INSTITUTE OF PHYSICS PUBLISHING NONLINEARITY

Nonlinearity **18** (2005) 445–463 doi:10.1088/0951-7715/18/1/023

Finite type approximations of Gibbs measures on sofic subshifts

J-R Chazottes[1], **L Ramirez**[2] **and E Ugalde**[2]

[1] CPHT, CNRS-Ecole Polytechnique, 91128 Palaiseau Cedex, France
[2] IICO, Av. Karakorum 1470, Lomas 4A., 78210 San Luis Potosí, SLP, Mexico

Received 29 March 2004, in final form 7 September 2004
Published 29 October 2004
Online at stacks.iop.org/Non/18/445

Recommended by C Liverani

Abstract
Consider a Hölder continuous potential ϕ defined on the full shift $A^{\mathbb{N}}$, where A is a finite alphabet. Let $X \subset A^{\mathbb{N}}$ be a specified sofic subshift. It is well known that there is a unique Gibbs measure μ_ϕ on X associated with ϕ. In addition, there is a natural nested sequence of subshifts of finite type (X_m) converging to the sofic subshift X. To this sequence we can associate a sequence of Gibbs measures (μ_ϕ^m). In this paper, we prove that these measures converge weakly at exponential speed to μ_ϕ (in the classical distance metrizing weak topology). We also establish a mixing property that implies that μ_ϕ is Bernoulli. Finally, we prove that the measure-theoretic entropy of μ_ϕ^m converges to the one of μ_ϕ exponentially fast. We indicate how to extend our results to more general subshifts and potentials.

Mathematics Subject Classification: 37D35, 37B10

1. Introduction

The existence and uniqueness of equilibrium states/Gibbs measures associated with sufficiently regular potentials is established in the general context of expansive homeomorphisms acting on a compact metric space satisfying specification [1, 8]. This class of systems contains subshifts of finite type (coding axiom A diffeomorphisms) but more generally all specified subshifts such as topologically mixing sofic subshifts (on finite alphabets).

The usual way of proving existence and uniqueness is to construct a sequence of elementary Gibbs measures (which are atomic) and to argue that such a sequence must have an accumulation point in the weak topology. Then one proves that this accumulation point is unique. In the particular case of subshifts of finite types and Hölder continuous potentials, there is a complete theory of Gibbs measures [2].

The point of view adopted here for studying Gibbs measures on a specified subshift X is to approximate it using a nested sequence of subshifts of finite type (X_m) in the sense of a

Hausdorff metric (there is a canonical way to do this). This gives a natural sequence of Gibbs measures (finite type approximations) which converges weakly to a Gibbs measure whose properties we wish to analyse.

For the sake of definiteness, we assume that the given potential, ϕ, on $A^{\mathbb{N}}$ (A is a finite alphabet) is Hölder continuous and $X \subset A^{\mathbb{N}}$ is a specified sofic subshift. As we shall comment at the end of the paper, we are not restricted to that situation. The two crucial properties on which our method relies are specification and the presence of magic words (see definitions below). Sofic subshifts provide a natural class of subshifts with such properties.

Our main result can be phrased as follows: the sequence of finite type approximations (μ_ϕ^m) defined on (X_m) converges weakly, as $X_m \to X$, to a measure μ_ϕ at an exponential speed. Then this measure must be a Gibbs measure associated with ϕ. Moreover, we prove a strong mixing property (implying that μ_ϕ is Bernoulli). By a classical argument (Bowen), this implies uniqueness. We also prove that the measure-theoretic entropy $h(\mu_\phi^m)$ convergences to $h(\mu_\phi)$ exponentially fast (as well as the relative entropy $h(\mu_\phi \mid \mu_\phi^m)$) to 0. We use and prove the fact that the topological pressure $P(\phi, X_m)$ converges to $P(\phi, X)$ exponentially fast.

We use two tools. The first one is algebraic (contraction properties of iteration of primitive matrices with respect to the projective metric); the second one is symbolic dynamics. All our constants have explicit expressions in terms of the 'data' of the problem, that is, the cardinality of the alphabet, the supremum norm of the potential, its Hölder constants and the specification length of the subshift X.

We would like to mention some related papers. First of all, the paper [7] where the author is concerned with measures of maximal entropy. Informally speaking, the author states some sufficient conditions on the way a subshift is approximated in 'entropy' by subshifts of finite type so that the corresponding sequence of measure of maximal entropy has a unique limit. The main tool is graph theory. Second, let us mention the paper [3] where the authors estimate from above the $\overline{d}$-distance between a chain with complete connections and its Markov approximations, but $X_m = X = A^{\mathbb{N}}$. They use a coupling approach. Third, there is a construction of certain Gibbs measures on an arbitrary subshift of $A^{\mathbb{N}}$ by means of Markov approximations similar to ours appearing in [13]. But only the weak convergence of these approximations is proved since the author is concerned with existence and uniqueness.

The paper is organized as follows. In section 2 we record basic definitions and notations. Section 3 contains our main results. Section 4 is devoted to some preparatory lemmas that we use for the proof of our main results in section 5. In section 6 we indicate some straightforward generalizations of our results as well as examples. We can indeed handle potentials with polynomial variations (decaying fast enough). Consequently, the exponential speeds mentioned above become polynomial. We can also deal with more general specified subshifts (for instance, non-sofic but specified β-shifts).

2. Preliminary notions

2.1. Symbolic dynamics

Let A be a finite alphabet. For all integers $m, n, m \leqslant n$, in $\mathbb{N}_0$ ($\mathbb{N}_0 := \mathbb{N} \cup \{0\}$), we denote by $\mathbf{a}(m:n)$ the word $\mathbf{a}(m)\mathbf{a}(m+1)\cdots\mathbf{a}(n-1)\mathbf{a}(n)$ of length $n-m+1$. Let $0 < \theta < 1$. The distance

$$d_A(\mathbf{a}, \mathbf{b}) := \theta^{n(\mathbf{a},\mathbf{b})} \qquad \text{where } n(\mathbf{a}, \mathbf{b}) := \min\{n \geqslant 0 : \mathbf{a}(0:n) \neq \mathbf{b}(0:n)\}$$

makes the Cartesian product $A^{\mathbb{N}}$ a compact metric space.

As usual, the *shift transformation* $T : A^{\mathbb{N}} \to A^{\mathbb{N}}$ is the map such that $T\mathbf{a}(i) = \mathbf{a}(i+1)$.

A *subshift* is a T-invariant compact set $X \subset A^{\mathbb{N}}$. The subshift X is said to be of *finite type*, if it is defined by a finite collection of admissible words, which can be taken to be of the same length for the sake of simplicity (and without loss of generality). So, the subshift of finite type defined by the collection $\mathcal{L} \subset A^{n+1}$ of *admissible words* is the compact set

$$A_{\mathcal{L}} := \{\mathbf{a} \in A^{\mathbb{N}} : \mathbf{a}(j : j + n) \in \mathcal{L}\ \forall j \in \mathbb{N}_0\}.$$

For a given subshift of finite type $X \subset A^{\mathbb{N}}$, the *order* of the subshift is the smallest integer $n \in \mathbb{N}$ such that X is defined by a collection of admissible words of length n.

A sequence $\mathbf{a} \in X$ is *periodic* of period $p \geqslant 1$ if $T^p\mathbf{a} = \mathbf{a}$, and this is its minimal period if in addition $T^k\mathbf{a} \neq \mathbf{a}$ whenever $0 \leqslant k < p$. We will denote by $\mathrm{Per}_p(X)$ the collection of all periodic sequences of period p in X.

For a general subshift $X \subset A^{\mathbb{N}}$ and $n \geqslant 0$, the collection of X-admissible words of length $m + 1$ is the set

$$\mathcal{L}_m(X) := \{\mathbf{a}(0 : m) : \mathbf{a} \in X\}.$$

A sofic subshift $X \subset A^{\mathbb{N}}$ is a continuous T-invariant image of a subshift of finite type. More precisely, let $Y \subset A^{\mathbb{N}}$ be a subshift of finite type, B a finite alphabet and $\Pi : Y \to B^{\mathbb{N}}$ a continuous map (with respect to the distances d_A and d_B), commuting with T. The image $X = \Pi(Y)$, which in general is not of finite type, is a sofic subshift.

A more convenient way of characterizing a sofic subshift is as follows. Let $X \subset A^{\mathbb{N}}$ be a subshift, and let $\mathcal{L}^*(X) := \cup_{n=0}^{\infty}\mathcal{L}_n(X)$ be *the language* defined by X. For each $\mathbf{a} \in \mathcal{L}^*(X)$ let $f(\mathbf{a}) := \{\mathbf{b} \in \mathcal{L}^* :\ \mathbf{ab} \in \mathcal{L}^*\}$ be the set of *followers* of $\mathbf{a}$, and $p(\mathbf{a}) := \{\mathbf{b} \in \mathcal{L}^* :\ \mathbf{ba} \in \mathcal{L}^*\}$ the set of *predecessors* of $\mathbf{a}$. The subshift X is sofic if $\{f(\mathbf{a}) :\ \mathbf{a} \in \mathcal{L}^*\}$ is a finite collection, in which case $\{p(\mathbf{a}) :\ \mathbf{a} \in \mathcal{L}^*\}$ is finite as well [11].

A word $\mathbf{a} \in \mathcal{L}^*$ is a *magic word* for X if $\mathbf{b} \in p(\mathbf{a})$ and $\mathbf{c} \in f(\mathbf{a})$ implies $\mathbf{bac} \in \mathcal{L}^*$. It is a direct consequence of the finiteness of the collection of followers that every sofic subshift has a magic word (see [11, p 148]).

For a general subshift $X \subset A^{\mathbb{N}}$ and an X-admissible word $\mathbf{a} \in \mathcal{L}_m(X)$, the set

$$[\mathbf{a}] = [\mathbf{a}(0 : m)] := \{\mathbf{b} \in X :\ \mathbf{b}(0 : m) = \mathbf{a}\}$$

is the *cylinder* of length $m + 1$ determined by $\mathbf{a}$.

The subshift X is said to be *specified*, with specification length $\ell \geqslant 1$ ($\ell = 0$ means that we have a full shift) if for each pair of X-admissible words $\mathbf{a} \in \mathcal{L}_m(X)$ and $\mathbf{b} \in \mathcal{L}_n(X)$, and $k \geqslant \ell$, there exists a periodic sequence $\mathbf{c}$ of period $m + n + k + 2$, such that $\mathbf{c}(0 : m) = \mathbf{a}$ and $(T^{m+k+1}\mathbf{c})(0 : n) = \mathbf{b}$. Specification implies topological mixing and the abundance of periodic orbits in the sense that periodic orbits form a dense set in X. See [5] for more details on the specification property.

A notational remark

We shall use the symbols $\mathbf{a}$, $\mathbf{b}$, etc, both for infinite sequences and finite words for convenience. To avoid any confusion we shall always specify the nature of the $\mathbf{a}$ or $\mathbf{b}$.

2.2. Gibbs measures

The σ-field generated by cylinders of $X \subset A^{\mathbb{N}}$ coincides with the Borel σ-field $\mathcal{B}(X)$. The set $\mathcal{M}(X)$ of Borel probability measures on X is convex and compact in the weak topology. The weak topology can be metrized with the distance (see [17, p 148])

$$D(\mu, \nu) := \sum_{m=0}^{\infty} 2^{-(m+1)} \left(\sum_{\mathbf{a} \in \mathcal{L}_m(X)} |\mu[\mathbf{a}] - \nu[\mathbf{a}]| \right).$$

We denote by $\mathcal{M}_T(X)$ the set of T-invariant probability measures on X.

