega)\,. \tag{4.27}$$

It follows that $B_k(\omega;\Delta) = C_{k+n_\Delta}(\omega)$ for all $k \in \mathbf{Z}_+$. For the full shift $\mathrm{diam}(C_{n_\Delta}(\omega)) = \Delta_{n_\Delta+1}$, but in general we only have $\mathrm{diam}(C_{n_\Delta}(\omega)) \leq \Delta_{n_\Delta+1}$. Comparing the two definitions, we obtain the next lemma.

Lemma 4.2 *Let $X \subset \mathtt{A}^{\mathbf{Z}_+}$ be a shift space with finite alphabet $\mathtt{A}$, and d be a metric as in Definition 4.1. Assume that there exists $\gamma > 0$ so that*

$$\limsup_{m} \sup_{\omega\in X} \Big|\frac{\log \mathrm{diam}\, C_m(\omega)}{V_m} + \gamma\Big| = 0\,. \tag{4.28}$$

Then, for any $A \subset X$, $h_{\mathrm{top}}(A,T) = \gamma \cdot \dim_H(A)$.

A simple example is when X is the full shift and there exists $\gamma > 0$ with

$$\lim_{n\to\infty} \frac{\log \Delta_n}{V_n} = -\gamma\,. \tag{4.29}$$

Indeed, in this case $\mathrm{diam}(C_{n_\Delta}(\omega)) = \Delta_{n_\Delta+1}$. This is an example of a scale metric. In the case of the full shift, Tempelman [Te2] computed Hausdorff dimension with respect to scale metrics for sets of the type of those considered in section 5. Further examples of shift spaces which verify the hypothesis of Lemma 4.2 are given in section 5.

5 β-shifts and topological entropy

In this section we consider β-shifts. Basic references are [R], [P], [B-M] and [B].
Let $\beta > 1$ be fixed. For $t \in \mathbf{R}$ we define

$$\lfloor t \rfloor := \max\{i \in \mathbf{Z} : i \leq t\},\ \lceil t \rceil := \min\{i \in \mathbf{Z} : i \geq t\}. \tag{5.1}$$

We define $\mathtt{b} := \lceil \beta \rceil$. Consider the β-expansion of 1,

$$1 = \sum_{i=1}^{\infty} c_i \beta^{-i}\,,$$

which is given by the algorithm

$$r_0 := 1,\ c_{i+1} := \lceil \beta\, r_i \rceil - 1,\ r_{i+1} := \beta\, r_i - c_{i+1},\ i \in \mathbf{Z}_+, \tag{5.2}$$

which insures that $r_i > 0$ for all $i \in \mathbf{Z}_+$. Therefore the sequence $c := \{c_i\}_{i\geq 1}$ cannot end with zeros only. For sequences $\{a_i\}_{i\geq 1}$ and $\{b_i\}_{i\geq 1}$ the lexicographical order is defined by $\{a_i\} < \{b_i\}$ if and only if for the least index i with $a_i \neq b_i$, $a_i < b_i$. The β-shift X^β is the subshift of the full shift on the alphabet with $\mathtt{b}$ characters, $\mathtt{A} := \{0, \ldots, \mathtt{b}-1\} \subset \mathbf{Z}_+$, which is given by

$$X^\beta := \Big\{\omega = \{\omega_i\}_{i\geq 1} : \omega_i \in \mathtt{A},\ T^k\{\omega_i\} \leq \{c_i\}\,\forall\, k \in \mathbf{Z}_+\Big\}\,, \tag{5.3}$$

since $T^k\{c_i\} \leq \{c_i\}$ for all $k \in \mathbf{Z}_+$, X^β is a shift-invariant closed subset of the full shift. For convenience X^β is defined as a subset of $\mathtt{A}^{\mathbf{N}}$ instead of $\mathtt{A}^{\mathbf{Z}_+}$, as in [PfS1]. Hence, the elements of X_n are the **words of length** n of the shift space. For $\nu \in \mathcal{M}_1$, ν_n denotes, as above, the measure on X_n, $\nu_n := \nu \circ X_n^{-1}$. When β is an integer, X^β coincides with the full shift.
The subshift X^β occurs as the natural coding of the β-transformation, which we define as follows (left-continuous version). Let $M := (0,1]$. We decompose M into $\mathtt{b}$ sub-intervals,