A function $\phi : A^{\mathbb{N}} \to \mathbb{R}$ is Hölder continuous, w.r.t. the above defined distance, if for some $\theta \in (0, 1)$ and $C > 0$, we have $\max\{|\phi(\mathbf{a}) - \phi(\mathbf{b})| : \mathbf{a}(0 : m) = \mathbf{b}(0 : m)\} \leqslant C\theta^m$ for all $m \geqslant 0$. As usual, we shall call ϕ a potential. If $\phi \equiv 0$, one has to set $C = \theta = 0$ formally in the sequel. This potential gives the measure of maximal entropy (see e.g. [11]).

For $\phi : A^{\mathbb{N}} \to \mathbb{R}$ and $k \in \mathbb{N}_0$ define $S_k\phi : A^{\mathbb{N}} \to \mathbb{R}$ as

$$S_k\phi(\mathbf{a}) = \sum_{i=0}^{k} \phi \circ T^i(\mathbf{a}).$$

Given a Hölder continuous potential ϕ and a subshift $X \subset A^{\mathbb{N}}$, $\mu \in \mathcal{M}_T(X)$ is a *Gibbs measure* for the potential ϕ if there are constants $C_{\phi,X} > 0$ and $P(\phi, X) \in \mathbb{R}$ such that

$$C_{\phi,X}^{-1} \leqslant \frac{\mu_\phi[\mathbf{a}(0 : k)]}{\exp(S_k\phi(\mathbf{a}) - (k+1)P(\phi, X))} \leqslant C_{\phi,X} \tag{1}$$

for all $\mathbf{a} \in X$, $k \in \mathbb{N}_0$.

The constant $P(\phi, X)$ above is the so called *topological pressure* of the potential ϕ. For specified subshifts, it can be defined (see e.g. [1]) by the limit

$$P(\phi, X) := \lim_{k\to\infty} \frac{1}{k+1} \log \left(\sum_{\mathbf{a}\in \mathrm{Per}_{k+1}(X)} \exp(S_k\phi(\mathbf{a}^*)) \right),$$

where $\mathbf{a}^* \in X$ is an arbitrary sequence in $[\mathbf{a}(0 : k+1)]$.

For X of finite type and ϕ Hölder continuous, there exists a unique Gibbs measure μ_ϕ (a proof of this fact can be found in [2, p 9 ff.] or [10, chapter 5]. The existence and uniqueness of μ_ϕ for general specified subshifts is a particular instance of theorem 2.5 in [8].

3. Main results

Let $X \subset A^{\mathbb{N}}$ be a specified subshift. The *finite type approximation of order* $m \in \mathbb{N}$, to X, is the subshift of finite type

$$X_m := A_{\mathcal{L}_m(X)} = \{\mathbf{a} \in A^{\mathbb{N}} : \mathbf{a}(j : j+m) \in \mathcal{L}_m(X), \forall j \in \mathbb{N}_0\},$$

determined by the X-admissible words of length $m+1$. It is easy to verify that the sequence of compact sets $\{X_m\}_{m\in\mathbb{N}}$ converges in the Hausdorff metric to X (see e.g. [5, p 111]).

On the finite type approximation X_m, the potential $\phi : A^{\mathbb{N}} \to \mathbb{R}$ determines a unique Gibbs measure $\mu_\phi^m \in \mathcal{M}_T(X_m)$. These measures will be used as *finite type approximations of order* m of $\mu_\phi \in \mathcal{M}_T(X)$.

For any $m \in \mathbb{N}$, $p \in \mathbb{N}_0$, let $\mathcal{E}_{(m,p)} \in \mathcal{M}(X_m)$ be *the elementary Gibbs measure* with support on $\mathrm{Per}_{p+1}(X_m)$, such that

$$\mathcal{E}_{(m,p)}[\mathbf{b}] := \frac{\exp(S_p\phi(\mathbf{b}))}{\sum_{\mathbf{a}\in\mathrm{Per}_{p+1}(X_m)} \exp(S_p\phi(\mathbf{a}))}$$

for each $\mathbf{b} \in \mathrm{Per}_{p+1}(X_m)$. We will use the fact [9, p 635] that each Gibbs measure μ_ϕ^m can be obtained as a weak limit of the sequence of elementary Gibbs measures $\mathcal{E}_{(m,p)}$, as $p \to \infty$.

We have the following three main results.

Theorem 3.1 (speed of convergence of μ_ϕ^m). *Let $\phi : A^{\mathbb{N}} \to \mathbb{R}$ be a Hölder continuous potential, and $X \subset A^{\mathbb{N}}$ a* specified sofic subshift. *There exist an invariant measure $\mu^* \in \mathcal{M}_T(X)$, a polynomial Q_{FT} of degree three, and constants $\theta_{\mathrm{FT}} \in (0, 1)$, $m^* \in \mathbb{N}$ satisfying*

$$D(\mu_\phi^m, \mu^*) \leqslant Q_{\mathrm{FT}}(m)\theta_{\mathrm{FT}}^m \tag{2}$$

for all $m \geqslant m^*$.

Theorem 3.2 ('Gibbs property'). *Under the hypotheses of theorem 3.1, there exists a constant* $C_g = C_g(X, \phi) > 0$ *such that*

$$C_g^{-1} \leqslant \frac{\mu^*[\mathbf{a}(0:n)]}{\exp(S_n\phi(\mathbf{a}) - (n+1)P(\phi, X))} \leqslant C_g \tag{3}$$

for each $n \in \mathbb{N}_0$ *and* $\mathbf{a} \in X$.

Theorem 3.3 ('strong mixing' property). *Under the hypotheses of theorem 3.1, there exist constants* $C_\mu > 0$ *and* $\theta_\mu \in (0, 1)$ *such that, for all* $\mathbf{a}, \mathbf{b} \in \mathcal{L}^*(X)$ *there exists* $s^* := s^*(\mathbf{a}, \mathbf{b})$ *satisfying*

$$\left| \frac{\mu^*([\mathbf{a}] \cap T^{-s}[\mathbf{b}])}{\mu^*[\mathbf{a}]\ \mu^*[\mathbf{b}]} - 1 \right| \leqslant C_\mu\, s\, \theta_\mu^{\sqrt{s}} \tag{4}$$

for all $s \geqslant s^*$.

Combining the three previous theorems we get the following one.

Theorem 3.4. *Let* $\phi : A^{\mathbb{N}} \to \mathbb{R}$ *be a Hölder continuous potential, and* $X \subset A^{\mathbb{N}}$ *a* specified sofic subshift. *The weak limit* $\mu^* = \lim_{m\to\infty} \mu_\phi^m$ *is the unique Gibbs measure associated with the potential* ϕ*, i.e. the unique* T*-invariant measure on* X *satisfying (3). Hence* $\mu^* = \mu_\phi$*. Moreover, the finite type approximations* μ_ϕ^m *converge exponentially fast to* μ_ϕ *in the sense of (2) and* μ_ϕ *is mixing in the sense of (4) and Bernoulli.*

Proof. Theorems 3.1 and 3.2 ensure the existence of a measure μ^* in $\mathcal{M}_T(X)$ satisfying inequalities (3) and having exponentially fast converging finite type approximations. To prove uniqueness, i.e. $\mu^* = \mu_\phi$, we can follow the last part of the proof of theorem 1.16 in [2]. The mixing property (4) implies weak Bernoullicity (see e.g. [16, p 169]). The theorem is proved. □

We end this section with the following theorem on the speed of convergence of the entropy $h(\mu_\phi^m)$ to $h(\mu_\phi)$, and the relative entropy $h(\mu_\phi \mid \mu_\phi^m)$ to 0.

Theorem 3.5. *Under the hypotheses of theorem 3.1, there exist constants* $C_h > 0$, $C_P > 0$, $0 < \theta_h < 1$ *and* $0 < \theta_P < 1$ *such that*

$$|h(\mu_\phi) - h(\mu_\phi^m)| \leqslant C_h\, \theta_h^m \tag{5}$$

$$h(\mu_\phi \mid \mu_\phi^m) \leqslant \frac{C_P}{1-\theta_P}\theta_P^m. \tag{6}$$

We refer the reader to [17] for details on the entropy of invariant measures. The appendix at the end of the paper contains the necessary information on the entropy and relative entropy regarding our context.

Remark 3.1. All constants appearing in the above theorems, including the coefficients of the polynomials, have explicit (but somewhat tedious) expressions in terms of the data of the problem, that is $\#A$, $\|\phi\|$, C, θ (the Hölder condition) and ℓ (the specification length). These expressions are given in the proofs.

To the best of our knowledge, theorems 3.1–3.3 and 3.5 are new. The first three imply the existence and uniqueness of μ_ϕ. The only known mixing property for this measure is that $\mu_\phi([\mathbf{a}] \cap T^{-s}[\mathbf{b}])$ converges to $\mu_\phi[\mathbf{a}]\ \mu_\phi[\mathbf{b}]$, when $s \to \infty$ [9]. This mixing is much less strong than (4) and does not imply Bernoullicity. One could ask whether the sub-exponential speed of mixing we get is optimal. Indeed, it is not the case. The reader can get convinced by looking at the proof that we could have in (4) any power of s of the form $1 - 1/d$, with $d \geqslant 2$, instead of $\frac{1}{2}$. Unfortunately, the constant C_μ goes to $+\infty$ when $d \to \infty$. We believe that the real mixing rate is exponential but our method does not allow us to prove it.

4. Technical lemmas

In this section we establish some technical lemmas needed to prove the theorems of section 3. We shall use some results coming from the theory of primitive matrices (projective distance, Birkhoff's contraction coefficient), as well as some elementary facts about the weak distance between measures. The appendix contains these results and some related notions. From now on we assume to know those results and notions, as well as the notation established there.

Notation

From now on, an expression of the type $a = c^{\pm 1}$ stands for the inequalities $c^{-1} \leqslant a \leqslant c$. Similarly $a = \pm c$ stands for $-c \leqslant a \leqslant c$. By extension, $a = \exp(\pm b)$ will stand for $\exp(-b) \leqslant a \leqslant \exp(b)$.

Given a Hölder continuous potential $\phi : A^{\mathbb{N}} \to \mathbb{R}$, for each $n \in \mathbb{N}_0$ we define the finite range potential $\phi^n : A^{n+1} \to \mathbb{R}$ such that

$$\phi^n(\mathbf{a}) = \max\{\phi(\mathbf{b}) : \mathbf{b} \in [\mathbf{a}]\}. \tag{7}$$

For $n \geqslant m$ let $\mathcal{L}_{(m,n)} := \mathcal{L}_n(X_m)$ be the set of X_m-admissible words of length $n+1$, which of course contains $\mathcal{L}_n := \mathcal{L}_n(X)$. Let us define the transfer matrix $M_{(m,n)} : \mathcal{L}_{(m,n)} \times \mathcal{L}_{(m,n)} \to \mathbb{R}^+$ such that

$$M_{(m,n)}(\mathbf{a}, \mathbf{b}) = \begin{cases} \exp(\phi^{n+1}(\mathbf{ab}(n)) & \text{if } \mathbf{a}(1:n) = \mathbf{b}(0:n-1), \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

For a specified subshift X, the matrix $M_{(m,n)}$ is primitive with primitivity index $\ell + n + 1$, and has a unique maximal eigenvalue $\rho_{(m,n)} := \rho(M_{(m,n)})$. There are unique normalized right and left eigenvectors $\mathbf{v}_{(m,n)} := \mathbf{v}_{M_{(m,n)}}$ and $\mathbf{w}_{(m,n)} := \mathbf{w}_{M_{(m,n)}}$ associated with $\rho_{(m,n)}$.

The elementary measure $\mathcal{E}_{(m,p)}$ can be expressed in terms of the transfer matrices $M_{(m,n)}$ as follows.

For $p > m \geqslant n$, and $\mathbf{a} \in \mathcal{L}_{(m,n)}$, we have

$$\mathcal{E}_{(m,p)}[\mathbf{a}] = \frac{M_{(m,n)}^{p+1}(\mathbf{a}, \mathbf{a})}{\mathrm{Trace}(M_{(m,n)}^{p+1})} \times \exp(\pm 2(p+1)C\theta^{n+1}). \tag{9}$$

Now, given $m \leqslant n$, $\mathbf{a} \in \mathcal{L}_{(m,n)}$ define $R^{\mathbf{a}}, L_{\mathbf{a}} : \mathcal{L}_{(m,n)} \to \mathbb{R}^+$ as

$$R^{\mathbf{a}}(\mathbf{b}) = M_{(m,n)}^{\ell+n+1}(\mathbf{b}, \mathbf{a}) \qquad \text{and} \qquad L_{\mathbf{a}}(\mathbf{b}) = M_{(m,n)}^{\ell+n+1}(\mathbf{a}, \mathbf{b}). \tag{10}$$

Note that these vectors are positive.

We are able to give a uniform estimate of the values of elementary measures on cylinders, using corollary 6.2.

Lemma 4.1. *Let $X \subset A^{\mathbb{N}}$ be a specified subshift, with specification length ℓ, and $\phi : A^{\mathbb{N}} \to \mathbb{R}$ be a Hölder continuous potential. There are constants $C_{\mathcal{E}} > 0$ and $\theta_{\mathcal{E}} \in (0,1)$ such that, for all integers m, n, p, such that $m \leqslant n$, $(n+1)(n+\ell+1) \leqslant p$ and $\mathbf{a} \in \mathcal{L}_{(m,n)}$, we have*

$$\mathcal{E}_{(m,p)}[\mathbf{a}] = \mathbf{w}_{(m,n)}(\mathbf{a})\mathbf{v}_{(m,n)}(\mathbf{a}) \times \exp(\pm(p+1)\, C_{\mathcal{E}}{\theta_{\mathcal{E}}}^{n+1}).$$

Proof. The proof of this lemma is based on theorem 6.1 and corollary 6.1. We are interested in bounds for the first factor of the right-hand side of (9). To this end, following corollary 6.1, we need upper bounds for both $\tau := \tau(M_{(m,n)}^{n+\ell+1})$ and $(1-\tau)^{-1}$. By definition (see (13) in the appendix)

$$\frac{1}{1-\tau} = 1 - \frac{1+\Gamma}{1-\Gamma} = \frac{1+\Gamma}{2\Gamma}$$

and since $0 < \Gamma < 1$ in our case, we have $(1-\tau)^{-1} \leqslant \Gamma^{-1}$. For simplicity of notation, we set $\bar{M} := M_{(m,n)}^{n+\ell+1}$.

Now we have

$$\Gamma^{-1} = \left(\min_{\mathbf{a},\mathbf{b},\mathbf{a}',\mathbf{b}' \in \mathcal{L}_{(m,n)}} \frac{\bar{M}(\mathbf{a},\mathbf{b})\bar{M}(\mathbf{a}',\mathbf{b}')}{\bar{M}(\mathbf{a},\mathbf{b}')\bar{M}(\mathbf{a}',\mathbf{b})} \right)^{-1/2} = \max_{\mathbf{a},\mathbf{b},\mathbf{a}',\mathbf{b}' \in \mathcal{L}_{(m,n)}} \left(\frac{\bar{M}(\mathbf{a},\mathbf{b})\bar{M}(\mathbf{a}',\mathbf{b}')}{\bar{M}(\mathbf{a},\mathbf{b}')\bar{M}(\mathbf{a}',\mathbf{b})} \right)^{1/2}.$$

Now, for arbitrary $\mathbf{a}, \mathbf{b}, \mathbf{a}', \mathbf{b}' \in \mathcal{L}_{(m,n)}$ we have

$$\begin{aligned}\frac{\bar{M}(\mathbf{a},\mathbf{b})\bar{M}(\mathbf{a}',\mathbf{b}')}{\bar{M}(\mathbf{a},\mathbf{b}')\bar{M}(\mathbf{a}',\mathbf{b})} &\leqslant \frac{\sum_{\mathbf{c}=\mathbf{a}\times\mathbf{b}\in\mathcal{L}_{(m,\ell+2n+1)}} \exp(S_{n+\ell}\,\phi^{n+1}(\mathbf{c}))}{\min_{\mathbf{c}=\mathbf{a}\times\mathbf{b}\in\mathcal{L}_{(m,\ell+2n+1)}} \exp(S_{n+\ell}\phi^{n+1}(\mathbf{c}))} \\ &\quad\times \frac{\sum_{\mathbf{c}'=\mathbf{a}'\times\mathbf{b}'\in\mathcal{L}_{(m,\ell+2n+1)}} \exp(S_{n+\ell}\phi^{n+1}(\mathbf{c}'))}{\min_{\mathbf{c}'=\mathbf{a}'\times\mathbf{b}'\in\mathcal{L}_{(m,\ell+2n+1)}} \exp(S_{n+\ell}\phi^{n+1}(\mathbf{c}'))} \\ &\leqslant ((\#A\,\mathrm{e}^{C})^{\ell} \times \mathrm{e}^{\Lambda\theta})^2,\end{aligned}$$

where we set $\Lambda := C/(1-\theta)$ from now on. Therefore

$$\frac{1}{1-\tau} \leqslant K_0 := (\mathrm{e}^{C}\#A)^{\ell} \times \mathrm{e}^{\Lambda\theta},$$

i.e. $\tau \leqslant 1 - K_0^{-1} < 1$.

For each $m \leqslant n$, let $d_{(m,n)}$ be the projective distance on the simplex of dimension $\#\mathcal{L}_{(m,n)}$, and $F_{(m,n)}$ the transformation defined on the simplex by the transition matrix $M_{(m,n)}$ (see the appendix). Note that

$$d_{(m,n)}(M_{(m,n)}R^{\mathbf{a}}, R^{\mathbf{a}}) = \log\left(\frac{\max_{\mathbf{b}} M_{(m,n)}^{n+\ell+2}(\mathbf{b},\mathbf{a})/M_{(m,n)}^{n+\ell+1}(\mathbf{b},\mathbf{a})}{\min_{\mathbf{b}} M_{(m,n)}^{n+\ell+2}(\mathbf{b},\mathbf{a})/M_{(m,n)}^{n+\ell+1}(\mathbf{b},\mathbf{a})} \right).$$

We have

$$\frac{M_{(m,n)}^{n+\ell+2}(\mathbf{b},\mathbf{a})}{M_{(m,n)}^{n+\ell+1}(\mathbf{b},\mathbf{a})} = \frac{\sum_{\mathbf{c}=\mathbf{b}\times\mathbf{a}\in\mathcal{L}_{(m,2n+\ell+2)}} \exp(S_{n+\ell+1}\phi^{n+1}(\mathbf{c}))}{\sum_{\mathbf{c}'=\mathbf{bya}\in\mathcal{L}_{(m,2n+\ell+1)}} \exp(S_{n+\ell}\phi^{n+1}(\mathbf{c}'))} = \mathrm{e}^{\pm\|\phi\|}(\#A\,\mathrm{e}^{C})^{\pm(\ell+1)}\,\mathrm{e}^{\Lambda\theta},$$

where $\|\phi\| := \max\{|\phi(\mathbf{a})| : \mathbf{a} \in A^{\mathbb{N}}\}$. From this we get

$$\max_{\mathbf{a}\in\mathcal{L}_{(m,n)}} d_{(m,n)}(F_{(m,n)}(R^{\mathbf{a}}), R^{\mathbf{a}}) \leqslant K_1$$

with

$$K_1 := 2((\ell+1)(\log(\#A) + C) + \Lambda\theta + \|\phi\|).$$

Finally, with (10), inequalities (9) may be rewritten as

$$\mathcal{E}_{(m,p)}[\mathbf{a}] = \frac{L_{\mathbf{a}}^{\dagger} M_{(m,n)}^{p+1-(n+\ell+1)} R^{\mathbf{a}}}{\sum_{\mathbf{b}\in\mathcal{L}_{(m,n)}} L_{\mathbf{b}}^{\dagger} M_{(m,n)}^{p+1-(n+\ell+1)} R^{\mathbf{b}}} \times \exp(\pm 2(p+1)C\theta^{n+1}),$$

where '†' means 'transposed'. Then, using corollary 6.2, we have

$$\begin{aligned}\mathcal{E}_{(m,p)}[\mathbf{a}] = &\frac{\mathbf{w}_{(m,n)}^{\dagger} R^{\mathbf{a}} L_{\mathbf{a}}^{\dagger} \mathbf{v}_{(m,n)}}{\sum_{\mathbf{b}\in\mathcal{L}_{(m,n)}} \mathbf{w}_{(m,n)}^{\dagger} R^{\mathbf{b}} L_{\mathbf{b}}^{\dagger} \mathbf{v}_{(m,n)}} \\ &\times \exp(\pm 2((p+1)C\theta^{n+1} + K_0K_1(n+\ell+1)(1-K_0^{-1})^{\lfloor (p+1)/(n+\ell+1) \rfloor})),\end{aligned}$$

where K_0 and K_1 are given above. On the other hand we have

$$L_{\mathbf{a}}^{\dagger}\mathbf{v}_{(m,n)} = (M_{(m,n)}^{n+\ell+1}\mathbf{v}_{(m,n)})(\mathbf{a}) = \rho_{(m,n)}^{n+\ell+1}\mathbf{v}_{(m,n)}(\mathbf{a}),$$

$$\mathbf{w}_{(m,n)}^{\dagger}R^{\mathbf{a}} = (\mathbf{w}_{(m,n)}^{\dagger}M_{(m,n)}^{n+\ell+1})(\mathbf{a}) = \rho_{(m,n)}^{n+\ell+1}\mathbf{w}_{(m,n)}(\mathbf{a})$$

and $\mathbf{w}_{(m,n)}^{\dagger}\mathbf{v}_{(m,n)} = 1$. Then, taking into account that $p+1 \geqslant (n+1)(n+\ell+1)$ and $m \leqslant n$, we obtain

$$\mathcal{E}_{(m,p)}[\mathbf{a}] = \mathbf{w}_{(m,n)}(\mathbf{a})\mathbf{v}_{(m,n)}(\mathbf{a}) \times \exp(\pm(p+1)\, C_{\mathcal{E}}\, \theta_{\mathcal{E}}{}^{n+1})$$

with $C_{\mathcal{E}} := 2(C + K_0K_1)$ and $\theta_{\mathcal{E}} := \max(1 - K_0^{-1}, \theta)$. The lemma is proved. □

Lemma 4.2. *Let $X \subset A^{\mathbb{N}}$ be a specified sofic subshift and $\phi : A^{\mathbb{N}} \to \mathbb{R}$ be a Hölder continuous potential. Then there are constants $m_X \in \mathbb{N}$, $C_X > 0$ and $\theta_X \in (0,1)$ such that for $m_X \leqslant m \leqslant p$*

$$1 - (p+1)C_X\theta_X^m \leqslant \frac{\mathcal{E}_{(m,p)}[\mathbf{a}]}{\mathcal{E}_{(m+1,p)}[\mathbf{a}]} \leqslant 1$$

for each $\mathbf{a} \in \mathrm{Per}_{p+1}(X_{m+1})$.