$$J_0 := \Big(0, \frac{1}{\beta}\Big]\,,\ J_1 := \Big(\frac{1}{\beta}, \frac{2}{\beta}\Big]\,, \cdots\,,\ J_{\mathtt{b}-1} := \Big(\frac{\mathtt{b}-1}{\beta}, 1\Big]\,, \tag{5.4}$$

and define $\varphi_\beta : M \to M$ by

$$\varphi_\beta(x) := \beta x - p \quad \text{if } x \in J_p\,. \tag{5.5}$$

To each orbit of $x \in M$, $(x, \varphi_\beta(x), \varphi_\beta^2(x), \ldots)$, we associate a coding sequence $\omega(x) \in \mathtt{A}^{\mathbf{N}}$, defined by

$$\omega_n(x) := p \iff \varphi_\beta^{n-1}(x) \in J_p\,. \tag{5.6}$$

In particular the coding sequence of $1 \in M$ is the sequence $c \in X^\beta$. X^β is the closure of the set of all coding sequences, and $\omega \in X^\beta$ is a coding sequence of an orbit of the β-transformation if and only if $\limsup_i \omega_i > 0$. If this is the case, then ω is the coding sequence of $x = \sum_{i\geq 1} \omega_i \beta^{-i}$.
There is a simple presentation of X^β by a labelled graph $\mathcal{G}^\beta$ [BH]. We set for $i, j \in \mathbf{N}$, $i < j$, $[i,j] := \{k \in \mathbf{N} : i \leq k \leq j\}$ and $\omega_i^j := J_{[i,j]}(\omega)$. Suppose first that the sequence c is not ultimately periodic. The set of vertices of the labelled graph $\mathcal{G}^\beta$ is countable, $\{q_i : i \in Z_+\}$, and is into 1-1-correspondence with the set of prefixes of $\{c_i\}$, including the empty word ϵ of length 0. The prefix ϵ corresponds to q_0 and the prefix c_1^m to q_m, $m \geq 1$. There are $\mathtt{b}-1$ (directed) edges from q_0 to q_0, which are labelled by $0, 1, \ldots, \mathtt{b}-2$, and one edge from q_0 to q_1 which is labelled by $c_1 = \mathtt{b}-1$. There is one edge from q_m to q_{m+1}, which is labelled by c_{m+1}, and c_{m+1} edges (possibly no edge if $c_{m+1} = 0$) from q_m to q_0, which are labelled by $0, 1, \ldots c_{m+1} - 1$. X^β corresponds to the set of all labelled infinite paths starting at q_0. When c is ultimately periodic, c can be written as the concatenation of words, $c \equiv c_1^m u_1^k u_1^k \cdots$. The graph $\mathcal{G}^\beta$ has only finitely many vertices, $q_0, q_1, \ldots, q_{m+k-1}$. The definition of $\mathcal{G}^\beta$ is as above, with the only modification that there is an edge from q_{m+k-1} to q_m which is labelled by $c_{m+k} = u_k$, rather than an infinite continuation.

Proposition 5.1 *The β-shifts have the approximate product property.*

Proof: We verify Definition 4.2 with $\delta_1 > 0$ and $\delta_2 = 0$. Suppose that n is large enough so that $\delta_1 n \geq 1$. Let $\{\mathrm{w}^i\}_{i\in L}$ be an arbitrary collection of words of length n. Each word is presented by a path of length n on $\mathcal{C}^\beta$ starting at q_0. Suppose that the path presenting