Proof. First note that

$$\frac{\mathcal{E}_{(m,p)}[\mathbf{a}]}{\mathcal{E}_{(m+1,p)}[\mathbf{a}]} = \frac{\sum_{\mathbf{b}\in\mathrm{Per}_{p+1}(X_{m+1})} \mathrm{e}^{S_p\phi(\mathbf{b})}}{\sum_{\mathbf{b}\in\mathrm{Per}_{p+1}(X_m)} \mathrm{e}^{S_p\phi(\mathbf{b})}}$$
$$= 1 - \frac{\sum_{\mathbf{b}\in\mathrm{Per}_{p+1}(X_m\setminus X_{m+1})} \mathrm{e}^{S_p\phi(\mathbf{b})}}{\sum_{\mathbf{b}\in\mathrm{Per}_{p+1}(X_m)} \mathrm{e}^{S_p\phi(\mathbf{b})}}$$
$$\geqslant 1 - (p+1)\frac{\sum_{\mathbf{b}\in\mathrm{Per}_{p+1}(\partial X_m)} \exp(S_p\phi(\mathbf{b}))}{\sum_{\mathbf{b}\in\mathrm{Per}_{p+1}(X_m)} \exp(S_p\phi(\mathbf{b}))},$$

where $\partial X_m := \{\mathbf{a} \in X_m : \mathbf{a}(0 : m+1) \notin \mathcal{L}_{m+1}\}$.

Let $\partial\mathcal{L}_m := \mathcal{L}_{(m,m+1)}\setminus\mathcal{L}_{m+1}$. Using the specification property we obtain

$$\frac{\sum_{\mathbf{b}\in\mathrm{Per}_{p+1}(\partial X_m)} \mathrm{e}^{S_p\phi(\mathbf{b})}}{\sum_{\mathbf{b}\in\mathrm{Per}_{p+1}(X_m)} \mathrm{e}^{S_p\phi(\mathbf{b})}} \leqslant \frac{\sum_{\mathbf{a}\in\partial\mathcal{L}_m} \mathrm{e}^{S_{m+1}\phi(\mathbf{a}^*)}}{\sum_{\mathbf{a}\in\mathcal{L}(m,m+1)} \mathrm{e}^{S_{m+1}\phi(\mathbf{a}^*)}} \times (\#A\,\mathrm{e}^{2\|\phi\|})^{\ell}\,\mathrm{e}^{4\Lambda}$$

for any $\mathbf{a}^* \in [\mathbf{a}]$. We will prove that the quotient

$$\frac{\sum_{\mathbf{a}\in\partial\mathcal{L}_m} \exp(S_{m+1}\phi(\mathbf{a}^*))}{\sum_{\mathbf{a}\in\mathcal{L}(m,m+1)} \exp(S_{m+1}\phi(\mathbf{a}^*))}$$

is exponentially small with m. This is the point at which we use the existence of magic words.

Fix a magic word $\mathbf{w} \in \mathcal{L}_k$ with $k \geqslant \ell + 1$. This is always possible since for a magic word $\mathbf{a} \in \mathcal{L}^*$, the concatenated word $\mathbf{ab}$ is again magic, for any $\mathbf{b} \in f_X(\mathbf{a})$ ($f_X(\mathbf{a})$ is the set of followers of $\mathbf{a}$, which contains arbitrary long words). Let $m \geqslant 2k(k+\ell)$, so that $\lfloor (m+1)/(k+\ell+1)\rfloor \geqslant m/k$ (we will use this condition at the final step of the proof). Note that if $\mathbf{a} \in \partial\mathcal{L}_m$, then $\mathbf{a}(i : i+k) \neq \mathbf{w}$ for each $1 \leqslant i \leqslant m-k$. This is because if $\mathbf{a}(i : i+k) = \mathbf{w}$ then $\mathbf{a}(0 : i+k), \mathbf{a}(i : m+1) \in \mathcal{L}^*(X)$, implying that $\mathbf{a} \in \mathcal{L}^*(X)$ which contradicts the hypothesis.

Letting $q := k + \ell + 1$, define

$$\partial\mathcal{L}_m^{\mathbf{w}} := \{\mathbf{a} \in \mathcal{L}_{(m,m+1)} : \mathbf{a}(jq : jq+k) \neq \mathbf{w},\ 0 \leqslant j \leqslant \lfloor (m+1)/q\rfloor - 1\}.$$

It is clear that $\partial\mathcal{L}_m \subset \partial\mathcal{L}_m^{\mathbf{w}}$. Define also

$$\epsilon^{\mathbf{w}} := \frac{\exp(S_k\phi(\mathbf{b}^-))}{\sum_{\mathbf{b}\in\mathcal{L}_k\setminus\{\mathbf{w}\}}\exp(S_k\phi(\mathbf{b}^+))} \times (\#A\,\mathrm{e}^{2\|\phi\|})^{-\ell},$$

where for each $\mathbf{b} \in \mathcal{L}_k$, the sequences $\mathbf{b}^-, \mathbf{b}^+ \in [\mathbf{b}]$ are such that $S_k\phi(\mathbf{b}^-) = \min_{\mathbf{b}^*\in[\mathbf{b}]} S_k\phi(\mathbf{b}^*)$ and $S_k\phi(\mathbf{b}^+) = \max_{\mathbf{b}^*\in[\mathbf{b}]} S_k\phi(\mathbf{b}^*)$.

Let $r := \lfloor (m+1)/q \rfloor$. For each $\omega \in \{0,1\}^r$ define

$$\mathcal{L}^{\omega}_{(m,m+1)} := \{\mathbf{a} \in \mathcal{L}_{(m,m+1)} : \mathbf{a}(jq : jq+k) = \mathbf{w} \text{ if and only if } \omega(j) = 1\}.$$

It is clear that the collection $\{\mathcal{L}^{\omega}_{(m,m+1)} : \omega \in \{0,1\}^r\}$ is a partition of $\mathcal{L}_{(m,m+1)}$. Now, it follows from the specification property that for each $\omega \in \{0,1\}^r$

$$\sum_{\mathbf{b}\in\mathcal{L}^{\omega}_{(m,m+1)}} \exp(S_{m+1}\phi(\mathbf{b}^*)) \geqslant (\epsilon^{\mathbf{w}})^{|\omega|_1} \times \sum_{\mathbf{b}\in\partial\mathcal{L}_m^{\mathbf{w}}} \exp(S_{m+1}\phi(\mathbf{b}^-)),$$

where, as before, $\mathbf{b}^- \in [\mathbf{b}]$ minimizes $S_{m+1}\phi$, and $|\omega|_1 := \sum_{j=0}^{r-1}\omega(i)$. From the previous inequality we readily derive

$$\sum_{\mathbf{b}\in\mathcal{L}_{(m,m+1)}} \exp(S_{m+1}\phi(\mathbf{b}^*)) \geqslant (1+\epsilon^{\mathbf{w}})^r \times \sum_{\mathbf{b}\in\partial\mathcal{L}_m^{\mathbf{w}}} \exp(S_{m+1}\phi(\mathbf{b}^-)).$$

Finally,

$$\frac{\sum_{\mathbf{a}\in\partial\mathcal{L}_m}\exp(S_{m+1}\phi(\mathbf{a}^*))}{\sum_{\mathbf{a}\in\mathcal{L}(m,m+1)}\exp(S_{m+1}\phi(\mathbf{a}^*))} \leqslant \frac{\sum_{\mathbf{a}\in\partial\mathcal{L}_m^{\mathbf{w}}}\exp(S_{m+1}\phi(\mathbf{a}^*))}{\sum_{\mathbf{a}\in\mathcal{L}(m,m+1)}\exp(S_{m+1}\phi(\mathbf{a}^*))} \leqslant (1+\epsilon^{\mathbf{w}})^{-r}.$$

Since $m \geqslant 2k(k+\ell)$ then $(1+\epsilon^{\mathbf{w}})^{-r} \leqslant (1+\epsilon^{\mathbf{w}})^{-m/k}$, and the result follows with

$$C_X := (\#A\,\mathrm{e}^{2\|\phi\|})^{\ell}\,\mathrm{e}^{4\Lambda}, \qquad \theta_X := (1+\epsilon^{\mathbf{w}})^{-1/k}, \qquad m_X = 2k(k+\ell).$$

The lemma is proved. □

The following lemma is of independent interest.

Lemma 4.3. *Let $X \subset A^{\mathbb{N}}$ be a specified sofic subshift and $\phi : A^{\mathbb{N}} \to \mathbb{R}$ be a Hölder continuous potential. Then there are constants $m_P \in \mathbb{N}$, $C_P > 0$ and $\theta_P \in (0,1)$ such that*

$$0 \leqslant P(\phi, X_m) - P(\phi, X_{m+1}) \leqslant C_P\theta_P^m$$

for all $m \geqslant m_P$.

Proof. Proceeding as in the proof of the previous lemma, we obtain

$$\begin{aligned} 0 &\leqslant \frac{1}{p+1}\log\left(\frac{\sum_{\mathbf{a}\in\mathrm{Per}_{p+1}(X_m)}\exp(S_{p+1}\phi(\mathbf{a}))}{\sum_{\mathbf{a}\in\mathrm{Per}_{p+1}(X_{m+1})}\exp(S_{p+1}\phi(\mathbf{a}))}\right) \\ &\leqslant \frac{1}{p+1}\log\left(1 + \frac{(p+1)C_X\theta_X^m}{1-(p+1)C_X\theta_X^m}\right) \\ &\leqslant \frac{C_X\theta_X^m}{1-(p+1)C_X\theta_X^m} \end{aligned}$$

for $m \geqslant m_X$.

To make use of the previous inequality, we need to know the speed of convergence of

$$\frac{1}{p+1}\log\left(\sum_{\mathbf{a}\in\mathrm{Per}_{p+1}(X_m)}\exp(S_p\phi(\mathbf{a}))\right) \qquad \text{to } P(\phi, X_m).$$

Some computations like the ones done to prove lemma 4.1 give

$$\sum_{\mathbf{a}\in\mathrm{Per}_p(X_m)} \exp(S_{p+1}\phi(\mathbf{a})) = \mathrm{Trace}(M_{(m,n)}^{p+1}) \times \exp(\pm(p+1)C\theta^{n+1})$$
$$= \left(\sum_{\mathbf{b}\in\mathcal{L}_{(m,n)}} \mathbf{w}_{(m,n)}^{\dagger} R^{\mathbf{b}} L_{\mathbf{b}}^{\dagger} \mathbf{v}_{(m,n)}\right) \times \rho_{(m,n)}^{p+1-2(n+\ell+1)}$$
$$\times \exp(\pm C_{\mathcal{E}}(p+1)\theta_{\mathcal{E}}^{n+1})$$
$$= \rho_{(m,n)}^{p+1} \times \exp(\pm C_{\mathcal{E}}(p+1)\theta_{\mathcal{E}}^{n+1})$$

for each $m < n$ and $(n+1)(n+\ell+1) \leqslant p$. Therefore

$$\frac{1}{p+1}\log\left(\sum_{\mathbf{a}\in\mathrm{Per}_{p+1}(X_m)} \exp(S_p\phi(\mathbf{a}))\right) = \log\rho_{(m,n)} \pm C_{\mathcal{E}}\theta_{\mathcal{E}}^{n+1}.$$

Let us now prove that $\{\rho_{(m,n)}\}_{n>m}$ converges exponentially fast. By definition, the limit has to be equal to $\exp(P(\phi, X_m))$.

Let us define $N : \mathcal{L}_{(m,n+1)} \times \mathcal{L}_{(m,n+1)} \to \mathbb{R}^+$ such that

$$N(\mathbf{a},\mathbf{b}) = \begin{cases} \exp(\phi^{n+1}(\mathbf{a})) & \text{if } \mathbf{a}(1:n+1) = \mathbf{b}(0:n), \\ 0 & \text{otherwise.} \end{cases}$$

Note that $M_{(m,n+1)} = N\exp(\pm C\theta^{n+1})$ coordinate-wise and $\rho_{(m,n)} = \exp(\pm C\theta^{n+1})\rho_N$. This can be derived easily from corollary 6.2, taking into account that $\rho_M = \lim_{n\to\infty}(\mathbf{y}^{\dagger}M^n\mathbf{x})^{1/n}$ for a primitive matrix M, and arbitrary positive vectors $\mathbf{x}, \mathbf{y}$. Let $\mathbf{v} : \mathcal{L}_{(m,n)} \to \mathbb{R}^+$ such that $\mathbf{v}(\mathbf{a}) = \exp(\phi^{n+1}(\mathbf{a})) \times \mathbf{v}_{(m,n)}(\mathbf{a}(1:n+1))$; we have

$$(N\mathbf{v})(\mathbf{a}) = \exp(\phi^{n+1}(\mathbf{a}))(M_{(m,n)}\mathbf{v}_{(m,n)})(\mathbf{a}(1:n+1))$$
$$= \exp(\phi^{n+1}(\mathbf{a}))\rho_{(m,n)}\mathbf{v}_{(m,n)}(\mathbf{a}(1:n+1)) = \rho_{(m,n)}\mathbf{x}(\mathbf{a}).$$

Hence, $\mathbf{v}$ is a positive eigenvector for the matrix N, associated with the positive eigenvalue $\rho_{(m,n)}$. Since N is primitive, corollary 6.2 implies that $\rho_N = \rho_{(m,n)}$, and therefore $\rho_{(m,n+1)} = \exp(\pm C\theta^{n+1})\rho_{(m,n)}$. From this we obtain

$$\frac{\rho_{(m,n)}}{\exp(P(\phi, X_m))} = \exp(\pm\Lambda\theta^{n+1}).$$

Since $X_m \supset X_{m+1}$, then $P(\phi, X_m) \geqslant P(\phi, X_{m+1})$. The previous computations imply on the other hand that

$$P(\phi, X_m) - P(\phi, X_{m+1}) \leqslant \frac{C_X\theta_X^m}{1 - ((m+2)(m+\ell+2)+1)C_X\theta_X^m} + C_{\mathcal{E}}\theta_{\mathcal{E}}^{m+2} + \Lambda\theta^{m+2}$$

for $m \geqslant m_X$, taking $n = m+1$ and $p = (n-1)(n+\ell+1)$. Thus, the lemma follows with

$$\theta_P := \max(\theta, \theta_X, \theta_{\mathcal{E}}),$$
$$C_P := 2C_X + C_{\mathcal{E}}\theta_{\mathcal{E}}^2 + \Lambda\theta^2$$

and $m_P := \max(m_X, m_0)$, with m_0 such that $2C_X((m+2)(m+\ell+2)+1)\theta_X^m \leqslant 1$ for all $m \geqslant m_0$. The lemma is proved. □

5. Proof of the main results

This section is devoted to the proof of theorems 3.1–3.3 and 3.5.

5.1. Proof of theorem 3.1

Lemma 4.1 implies that

$$\mathcal{E}_{(m,p+1)}[\mathbf{a}] = \exp(\pm 2(p+2)C_{\mathcal{E}}\theta^{\sqrt{p}-(\ell/2+1)})\mathcal{E}_{(m,p)}[\mathbf{a}]$$

for each $\mathbf{a} \in \cup_{k=1}^{\lfloor\sqrt{p}-(\ell/2+1)\rfloor}\mathcal{L}_{(m,k)}$. Then lemma 6.1 applies, and we obtain

$$\begin{aligned}D(\mathcal{E}_{(m,p)}, \mathcal{E}_{(m,p+1)}) &\leqslant (\exp(2C_{\mathcal{E}}(p+2)\theta_{\mathcal{E}}^{\sqrt{p}-(\ell/2+1)}) - 1) + 2^{(\ell/2+1)-\sqrt{p}}\\ &\leqslant 4C_{\mathcal{E}}(p+2)\,\theta_{\mathcal{E}}^{\sqrt{p}-(\ell/2+1)} + 2^{(\ell/2+1)-\sqrt{p}}\end{aligned}$$

for each $p \geqslant \max(p_0, (m+2)(m+\ell+2))$, with

$$p_0 := \min\{p \in \mathbb{N} :\ 2C_{\mathcal{E}}(k+2)\theta_{\mathcal{E}}^{\sqrt{k}-(\ell/2+1)} \leqslant 1 \text{ for all } k \geqslant p\}.$$

Since $\sum_{p=0}^{\infty}(p+2)\,\theta_{\mathcal{E}}^{\sqrt{p}} < \infty$, there exists a limit measure $\mu^m := \lim_{p\to\infty}\mathcal{E}_{(m,p)}$ belonging to $\mathcal{M}_T(X_m)$. The convergence is such that

$$D(\mu^m, \mathcal{E}_{(m,p)}) \leqslant 4C_{\mathcal{E}}\theta_{\mathcal{E}}^{-(\ell/2+1)}Q(\sqrt{p})\theta_{\mathcal{E}}^{\sqrt{p}} + 2^{(\ell/2+3)}2^{-\sqrt{p}}(\sqrt{p}+3)$$

for each $p \geqslant \max(p_0, (m+2)(m+\ell+2))$. Here

$$Q(x) := -\frac{2x(x^2+3)}{\log(\theta_{\mathcal{E}})} + \frac{6(x^2+1)}{\log^2(\theta_{\mathcal{E}})} - \frac{12x}{\log^3(\theta_{\mathcal{E}})} + \frac{12}{\log^4(\theta_{\mathcal{E}})}.$$

Let us now prove that the limiting measure μ^m coincides with the unique Gibbs measure $\mu_\phi \in \mathcal{M}_T(X_m)$. From the specification property we can derive the inequalities

$$\begin{aligned}\mathcal{E}_{(m,p)}[\mathbf{a}] = \exp&\left(S_n\phi(\mathbf{a}^*) - \log\left(\sum_{\mathbf{b}\in\mathrm{Per}_{n+1}(X_m)}\exp(S_n\phi(\mathbf{b}))\right)\right)\\ &\times \exp\left(\pm\,(3\ell\|\phi\|\log(\#A) + 5\Lambda)\right),\end{aligned}$$

which hold for any $n \leqslant p$, $\mathbf{a} \in \mathcal{L}_{(m,n)}$ and $\mathbf{a}^* \in [\mathbf{a}]$. On the other hand, the computations performed in the proof of lemma 4.3 lead to the inequalities

$$\log\left(\sum_{\mathbf{b}\in\mathrm{Per}_{n+1}(X_m)}\exp(S_n\phi(\mathbf{b}))\right) = (n+1)(P(\phi, X_m) \pm (C_{\mathcal{E}}\theta_{\mathcal{E}}^{n+1} + \Lambda\theta^{n+1})),$$

for each $m \leqslant n$ and n such that $(n+1)(n+\ell+1) \leqslant p$. Since $\mathcal{E}_{(m,p+1)}[\mathbf{a}] = \exp(\pm 2 \times (p+2)C_{\mathcal{E}}\theta^{\sqrt{p}-(\ell/2+1)})\mathcal{E}_{(m,p)}[\mathbf{a}]$, it follows by induction that

$$\mu^m[\mathbf{a}] = \mathcal{E}_{(m,p)}[\mathbf{a}] \times \exp(\pm 2C_{\mathcal{E}}\theta_{\mathcal{E}}^{-(\ell/2+1)}Q(\sqrt{p})\theta_{\mathcal{E}}^{\sqrt{p}})$$

for each $\mathbf{a} \in \cup_{k=1}^{\lfloor\sqrt{p}-(\ell/2+1)\rfloor}\mathcal{L}_{(m,k)}$. Therefore, for each $m \leqslant n$, $\mathbf{a} \in \mathcal{L}_{(m,n)}$, and $\mathbf{a}^* \in [\mathbf{a}]$, we have

$$\frac{\mu^m[\mathbf{a}]}{\exp(S_n\phi(\mathbf{a}^*) - (n+1)P(\phi, X_m))} = \exp(\pm C_{\mathrm{FT}})$$

with

$$\begin{aligned}C_{\mathrm{FT}} := {}& 2C_{\mathcal{E}}\theta_{\mathcal{E}}^{-(\ell/2+1)}\max\{Q(k)\theta_{\mathcal{E}}^{k} :\ k \in \mathbb{N}\} + 3\ell\|\phi\|\log(\#A)\\ &+ 5\Lambda + \max\{(n+1)(C_{\mathcal{E}}\theta_{\mathcal{E}}^{n+1} + \Lambda\theta^{n+1}) :\ n \in \mathbb{N}\}.\end{aligned}$$

Now, for $\mathbf{a} \in \mathcal{L}_n$ with $n \leqslant m$, we obtain

$$\begin{aligned}\mu^m[\mathbf{a}] &= \sum_{\mathbf{b}\in\mathcal{L}_{m,n+k}\cap[\mathbf{a}]} \mu^m[\mathbf{b}]\\ &= \exp(\pm C_{\mathrm{FT}}) \sum_{\mathbf{b}\in\mathcal{L}_{m,n+k}\cap[\mathbf{a}]} \exp(S_{n+k}\phi(\mathbf{b}^*) - (n+k+1)P(\phi, X_m))\\ &= \exp[S_n\phi(\mathbf{a}^*) - (n+1)P(\phi, X_m) \pm (C_{\mathrm{FT}} + \ell(\log(\#A) + \|\phi\|) + \Lambda)]\\ &\quad\times \sum_{\mathbf{b}\in\mathcal{L}_{m,k-1}} \mathrm{e}^{S_{k-1}\phi(\mathbf{b}^*) - kP(\phi, X_m)}\\ &= \exp[S_n\phi(\mathbf{a}^*) - (n+1)P(\phi, X_m) \pm (2C_{\mathrm{FT}} + \ell(\log(\#A) + \|\phi\|) + \Lambda)] \sum_{\mathbf{b}\in\mathcal{L}_{m,k-1}} \mu^m[\mathbf{b}]\\ &= \exp(S_n\phi(\mathbf{a}^*) - (n+1)P(\phi, X_m)) \exp(\pm C_g)\end{aligned}$$

by using the specification property, and for k sufficiently large. Here

$$C_g := 2C_{\mathrm{FT}} + \ell(\log(\#A) + \|\phi\|) + \Lambda).$$

This way we prove that μ^m satisfies 'Gibbs inequalities' (3). Theorem 1.16 in [2], establishes the existence and uniqueness of the Gibbs measure $\mu_\phi^m \in \mathcal{M}_T(X_m)$, implies that $\mu^m := \mu_\phi^m$.