w^i ends at q_0. Then the concatenation of the words w^i and w^{i+1} is a word of length $2n$ of X^β. On the other hand, if the path presenting a word w ends at q_m, let $\mathrm{w} \equiv u_1^n$. This occurs if and only if there exists a decomposition of u_1^n as $u_1^k c_1^{n-k}$ with $k < n$ $(m = n-k)$. Let c_j be the last character in the word c_1^{n-k} which is different from 0. We change this character c_j into the character $c_j - 1$. We get a new word w' which is presented by a path from q_0 to q_0, and which can be concatenated with any word. Let

$$\widehat{\mathrm{w}} := \begin{cases} \mathrm{w} & \text{if w is presented by a path from } q_0 \text{ to } q_0, \\ \mathrm{w}' & \text{if w is presented by a path from } q_0 \text{ to } q_m,\ m \neq 0. \end{cases} \tag{5.7}$$

By definition the concatenation of all words $\{\widehat{\mathrm{w}}^i\}$ gives $\omega = \widehat{\mathrm{w}}^1\widehat{\mathrm{w}}^2\widehat{\mathrm{w}}^3 \cdots \in X^\beta$ with the desired properties. □

The β-shift X^β has a unique ergodic measure of maximal entropy, which we denote by ν^β, [H1]. The corresponding measure of maximal entropy for the β-transformation of the interval $[0,1]$ is known to be absolutely continuous with respect to Lebesgue measure, with a density which is bounded above by $\frac{\beta}{\beta-1}$ and below by $\frac{\beta-1}{\beta}$, [R] and [P]. We use this fact for estimating $\nu_n^\beta(\mathrm{w})$ for $\mathrm{w} \in X_n$. We first define

$$z^\beta(c_1^\ell) := \begin{cases} 0 & \text{if } c_{\ell+1} > 0, \text{ or if } c_1^\ell = \epsilon \\ p & \text{if } c_{\ell+1} = \ldots = c_{\ell+p} = 0 \text{ and } c_{\ell+p+1} > 0. \end{cases} \tag{5.8}$$

The meaning of $z^\beta(c_1^\ell)$ is simple: $z^\beta(c_1^\ell)$ is a measure of the obstruction to go from the vertex q_ℓ to the vertex q_0 of the labelled graph $\mathcal{G}^\beta$. The length of the shortest path in $\mathcal{G}^\beta$ from vertex q_ℓ to vertex q_0 is $z^\beta(c_1^\ell) + 1$. We set

$$\overline{z}_n^\beta := \max_{1\leq \ell \leq n} z^\beta(c_1^\ell)\,. \tag{5.9}$$

Lemma 5.1 *Let $\mathrm{w} \in X_n$, and write $\mathrm{w} = \omega_1^k c_1^{n-k}$, where c_1^{n-k} is the largest possible prefix of c (possibly ϵ if $k = n$). Then*

$$\nu_n^\beta(\mathrm{w}) \leq \frac{\beta}{\beta-1}\beta^{-n}\,, \tag{5.10}$$

and

$$\nu_n^\beta(\mathrm{w}) \geq \frac{\beta-1}{\beta}\beta^{-(n+z^\beta(c_1^{n-k})+1)}\,. \tag{5.11}$$

Proof: We estimate $\nu_n^\beta(\mathrm{w})$ by estimating the Lebesgue measure of the set of points $x \in M$ whose coding sequences begin with w. The basic elementary observation is that, if $(a_1, a_2] \subset J_p$ and there exists $b_2 \in J_q$ so that $\varphi_\beta(b_2) = a_2$, then there exists $b_1 \in J_q$ so that $\varphi_\beta(b_1) = a_1$ and $\beta(b_2 - b_1) = a_2 - a_1$. Therefore, if the coding sequence of $x \in M$ begins with $\mathrm{w} = \omega_1^n$, then

$$\varphi_\beta^{-1}\Big(\cdots\varphi_\beta^{-1}\Big(\varphi_\beta^{-1}\Big((\omega_n\beta^{-1}, \varphi_\beta^{n-1}(x)]\Big) \cap J_{\omega_{n-1}}\Big) \cap J_{\omega_{n-2}}\cdots\Big) \tag{5.12}$$