Let $\tilde{m} = \min\{m \in \mathbb{N} : 4((k+\ell+1)^2+1)C_X\theta_X^k \leqslant 1 \text{ for all } k \geqslant m\}$, and m_X be as in lemma 4.2. From lemmas 6.2 and 4.2, and following the computations in the first part of this proof, we obtain

$$\begin{aligned}&D(\mathcal{E}_{(m,(m+\ell+1)^2)}, \mathcal{E}_{(m+1,(m+\ell+2)^2)})\\ &\quad\leqslant D(\mathcal{E}_{(m,(m+\ell+1)^2)}, \mathcal{E}_{(m,(m+\ell+2)^2)}) + D(\mathcal{E}_{(m,(m+\ell+2)^2)}, \mathcal{E}_{(m+1,(m+\ell+2)^2)})\\ &\quad\leqslant 4((m+\ell+2)^2+1)C_{\mathcal{E}}\theta_{\mathcal{E}}^{m+1} + 2^{-m} + 8((m+\ell+2)^2+1)C_X\theta_X^m\end{aligned}$$

for all $m \geqslant m^*$, with $m^* = \max(m_X, \tilde{m})$.

Since $\sum_{m=0}^{\infty}(m+\ell+2)^2\max(\theta_{\mathcal{E}}, \theta_X)^m$ is finite, $\mu^* := \lim_{m\to\infty}\mathcal{E}_{(m,(m+\ell+1)^2)}$ is a well-defined measure in $\mathcal{M}_T(X)$. Furthermore, the convergence is such that

$$D(\mu^*, \mathcal{E}_{(m,(m+\ell+1)^2)}) \leqslant 2^{-m+1} + 4C_{\mathcal{E}}Q_{\mathcal{E}}(m)\theta_{\mathcal{E}}^m + 8C_XQ_X(m)\theta_X^{m-1}$$

with

$$Q_{\mathcal{E}}(x) := -\frac{(x+\ell+2)^2+1}{\log(\theta_{\mathcal{E}})} + \frac{2(x+\ell+2)}{\log^2(\theta_{\mathcal{E}})} - \frac{2}{\log^3(\theta_{\mathcal{E}})},$$

$$Q_X(x) := -\frac{(x+\ell+2)^2+1}{\log(\theta_X)} + \frac{2(x+\ell+2)}{\log^2(\theta_X)} - \frac{2}{\log^3(\theta_X)}.$$

Therefore, for any $m \geqslant m^*$, one has

$$D(\mu^*, \mu_\phi^m) \leqslant D(\mu^*, \mathcal{E}_{(m,(m+\ell+1)^2)}) + D(\mu_\phi^m, \mathcal{E}_{(m,(m+\ell+1)^2)}) \leqslant Q_{\mathrm{FT}}(m)\theta_{\mathrm{FT}}^m$$

with

$$Q_{\mathrm{FT}}(m) := 4C_{\mathcal{E}}(\theta_{\mathcal{E}}^{-(\ell/2+1)}Q(m) + \theta_{\mathcal{E}}^{-1}Q_{\mathcal{E}}(m)) + 8C_X\theta_X^{-1}Q_X(m) + (m+3)2^{(\ell/2+3)} + 2$$

and

$$\theta_{\mathrm{FT}} := \max\left(\theta_{\mathcal{E}}, \theta_X, \tfrac{1}{2}\right).$$

Remark 5.1. In the previous proof, the polynomials Q, $Q_{\mathcal{E}}$ and Q_X were obtained by upper bounding the series $\sum_{k=m}^{\infty} P(k)\eta^k$, with $P(x)$ an increasing polynomial, and $\eta \in (0,1)$, by the integral $\eta^{-1}\int_m^{\infty} P(x)\eta^x\,\mathrm{d}x$. Then we used the identity

$$\int_m^{\infty} P(x)\eta^x\,\mathrm{d}x = \eta^m \times \sum_{k=0}^{\deg(P)} \left(-\frac{1}{\log(\eta)}\right)^{k+1} P^{(k)}(m),$$

where $P^{(k)}$ is the kth derivative of P.

5.2. Proof of theorem 3.2

In the previous proof we derived the inequalities

$$\frac{\mu_\phi^m[\mathbf{a}]}{\exp(S_n\phi(\mathbf{a}^*) - (n+1)P(\phi, X_m))} = \exp(\pm C_g)$$

for all $n \in \mathbb{N}$, $\mathbf{a} \in \mathcal{L}_{(m,n)}$ and $\mathbf{a}^* \in [\mathbf{a}]$.

On the other hand, lemma 4.3 ensures that $P(\phi, X_m) = P(\phi, X) \pm C_P\theta_P^m$; therefore

$$\frac{\mu_\phi^m[\mathbf{a}]}{\exp(S_n\phi(\mathbf{a}^*) - (n+1)P(\phi, X_m))} = \exp(\pm(C_g + (n+1)C_P\theta_P^m))$$

is valid for each $n \in \mathbb{N}$, $\mathbf{a} \in \mathcal{L}_{(m,n)}$ and $\mathbf{a}^* \in [\mathbf{a}]$. Taking the limit $m \to \infty$, we obtain the desired result.

5.3. Proof of theorem 3.3

Proceeding as in the proof of lemma 4.2, the specification property implies for all $n < m < m'$

$$\begin{aligned}\frac{\mathcal{E}_{(m,(m'+\ell+1)^2)}[\mathbf{a}]}{\mathcal{E}_{(m',(m'+\ell+1)^2)}[\mathbf{a}]} &= \frac{\sum_{\mathbf{b}\in \mathrm{Per}_{(m'+\ell+1)^2+1}(X_{m'})} \exp(S_p\phi(\mathbf{b}))}{\sum_{\mathbf{b}\in \mathrm{Per}_{(m'+\ell+1)^2+1}(X_m)} \exp(S_p\phi(\mathbf{b}))} \\ &\quad\times \frac{\sum_{\mathbf{b}\in \mathrm{Per}_{(m'+\ell+1)^2+1}(X_m)\cap[\mathbf{a}]} \exp(S_p\phi(\mathbf{b}))}{\sum_{\mathbf{b}\in \mathrm{Per}_{(m'+\ell+1)^2+1}(X_{m'})\cap[\mathbf{a}]} \exp(S_p\phi(\mathbf{b}))} \\ &= \exp\left(\pm\left(4\ell(\log(\#A) + \|\phi\| + \Lambda) + 4C_X\sum_{k=m}^{m'}((k+\ell+1)^2+1)\theta_X^k\right)\right)\end{aligned}$$

for each $\mathbf{a} \in \mathcal{L}_{(m,n)} \equiv \mathcal{L}_n$, as long as $m \geqslant m^*$ (where m^* is defined in the course of the proof of theorem 3.1). These inequalities can be viewed as extensions to cylinders of the inequalities of lemma 4.2.

On the other hand, lemma 4.1 ensures that

$$\mathcal{E}_{(m,(m+\ell+1)^2)}[\mathbf{a}] = \mathcal{E}_{(m,(m+\ell+2)^2)}[\mathbf{a}] \times \exp(\pm 2((m+\ell+2)^2+1)C_{\mathcal{E}}\theta_{\mathcal{E}}^{m+1}).$$

These and the previous inequalities imply that $\mu^*[\mathbf{a}] = \mathcal{E}_{(m,(m+\ell+1)^2)}[\mathbf{a}]\exp(\pm\gamma_{\mathrm{FT}})$ for each $m \geqslant m^*$, $m \geqslant n$ and $\mathbf{a} \in \mathcal{L}_n$. Here

$$\begin{aligned}\gamma_{\mathrm{FT}} &:= 4\ell(\log(\#A) + \|\phi\|) + 4\Lambda + \sum_{k=m^*}^{\infty}((k+1)^3+1)(4C_X\theta_X^k + 2C_{\mathcal{E}}\theta_{\mathcal{E}}^{k+1}) \\ &= 4\ell(\log(\#A) + \|\phi\|) + 4\Lambda + 4C_X Q_X(m^*)\,\theta_X^{m^*-1} + 2C_{\mathcal{E}}Q_{\mathcal{E}}(m^*)\theta^{m^*}.\end{aligned}$$

Because of the previous inequalities,

$$\begin{aligned}&|\mu^*([\mathbf{a}]\cap T^{-s}[\mathbf{b}]) - \mu^*[\mathbf{a}]\mu^*[\mathbf{b}]| \\ &\qquad\leqslant \mathrm{e}^{2\gamma_{\mathrm{FT}}}|\mathcal{E}_{(m,(m+\ell+1)^2)}([\mathbf{a}]\cap T^{-s}[\mathbf{b}]) - \mathcal{E}_{(m,(m+\ell+1)^2)}[\mathbf{a}]\mathcal{E}_{(m,(m+\ell+1)^2)}[\mathbf{b}]|\end{aligned}$$

for every $\mathbf{a} \in \mathcal{L}_n$, $\mathbf{b} \in \mathcal{L}_{n'}$ and $s \in \mathbb{N}$, as long as $n + n' + s \leqslant m$.

Let us now prove a product formula for the elementary measure $\mathcal{E}_{(m,(m+\ell+1)^2)}$. Take $\mathbf{a}, \mathbf{b} \in \mathcal{L}_m$, and let $p := (m+\ell+1)^2$. Let $L_{\mathbf{a}}$ and $R^{\mathbf{a}}$ be as in the proof of lemma 4.1;

then, for $s \geqslant m + \ell + 1$ we have

$$\mathcal{E}_{(m,p)}([\mathbf{a}] \cap T^{-s}[\mathbf{b}]) = \frac{\sum_{\mathbf{c} \in \mathrm{Per}_{p+1}(X_m):\ \mathbf{c}(0:m)=\mathbf{a},\ \mathbf{c}(m+s+1:2m+s+1)=\mathbf{b}} \exp(S_p\phi(\mathbf{c}))}{\sum_{\mathbf{c} \in \mathrm{Per}_{p+1}(X_m)} \exp(S_p\phi(\mathbf{c}))}$$

$$= \frac{L_{\mathbf{a}}^{\dagger} M_{(m,m)}^{p_1} R^{\mathbf{b}} \times L_{\mathbf{b}}^{\dagger} M_{(m,m)}^{p_2} R^{\mathbf{a}}}{\sum_{\mathbf{a} \in \mathcal{L}_{(m,m)}} \sum_{\mathbf{b} \in \mathcal{L}_{(m,m)}} L_{\mathbf{a}}^{\dagger} M_{(m,m)}^{p_1} R^{\mathbf{b}} \times L_{\mathbf{b}}^{\dagger} M_{(m,m)}^{p_2} R^{\mathbf{a}}} \times \exp(\pm 2(p+1)C\theta^{m+1}),$$

where $p_1 = s - 2\ell + m$ and $p_2 = p + 1 - s + 2\ell + m$. Suppose that s is such that $p_1 \leqslant p_2$; then, taking into account that $C_{\mathcal{E}} \geqslant C$ and $\theta_{\mathcal{E}} \geqslant \theta$, we obtain

$$\mathcal{E}_{(m,p)}([\mathbf{a}] \cap T^{-s}[\mathbf{b}]) = \frac{\mathbf{v}_{(m,m)}(\mathbf{a})\mathbf{w}_{(m,m)}(\mathbf{b}) \times \mathbf{w}_{(m,m)}(\mathbf{a})\mathbf{v}_{(m,m)}(\mathbf{b})}{\sum_{\mathbf{a} \in \mathcal{L}_{(m,m)}} \sum_{\mathbf{b} \in \mathcal{L}_{(m,m)}} \mathbf{v}_{(m,m)}(\mathbf{a})\mathbf{w}_{(m,m)}(\mathbf{b}) \times \mathbf{w}_{(m,m)}(\mathbf{a})\mathbf{v}_{(m,m)}(\mathbf{b})} \times \exp(\pm(4(p+1)C_{\mathcal{E}}\theta_{\mathcal{E}}^{\lfloor p_1/(m+\ell+1)\rfloor})).$$