is a sub-interval of J_{ω_1} of length $\beta^{-n+1}|\varphi_\beta^{n-1}(x) - \omega_n\beta^{-1}|$. Moreover, each x in that interval has a coding sequence starting with the prefix w. Consequently

$$\frac{\beta-1}{\beta}\beta^{-n+1}|\varphi_\beta^{n-1}(x) - \omega_n\beta^{-1}| \leq \nu_n^\beta(\mathrm{w}) \leq \frac{\beta}{\beta-1}\beta^{-n}\,. \tag{5.13}$$

Let $x = 1$ and $\mathrm{w} = c_1^n$. Then, by definition of $z^\beta(c_1^n)$,

$$\beta^{z^\beta(c_1^n)}(\varphi_\beta^{n-1}(1) - c_n\beta^{-1}) \leq \beta^{-1} \tag{5.14}$$

and

$$\beta^{z^\beta(c_1^n)+1}(\varphi_\beta^{n-1}(1) - c_n\beta^{-1}) > \beta^{-1}\,. \tag{5.15}$$

This proves (5.11) in this particular case. In the general case $\mathrm{w} = \omega_1^k c_1^{n-k}$. Then ω' such that

$$\omega_i' := \begin{cases} \omega_i & \text{if } i \leq k \\ c_{i-k} & \text{if } i > k, \end{cases} \tag{5.16}$$

is a coding sequence of some point $x' \in M$. Hence $\varphi_\beta^{n-1}(x') = \varphi_\beta^{n-k-1}(1)$ (the preimage of 1 in J_{ω_n} if $k = n$). From this the result follows. □

Theorem 5.1 *Let $\beta > 1$, and ν^β denote the ergodic measure of maximal entropy for the β-shift X^β. Then the constant function*

$$e_\beta := \log\beta \equiv h(T,\nu^\beta) \equiv h_{\text{top}}(X^\beta, T) \tag{5.17}$$

is an upper and a weak lower-energy function for ν^β. If $\lim \overline{z}_n^\beta/n = 0$, then this e_β is a lower-energy function.

Proof: The upper-energy property follows from (5.10). If $\lim \overline{z}_n^\beta/n = 0$, (5.11) implies the lower-energy property. To deduce the weak lower-energy property, we use Proposition 4.3, noting that $\widehat{\mathrm{w}}$ given by (5.7) differs from w by not more than one character and $\nu_n^\beta(\widehat{\mathrm{w}})$ satisfies (5.11) with $n = k$. □

Corollary 5.1 *For any $\beta > 1$ the empirical measures verify a large deviations principle (Theorems 3.1 and 3.2) for the measure of maximal entropy ν^β and the rate function $h(T,\rho) - \langle \log\beta, \rho \rangle \equiv h(T,\rho) - \log\beta$ is upper semicontinuous.*

In [PfS1] we computed the Billingsley dimension ([Bi]) of various sets of generic points in the case of subshifts, which verify three hypothesis H1, H2, H3. Billingsley dimension is defined as Hausdorff dimension (see section 4), by replacing $\mathrm{diam}(C_m(\omega))$ by $\nu(C_m(\omega))$, where ν is a shift-invariant probability measure. The results of [PfS1] apply in particular to any irreducible (possible periodic) sofic shifts, but in fact to a much larger class of subshifts, which do not verify the specification property. Condition H2 implies the approximate product property. Condition H1 says that the ergodic measures are entropy-dense. Hence, Theorem 2.1 of this paper asserts that H2 implies H1. Condition H3 says that the reference measure ν for computing the Billingsley dimension has a continuous energy function (upper and lower) e_ν. We can apply directly the results of [PfS1] to most β-shifts if we choose as reference probability measure the measure ν^β. Indeed, condition H2 of [PfS1] implies that $\lim \overline{z}_n^\beta/n = 0$; it follows that

$$\lim_n \sup \sup_{\omega\in X^\beta} \Big|\frac{\log\nu^\beta(C_n(\omega))}{n} + \log\beta\Big| = 0\,, \tag{5.18}$$

hence ν^β has the energy function $e_\beta = \log\beta$. It is not difficult to prove, using Theorem C of [Sc], that if β does not verify hypothesis H2 of [PfS1], then it cannot be in class $\tilde{C}_5$ of

[Sc]. Therefore, condition H2 of [PfS1] is verified for a set of β of full Lebesgue measure, but it does not hold for all β. From Lemma 4.2, if $\lim_{n\to\infty}\overline{z}_n^\beta/n = 0$, then

$$h_{\mathrm{top}}(A,T) = \log\beta\cdot\dim_{\nu^\beta}(A)\,. \tag{5.19}$$

In [PfS1] we computed the Billingsley dimension of the following sets. Let F be a nonempty closed subset of $\mathcal{M}_1^T$. We define

$${}^F G := \{\omega\in X^\beta : \mathcal{E}_n(\omega) \text{ has a limit point in } F\}\,. \tag{5.20}$$

and, when F is also connected,

$$G_F := \{\omega\in X^\beta : \text{the limit point set of } \mathcal{E}_n(\omega) \text{ equals } F\}\,. \tag{5.21}$$

An important particular case is for $F := \{\nu\}$. $G_{\{\nu\}}$ is the set of generic points of ν.

Theorem 5.2 *Let $\beta > 1$ and ν^β be the ergodic measure of maximal entropy for the β-shift X^β.*

1) Let $F\subset\mathcal{M}_1^T$ be a closed set. Then $h_{\mathrm{top}}({}^F G,T) = \sup_{\rho\in F} h(T,\rho)$.

2) Let $F\subset\mathcal{M}_1^T$ be a closed connected set. Then $h_{\mathrm{top}}(G_F,T) = \inf_{\rho\in F} h(T,\rho)$.

Proof: When $\lim\overline{z}_n^\beta/n = 0$ this is simply a transcription of the results of [PfS1] using (5.19). If β does not imply $\lim\overline{z}_n^\beta/n = 0$, then one first notices that the proof of Proposition 2.1 in [PfS1] gives

$$h_{\mathrm{top}}({}^F G,T) \leq \sup_{\rho\in F} h(T,\rho)\,. \tag{5.22}$$

Moreover, we can adapt the proof of Proposition 2.2 in [PfS1] to get

$$h_{\mathrm{top}}(G_F,T) \geq \inf_{\rho\in F} h(T,\rho)\,. \tag{5.23}$$

Indeed, we can use only words which are presented by paths from q_0 to q_0 in $\mathcal{G}^\beta$. For those words, we get from Lemma 5.1

$$\nu_n^\beta(\mathrm{w}) \geq \frac{\beta-1}{\beta}\beta^{-(n+1)}\,. \tag{5.24}$$

The construction of the set $B\subset G_F$ in [PfS1] becomes even simpler, since two words w and w$'$ which are presented by a path from q_0 to q_0 can be concatenated to give a word ww$'$ which is presented by a path from q_0 to q_0. Because of (5.24) we still have for the particular set B, $h_{\mathrm{top}}(B,T) = \log\beta\dim_{\nu^\beta}(B)$; hence (5.23) is still valid. Then

$$\sup_{\rho\in F} h(T,\rho) \leq \sup_{\rho\in F} h_{\mathrm{top}}(G_{\{\rho\}},T) \leq h_{\mathrm{top}}({}^F G,T) \leq \sup_{\rho\in F} h(T,\rho)\,, \tag{5.25}$$

and

$$\inf_{\rho\in F} h(T,\rho) \geq \inf_{\rho\in F} h_{\mathrm{top}}({}^{\{\rho\}}G,T) \geq h_{\mathrm{top}}(G_F,T) \geq \inf_{\rho\in F} h(T,\rho)\,. \tag{5.26}$$

□

The following corollary is well-known for ergodic measures for compact dynamical systems (X,d,T) with T continuous [Bo2]. If ν is ergodic, then $\nu(G_{\{\nu\}}) = 1$, and this fact is essential in [Bo2]. When ν is not ergodic, then $\nu(G_{\{\nu\}}) = 0$, so the situation is completely different.