Here we have used, as in the proof of lemma 4.1, the fact that

$$L_{\mathbf{a}}^{\dagger} M_{(m,m)}^{p_1} R^{\mathbf{b}} = \rho_{(m,m)}^{p_1+2(\ell+m+1)} \mathbf{w}_{(m,m)}(\mathbf{a})\mathbf{v}_{(m,m)}(\mathbf{b}) \exp(\pm C_{\mathcal{E}}(m+\ell+1)\theta_{\mathcal{E}}^{\lfloor p_1/(m+\ell+1)\rfloor}).$$

Now, since

$$\sum_{\mathbf{a} \in \mathcal{L}_{(m,m)}} \sum_{\mathbf{b} \in \mathcal{L}_{(m,m)}} \mathbf{v}_{(m,m)}(\mathbf{a})\mathbf{w}_{(m,m)}(\mathbf{b}) \times \mathbf{w}_{(m,m)}(\mathbf{a})\mathbf{v}_{(m,m)}(\mathbf{b}) = \left(\sum_{\mathbf{a} \in \mathcal{L}_{(m,m)}} \mathbf{v}_{(m,m)}(\mathbf{a})\mathbf{w}_{(m,m)}(\mathbf{a}) \right)^2 = 1,$$

we finally obtain

$$\mathcal{E}_{(m,p)}([\mathbf{a}] \cap T^{-s}[\mathbf{b}]) = \mathcal{E}_{(m,p)}[\mathbf{a}]\mathcal{E}_{(m,p)}[\mathbf{b}] \times \exp(\pm(4(p+1)C_{\mathcal{E}}\theta_{\mathcal{E}}^{\lfloor (s-(2\ell+m))/(m+\ell+1)\rfloor}))$$

for each $\mathbf{a}, \mathbf{b} \in \mathcal{L}_m$. Because of the additivity of the measure $\mathcal{E}_{(m,p)}$, these inequalities extend to any $\mathbf{a}, \mathbf{b} \in \cup_{k=0}^{m} \mathcal{L}_k$.

Combining the inequalities we have just derived with those we obtained in the first part of the proof, we get

$$\left| \frac{\mu^*([\mathbf{a}] \cap T^{-s}[\mathbf{b}])}{\mu^*[\mathbf{a}]\mu^*[\mathbf{b}]} - 1 \right| \leqslant \mathrm{e}^{4\gamma_{\mathrm{FT}}}(\exp(4(p+1)C_{\mathcal{E}}\theta_{\mathcal{E}}^{\lfloor (s-(2\ell+m))/(m+\ell+1)\rfloor}) - 1)$$

for each $\mathbf{a}, \mathbf{b} \in \cup_{k=0}^{m} \mathcal{L}_k$.

To finish the proof, it only remains for us to bind the term

$$\exp(4(p+1)C_{\mathcal{E}}\theta_{\mathcal{E}}^{\lfloor (s-(2\ell+m))/(m+\ell+1)\rfloor}) - 1$$

by a function depending only on s. For this, note that $m + \ell + 1 \leqslant s$ and $2s \leqslant (m+\ell+1)^2 - 2(2\ell+m) + 1$. Let $\bar{m} \geqslant m^*$ be such that $(m+\ell+1)^2 - 2(2\ell+m) + 1 \geqslant 10m^2/11$ and $5m^2/11 \geqslant m + \ell + 1$ for $m \geqslant \bar{m}$. Then, for $s \geqslant s_0 := 5\bar{m}^2/11$ define $m(s)$ to be the only positive integer satisfying $5(m(s) - 1)^2/11 \leqslant s < 5m(s)^2/11$. By taking $m = m(s)$ in the previous inequalities we have

$$\left| \frac{\mu^*([\mathbf{a}] \cap T^{-s}[\mathbf{b}])}{\mu^*[\mathbf{a}]\mu^*[\mathbf{b}]} - 1 \right| \leqslant \mathrm{e}^{4\gamma_{\mathrm{FT}}}(\exp(4((\sqrt{11s/5} + \ell + 1)^2 + 1)C_{\mathcal{E}}\theta_{\mathcal{E}}^{\lfloor (s-2\ell-\sqrt{11s/5})/(\sqrt{11s/5}+\ell)\rfloor}) - 1)$$

for each $\mathbf{a}, \mathbf{b} \in \cup_{k=0}^{m(s)} \mathcal{L}_k$.

As a final step we propose a simplification of the previous expression, a simplification which will hold for s large enough. Let $s_1 \geqslant s_0$ be such that $(s-2\ell-\sqrt{11s/5})/(\sqrt{11s/5}+\ell) \geqslant \sqrt{5s/12}$ and $(\sqrt{11s/5}+\ell+1)^2+1 \leqslant 12s/5$ for every $s \geqslant s_1$. Then, for $s \geqslant s_1$ we have

$$\left|\frac{\mu^*([\mathbf{a}]\cap T^{-s}[\mathbf{b}])}{\mu^*[\mathbf{a}]\mu^*[\mathbf{b}]}-1\right| \leqslant \mathrm{e}^{4\gamma_{\mathrm{FT}}}\left(\exp\left(\frac{48s}{5}C_{\mathcal{E}}\theta_{\mathcal{E}}^{\lfloor\sqrt{5s/12}\rfloor}\right)-1\right)$$

for each $\mathbf{a},\mathbf{b}\in\cup_{k=0}^{m(s)}\mathcal{L}_k$. Finally, taking $s_2 \geqslant s_1$ such that

$$\frac{48s}{5}C_{\mathcal{E}}\theta_{\mathcal{E}}^{\lfloor\sqrt{5s/12}\rfloor} \leqslant 1$$

for every $s \geqslant s_2$, and since $\mathrm{e}^{\epsilon}-1 \leqslant 2\epsilon$ for $0 \leqslant \epsilon \leqslant 1$, we can ensure that

$$\left|\frac{\mu^*([\mathbf{a}]\cap T^{-s}[\mathbf{b}])}{\mu^*[\mathbf{a}]\mu^*[\mathbf{b}]}-1\right| \leqslant \frac{96}{5}C_{\mathcal{E}}\theta_{\mathcal{E}}^{-1}\,\mathrm{e}^{4\gamma_{\mathrm{FT}}}\ s\times\theta_{\mathcal{E}}^{\sqrt{5s/12}}$$

for $s \geqslant s_2$ and $\mathbf{a},\mathbf{b}\in\cup_{k=0}^{m(s)}\mathcal{L}_k$. The result follows with

$$\theta_\mu := \theta_{\mathcal{E}}^{\sqrt{5/12}}$$

$$s^*(\mathbf{a},\mathbf{b}) := \max(s_2, \lfloor 5\max(n,n')^2/11\rfloor)$$

$$C_\mu := \frac{96}{5}\,C_{\mathcal{E}}\theta_{\mathcal{E}}^{-1}\,\mathrm{e}^{4\gamma_{\mathrm{FT}}}.$$

5.4. *Proof of theorem 3.5*

By [2], each measure μ_ϕ^m satisfies the variational principle, as well as the measure μ_ϕ by [1]. This means in particular the following:

$$P(\phi,X_m)=\int_{X_m}\phi\,\mathrm{d}\mu_\phi^m+h(\mu_\phi^m) \qquad \text{and} \qquad P(\phi,X)=\int_X\phi\,\mathrm{d}\mu_\phi+h(\mu_\phi). \tag{11}$$

Hence we have

$$|h(\mu_\phi^m)-h(\mu_\phi)| \leqslant |P(\phi,X_m)-P(\phi,X)|+\left|\int_X\phi\,\mathrm{d}\mu_\phi-\int_{X_m}\phi\,\mathrm{d}\mu_\phi^m\right|.$$

It is obvious from lemma 4.3 that

$$0<P(\phi,X_m)-P(\phi,X)\leqslant\frac{C_P}{1-\theta_P}\theta_P^m. \tag{12}$$

On the other hand,

$$\left|\int_X\phi\,\mathrm{d}\mu_\phi-\int_{X_m}\phi\,\mathrm{d}\mu_\phi^m\right|\leqslant C\theta^m.$$

Statement (5) is thus proved.

Now, applying (15) (see the appendix) and using (11) and (12) we get

$$h(\mu_\phi\mid\mu_\phi^m)=P(\phi,X_m)-\left(\int_X\phi\,\mathrm{d}\mu_\phi+h(\mu_\phi)\right)=P(\phi,X_m)-P(\phi,X)\leqslant\frac{C_P}{1-\theta_P}\theta_P^m.$$

This proves (6). The proof of the theorem is now complete.

6. Examples and generalizations

A natural class of specified sofic subshifts is provided by β-shifts coding the dynamics of the map on the unit interval $T_\beta : x \mapsto \beta x \bmod 1$, where $\beta > 1$ is a real number. For certain β, the corresponding β-shift is a specified sofic subshift. In [12], the author constructs a sofic coding of hyperbolic automorphisms of the torus. In both cases, the Lebesgue measure on the unit interval or the torus is sent to the measure of maximal entropy on the coding subshift.

In this paper we assumed, for the sake of definiteness, that the potential ϕ was Hölder continuous and the subshift $X \subset A^{\mathbb{N}}$ was a specified sofic subshift. Nevertheless, both assumptions can be weakened. In the proof of theorem 3.1, and in all other computations, the exponential decay

$$\max\{|\phi(\mathbf{a}) - \phi(\mathbf{b})| : \mathbf{a}(0:m) = \mathbf{b}(0:m)\} \leqslant C\theta^m$$

can be replaced by a polynomial decay

$$\max\{|\phi(\mathbf{a}) - \phi(\mathbf{b})| : \mathbf{a}(0:m) = \mathbf{b}(0:m)\} \leqslant Cm^{-\alpha},$$

as long as $\alpha > 4$. In this case, the speed of convergence of the topological pressure (lemma 4.3) as well as the speed of convergence of entropy in theorem 3.5 becomes polynomial.

Regarding the nature of the subshift, the reader can verify that the essential assumptions are the specification and presence of magic words. Moreover, the latter assumption is only used in lemma 4.2. Specified sofic subshifts form a natural class of subshifts having the specification property as well as magic words, but there are huge classes of non-sofic specified subshifts with magic words. Among them, we can mention the class of non-sofic specified β-shifts (see [14]). One can straightforwardly prove that for each non-sofic specified β-shift there exists $k \in \mathbb{N}$ such that 0^k is a magic word.

On the other hand, following the examples in [6] we can obtain non-sofic specified subshifts with magic words, as finitary codings of Bernoulli shifts. Take for example, the finitary coding $\pi : \{0, 1, 2, 3\}^{\mathbb{N}} \to \{0, 1, 2, 3\}^{\mathbb{N}}$ such that

$$(\pi\mathbf{a})_n = \begin{cases} 0 & \text{if } \mathbf{a}(0:2k+1) = 32^k1^k0 \text{ for some } k \in \mathbb{N}, \\ \mathbf{a}(n) & \text{otherwise.} \end{cases}$$

The image subshift $X := \overline{\pi\{0, 1, 2, 3\}^{\mathbb{N}}}$ is not sofic: its description involves a non-regular language. Nevertheless it has the specification property: we may connect any two admissible words by words of the kind $12^\ell 1$, and 3 is a magic letter. Any product measure on $\{0, 1, 2, 3\}^{\mathbb{N}}$ induces a Gibbs measure in X, which can be approximated by our method.