Corollary 5.2 *Let $\beta > 1$, and ν be any shift-invariant measure on the β-shift X^β. Then*

$$h_{\mathrm{top}}(G_{\{\nu\}}, T) = h(T, \nu)\,. \tag{5.27}$$

The next result shows how Theorem 5.2 can be used to compute the topological entropy of sets involving ergodic averages. In the terminology in usage in large deviations theory it is called a contraction principle.

Definition 5.1 *Let $\mathcal{A}$ be an arbitrary nonempty parameter set. Let*

$$\mathbf{f} := \{(f_\alpha, c_\alpha, d_\alpha) \colon \alpha \in \mathcal{A}\}, \tag{5.28}$$

where $f_\alpha \colon \mathbf{R} \to \mathbf{R}$ is continuous and $c_\alpha \leq d_\alpha \in \mathbf{R}$ for all $\alpha \in \mathcal{A}$. Define

$$\begin{aligned} X_{\mathbf{f}} := \{x \in X \colon \qquad & \liminf \langle\, f_\alpha, \mathcal{E}_n(x)\,\rangle \geq c_\alpha, \\ & \limsup \langle\, f_\alpha, \mathcal{E}_n(x)\,\rangle \leq d_\alpha,\ \forall\, \alpha \in \mathcal{A}\} \end{aligned} \tag{5.29}$$

and

$$F_{\mathbf{f}} := \{\rho \in \mathcal{M}_1^T \colon \langle\, f_\alpha, \rho\,\rangle \in [c_\alpha, d_\alpha]\ \forall\, \alpha \in \mathcal{A}\}. \tag{5.30}$$

Lemma 5.2 *We have $G(F_{\mathbf{f}}) = X_{\mathbf{f}}$, where*

$$G(F_{\mathbf{f}}) := \{x \in X \colon \{\mathcal{E}_n(x)\} \text{ has all its limit points in } F_{\mathbf{f}}\}. \tag{5.31}$$

Proof: Note that $F_{\mathbf{f}}$ is a (possibly empty) closed convex subset of $\mathcal{M}_1^T$. Let $x \in X_{\mathbf{f}}$. For any given $\alpha \in \mathcal{A}$,

$$\liminf \langle\, f_\alpha, \mathcal{E}_n(x)\,\rangle \geq c_\alpha, \tag{5.32}$$

so any weak* limit ρ satisfies $\langle\, f_\alpha, \rho\,\rangle \geq c_\alpha$; the corresponding inequality for $\limsup$ also obtains, so $\rho \in F_{\mathbf{f}}$. Thus $X_{\mathbf{f}} \subset G(F_{\mathbf{f}})$. Next assume $x \in G(F_{\mathbf{f}})$. Given $\alpha \in \mathcal{A}$, by compactness for the sequence $\{\langle\, f_\alpha, \mathcal{E}_n(x)\,\rangle \colon n \in \mathbf{Z}_+\}$ there exist $\{n_j\}$ and $\rho \in \mathcal{M}_1^T$ so that

$$\lim_j \mathcal{E}_{n_j}(x) = \rho,\ \lim_j \langle\, f_\alpha, \mathcal{E}_{n_j}(x)\,\rangle = \liminf_n \langle\, f_\alpha, \mathcal{E}_n(x)\,\rangle = \langle\, f_\alpha, \rho\,\rangle. \tag{5.33}$$

Since $\rho \in F_{\mathbf{f}}$, $\liminf_n \langle\, f_\alpha, \mathcal{E}_n(x)\,\rangle \geq c_\alpha$. The corresponding argument for $\limsup$ shows that $x \in X_{\mathbf{f}}$, so $G(F_{\mathbf{f}}) \subset X_{\mathbf{f}}$. □

Theorem 5.3 *If the conclusions of Theorem 5.2 hold and $F_{\mathbf{f}} \neq \emptyset$, then the following variational principle holds,*

$$h_{\mathrm{top}}(X_{\mathbf{f}}, T) = \sup\{h(T, \rho) \colon \rho \in F_{\mathbf{f}}\}\,. \tag{5.34}$$

Proof: Since for $\rho \in F_{\mathbf{f}}$,

$$G_{\{\rho\}} \subset X_{\mathbf{f}} = G(F_{\mathbf{f}}) \subset {}^{F_{\mathbf{f}}}G, \tag{5.35}$$

Theorem 5.2 implies

$$\sup_{\rho \in F_{\mathbf{f}}} h(T, \rho) \leq h_{\mathrm{top}}(X_{\mathbf{f}}, T) \leq h_{\mathrm{top}}({}^{F_{\mathbf{f}}}G, T) \leq \sup_{\rho \in F_{\mathbf{f}}} h(T, \rho)\,. \tag{5.36}$$

□

Remark. Results of this kind have been obtained by Tempelman [Te2] and Takens and Verbitskiy [TV2]. They compute Hausdorff dimensions or topological entropy for sets of the form

$$\{x \in X : \lim_{n\to\infty} \langle\, f_\alpha, \mathcal{E}_n(x) \,\rangle = a_\alpha \ ,\ \alpha = 1, \ldots n\}\,. \tag{5.37}$$

Tempelman treats the case when the continuous f_α have values in some Banach space and the Hausdorff dimension is computed with scale metrics. Tempelman considers also more general ergodic averages (Følner sequences). The more restrictive hypothesis in [Te2] is that the results are obtained for the full shift only, so that they apply only to β-shifts with $\beta \in \mathbf{N}$. Takens and Verbitskiy compute the topological entropy of sets (5.37) for continuous real-valued functions f_α. They obtain their results in the more general setting of dynamical systems with (X, d) compact and $T : X \to X$ continuous. However they assume that the specification property holds, so that their results, in the case of β-shifts, apply only to a set of β of Lebesgue measure 0.

Acknowledgements
We wish to dedicate this paper to the memory of John T. Lewis, our teacher, colleague and friend.
Very careful reading of the original manuscript by the referees helped us to make significant corrections and improvements.
This work was supported by Science Foundation Ireland under the National Development Plan.

References

[B-M] A. Bertrand-Mathis, Développement en base θ, répartition modulo un de la suite $(x\theta^n)_{n\geq 0}$, langages codes et θ-shift, *Bull. Soc. math. France* **114**, 271-323 (1986).

[Bi] P. Billingsley, Hausdorff dimension in probability theory, *Ill. J. Mat.* **4**, 187–209 (1960) and *Ill. J. Mat.* **5**, 291–298 (1961).

[B] F. Blanchard, β-expansions and symbolic dynamics, *Theor. Comput. Sci.* **65**, 131-141 (1989).

[BH] F. Blanchard, G. Hansel, Systèmes codés, *Theor. Comput. Sci.* **44**, 17-49 (1986).

[Bl] A.M. Blokh, Decomposition of dynamical systems on an interval, *Russ. Math. Surv.* **38**, 133-134 (1983).

[Bo1] R. Bowen, Entropy for group endomorphisms and homogeneous spaces *Trans. Amer. Math. Soc.* **153**, 401-414 (1971).

[Bo2] R. Bowen, Topological entropy for noncompact sets, *Trans. Amer. Math. Soc.* **184**, 125-136 (1973).

[Bu] J. Buzzi, Specification on the interval *Trans. Amer. Math. Soc.* **349**, 2737-2754 (1997).

[E] G.A. Edgar, Measure, Topology and Fractal Geometry, Springer, Berlin (1990).

[EKW] A. Eizenberg, Y. Kifer and B. Weiss, Large Deviations for $\mathbf{Z}^d$-actions, *Commun. Math. Phys.* **164** 433-454 (1994).

[FO] H. Föllmer, S. Orey, Large Deviations for the Empirical Field of a Gibbs Measure, *Annals of Prob.* **16**, 961-977 (1988).

[H1] F. Hofbauer, β-shifts have unique maximal measure, *Mh. Math.* **85**, 189-198 (1978).

[H2] F. Hofbauer, Generic properties of invariant measures for continuous piecewise monotonic transformations *Mh. Math.* **106**, 301-312 (1988).

[KH] A. Katok and B. Hasselblatt, Introduction to the Modern Theory of Dynamical Systems, Cambridge University Press, Cambridge (1995).

[K1] Y. Kifer, Large Deviations in Dynamical Systems and Stochastic Processes, *Trans. Amer. Math. Soc.* **321**, 505-524 (1990).

[K2] Y. Kifer, Large Deviations, Averaging and Periodic Orbits of Dynamical Systems, *Commun. Math. Phys.* **162**, 33-46 (1994).

[M] B. Markus, A note on periodic points for ergodic toral automorphismus *Mh. Math.* **89**, 121-129 (1980).

[Mi] M. Misiurewicz, A short proof of the variational principle for a $\mathbf{Z}_+^{\mathbf{N}}$-action on a compact metric space, *Astérisque*, **40**, 147-157 (1976).

[P] W. Parry, On the β-expansions of real numbers, *Acta Math. Hung.* **11**, 401-416 (1960).

[Pe] Y.E. Pesin, Dimension Theory in Dynamical Systems, Contemporary Views and Applications, Chicago Lectures in Mathematics, Univerity of chicago Press, Chicago (1997).

[Pf] C.-E. Pfister, Thermodynamical aspects of classical spin systems. Lectures delivered at the 4th brazilian school of probability, Mambucaba, RJ. *In and Out of Equilibrium, Probability with a Physics Flavor*, ed. V. Sidoravicius, Birkhäuser, 393-472 (2002).

[PfS1] C.-E. Pfister, W.G. Sullivan, Billingsley dimension on shift spaces, *Nonlinearity* **16**, 661-682 (2003).

[PfS2] C.-E. Pfister, W.G. Sullivan, Rényi Entropy and Guesswork Moments, to appear in *IEEE Transactions On Information Theory* (2004).

[R] A. Rényi, Representations for real numbers and their ergodic properties, *Acta Math. Hung.* **8**, 477-493 (1957).

[S] P.C. Shields, The Ergodic Theory of Discrete Sample Paths, Graduate Studies in Mathematics Volume **13**, American Mathematical Society(1996).

[Sc] J. Schmeling, Symbolic dynamics for β-shifts and self-normal numbers, *Ergod. Th. & Dynam. Sys.* **17**, 675-694 (1997).

[Schw] L. Schwartz, Radon measures on arbitrary topological measures and cylindrical measures, Oxford University Press, London (1973).

[TV1] F. Takens, E. Verbitskiy, Multifractal analysis of dimensions and entropies *Regular and Chaotic Dynamics*, **5**, 361-382 (2000).

[TV2] F. Takens, E. Verbitskiy, On the variational principle for the topological entropy of certain non-compact sets *Ergod. Th. & Dynam. Sys* **23**, 317-348 (2003).

[Te1] A.A. Tempelman, Dimension of Random Fractals in Metric Spaces, *Teor. Veroyatnost. i Primenen.* **44**, 589-616 (1999); translation in *Theory Probab. Appl.* **44**, 537-557 (2000).

[Te2] A.A. Tempelman, Multifractal Analysis of Ergodic Averages: A Generalization of Eggleston's Theorem, *J. Dynam. Control Systems* **7**, 535-551 (2001).

[W] P. Walters, An Introduction to Ergodic Theory, Graduate Texts in Mathematics **79**, Springer, Berlin (2000).

[Y] L.-S. Young, Some Large Deviation Results for Dynamical Systems, *Trans. Amer. Math. Soc.* **318**, 525-543 (1990).