Though the class of systems considered here is only a subclass of those covered by theorem 2.5 in [8], we are able to obtain a speed of convergence (in the weak distance) of finite type approximations of the Gibbs measure on the approximated subshift X. We are also able to prove a strong mixing property, implying Bernoullicity. Finally, we provide a speed of convergence of the entropy of the finite type approximations to the entropy of the Gibbs measure on X. We also emphasize that all constants appearing in the statements of section 3 have explicit expressions in terms of the data of the problem. It is also worth noting that we only used classical algebraic tools and symbolic dynamics, except for the uniqueness of μ_ϕ for which we used Bowen's argument.

Acknowledgments

We thank K Petersen for providing us with [7] and F Ledrappier for useful comments.

Appendices

Appendix 1. Primitive matrices

$M : \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\} \to [0, \infty)$ is said to be *primitive* if there exists an integer $\ell \geqslant 1$ such that $M^{\ell} > 0$. The smallest such integer is the *primitivity index* of M.

For M primitive let

$$\Gamma(M) := \begin{cases} \sqrt{\min_{i,j,k,l} \dfrac{M(i,j)M(k,l)}{M(i,l)M(k,j)}} & \text{if } M > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

The *Birkhoff's coefficient* for M is $\tau(M) := (1 - \Gamma(M))/(1 + \Gamma(M))$.

Consider the function $d : (\mathbb{R}^{+})^{n} \times (\mathbb{R}^{+})^{n} \to \mathbb{R}^{+}$ such that

$$d(\mathbf{x}, \mathbf{y}) = \log\left(\frac{\max_i \mathbf{x}(i)/\mathbf{y}(j)}{\min_i \mathbf{x}(i)/\mathbf{y}(i)}\right). \tag{14}$$

It is the projective distance when restricted to the simplex

$$\Delta_n := \left\{ \mathbf{x} : \{1, 2, \ldots, n\} \to (0, 1) : |\mathbf{x}|_1 := \sum_{i=1}^{n} \mathbf{x}(i) = 1 \right\}.$$

Birkhoff's coefficient gives the contraction rate of the iterated action of M on vectors in Δ_n.

Theorem 6.1. *Let M, Δ_n and d be defined as above and define $F_M : \Delta_n \to \Delta_n$ as*

$$F_M \mathbf{x} := \frac{M\mathbf{x}}{|M\mathbf{x}|_1}.$$

Then F_M is a contraction in (Δ_n, d) with contraction coefficient $\tau(M)$, i.e.

$$d(F_M \mathbf{x}, F_M \mathbf{y}) \leqslant \tau(M) d(\mathbf{x}, \mathbf{y}), \qquad \forall\, \mathbf{x}, \mathbf{y} \in \Delta_n.$$

A proof of this result can be derived easily from theorem 3.12 in [15, p 108].

The previous result directly implies the Perron–Frobenius theorem (see [15, chapter 1] for more details): a primitive matrix M has only one maximal eigenvalue $\rho_M > 0$. Associated with it there is a unique right eigenvector $\mathbf{v}_M \in \Delta_n$, and a unique left eigenvector $\mathbf{w}_M > 0$ such that $\mathbf{w}_M^{\dagger} \mathbf{v}_M = 1$.

A rather direct consequence of the previous theorem is the following corollary.

Corollary 6.1. *For M primitive with primitivity index ℓ, let $F := F_M$ and $\tau := \tau(M^{\ell})$. Then, for each $\mathbf{x} \in \Delta_n$ and $m \in \mathbb{N}$, we have*

$$d(F^m \mathbf{x}, \mathbf{v}_M) \leqslant \frac{\tau^{\lfloor m/\ell \rfloor}}{1 - \tau} \times d_M(\mathbf{x})$$

with $d_M(\mathbf{x}) := \min(\ell\, d(\mathbf{x}, F\mathbf{x}), d(\mathbf{x}, F^{\ell}\mathbf{x}))$.

From this we readily deduce the following corollary.

Corollary 6.2. *Let $M : \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\} \to \mathbb{R}^{+}$ be a primitive matrix with primitivity index ℓ, $F := F_M$ and $\tau := \tau(M^{\ell})$. Then, for each $\mathbf{x} \in \Delta_n$ and $m \in \mathbb{N}$ we have*

$$M^m \mathbf{x} = \rho_M^m (\mathbf{w}_M^{\dagger} \mathbf{x}) \mathbf{v}_M \exp\left(\pm \frac{\tau^{\lfloor m/\ell \rfloor} d_M(\mathbf{x})}{1 - \tau}\right)$$

with $d_M(\mathbf{x}) := \min(\ell\, d(\mathbf{x}, F\mathbf{x}), d(\mathbf{x}, F^{\ell}\mathbf{x}))$.

Proof. Since $M^m\mathbf{x} = |M^m\mathbf{x}|_1 F^m\mathbf{x}$,

$$d(F^m\mathbf{x}, \mathbf{v}_M) = \log\left(\frac{\max_i (M^m\mathbf{x})(i)/\mathbf{v}_M(i)}{\min_i (M^m\mathbf{x})(i)/\mathbf{v}_M(i)}\right).$$

With

$$C_m(\mathbf{x}) := \left(\max_i \frac{(M^m\mathbf{x})(i)}{\mathbf{v}_M(i)} \min_i \frac{(M^m\mathbf{x})(i)}{\mathbf{v}_M(i)}\right)^{1/2}$$

we have $M^m\mathbf{x} = C_m(\mathbf{x})\mathbf{v}_M \times \mathrm{e}^{\pm d(F^m\mathbf{x},\mathbf{v}_M)/2}$. Multiplying from the left by $\mathbf{w}_M^\dagger$ these inequalities yields $C_m(\mathbf{x}) = \rho_M^m(\mathbf{w}_M^\dagger\mathbf{x})\,\mathrm{e}^{\pm d(F^m\mathbf{x},\mathbf{v}_M)/2}$. Taking into account corollary 6.1, the desired result follows. □

Appendix 2. Weak distance

In this subsection $X \subset A^{\mathbb{N}}$ is any subshift. We have the following very simple lemmas.

Lemma 6.1. *If $\nu, \mu \in \mathcal{M}(X)$ are such that $\mu[\mathbf{a}] = \nu[\mathbf{a}]\ \exp(\pm\epsilon)$ for each $\mathbf{a} \in \mathcal{L}_k(X)$, then $D(\mu, \nu) \leqslant (\exp(\epsilon) - 1) + 2^{-k}$.*

Proof. For $j \leqslant k$ we have

$$\begin{aligned}\sum_{\mathbf{a}\in\mathcal{L}_j(X)} |\mu[\mathbf{a}] - \nu[\mathbf{a}]| &\leqslant \sum_{\mathbf{a}\in\mathcal{L}_j(X)}\left(\sum_{\mathbf{b}\in\mathcal{L}_k(X):\ \mathbf{b}(0:j)=\mathbf{a}} |\mu[\mathbf{b}] - \nu[\mathbf{b}]|\right)\\ &\leqslant \sum_{\mathbf{a}\in\mathcal{L}_j(X)}\left(\sum_{\mathbf{b}\in\mathcal{L}_k(X):\ \mathbf{b}(0:j)=\mathbf{a}} \mu[\mathbf{b}](\mathrm{e}^\epsilon - 1)\right) = \mathrm{e}^\epsilon - 1.\end{aligned}$$

Hence $D(\mu, \nu) \leqslant (\mathrm{e}^\epsilon - 1)\sum_{j=0}^k 2^{-(j+1)} + \sum_{j=k+1}^\infty 2^{-(j+1)}(\sum_{\mathbf{a}\in\mathcal{L}_j(X)} |\mu[\mathbf{a}] - \nu[\mathbf{a}]|)$. The result follows by taking into account the fact that $\sum_{\mathbf{a}\in\mathcal{L}_j(X)} |\mu[\mathbf{a}] - \nu[\mathbf{a}]| \leqslant 2$ for all $j \in \mathbb{N}$. □

Lemma 6.2. *Let $\nu, \mu \in \mathcal{M}(X)$ be atomic with support $S_\nu := \mathrm{supp}(\nu) \subset S_\mu := \mathrm{supp}(\mu)$. Suppose that $\mu\{x\} \leqslant \nu\{x\} \leqslant \mu\{x\}\exp(\epsilon)$ for each $x \in S_\nu$. Then $D(\mu, \nu) \leqslant \exp(\epsilon) - \exp(-\epsilon)$.*

Proof. For each $k \in \mathbb{N}$, since $\{[\mathbf{a}] : \mathbf{a} \in \mathcal{L}_k(X)\}$ is a partition of X, we have

$$\begin{aligned}\sum_{\mathbf{a}\in\mathcal{L}_k(X)} |\mu[\mathbf{a}] - \nu[\mathbf{a}]| &= \sum_{\mathbf{a}\in\mathcal{L}_k(X)} (\nu(S_\nu \cap [\mathbf{a}]) - \mu(S_\nu \cap [\mathbf{a}])) + \sum_{\mathbf{a}\in A_m} \mu((S_\mu\backslash S_\nu) \cap [\mathbf{a}])\\ &\leqslant (\mathrm{e}^\epsilon - 1)\mu(S_\nu) + \mu(S_\mu\backslash S_\nu) \leqslant (\mathrm{e}^\epsilon - 1) + \mu(S_\mu\backslash S_\nu).\end{aligned}$$

Now, $1 = \nu(S_\nu) \leqslant \exp(\epsilon)\mu(S_\nu)$; hence $\mu(S_\mu\backslash S_\nu) \leqslant 1 - \exp(-\epsilon)$ and the result follows. □

Appendix 3. Entropy and relative entropy

Let ν be a shift-invariant probability measure on a specified subshift $Y \subset A^{\mathbb{N}}$. The measure-theoretic entropy of ν is

$$h(\nu) = -\lim_{n\to\infty} \frac{1}{n+1} \sum_{\mathbf{a}\in\mathcal{L}_n(Y)} \nu[\mathbf{a}] \log \nu[\mathbf{a}].$$

Since $\nu(A^{\mathbb{N}}\backslash Y) = 0$, we can replace $\mathcal{L}_n(Y)$ by A^{n+1} by using the usual convention '$0 \log 0 = 0$'.

We now turn to relative entropy. We refer the reader to [4] for details. Therein, only subshifts of finite type are considered, but the extension to more general subshifts is straightforward. Let μ_ψ be a Gibbs measure (with Hölder continuous potential ψ defined on $A^{\mathbb{N}}$) on a specified subshift $Y' \supset Y$. The relative entropy of ν with respect to μ_ψ is defined as

$$h(\nu|\mu_\psi) = \lim_{n\to\infty} \frac{1}{n+1} \sum_{\mathbf{a}\in\mathcal{L}_n(Y')} \nu[\mathbf{a}] \log \frac{\nu[\mathbf{a}]}{\mu_\psi[\mathbf{a}]}.$$

Note that the hypothesis $Y \subset Y'$ is crucial for making $\nu[\mathbf{a}] \log(\nu[\mathbf{a}]/\mu_\psi[\mathbf{a}])$ well-defined. One can prove that

$$h(\nu|\mu_\psi) = P(\psi, Y') - \int_Y \psi \, \mathrm{d}\nu - h(\nu). \tag{15}$$

We note that this result is true whenever μ_ψ satisfies the 'Gibbs inequality' (1), ψ not being necessarily Hölder continuous.

References

[1] Bowen R 1974/75 Some systems with unique equilibrium states *Math. Syst. Theory* **8** 193–202

[2] Bowen R 1975 *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms* (*Lecture Notes in Mathematics* vol 470) (Berlin: Springer)

[3] Bressaud X, Fernández R and Galves A 1999 Speed of $\bar{d}$-convergence for Markov approximations of chains with complete connections. A coupling approach *Stochastic Process. Appl.* **83** 127–38

[4] Chazottes J-R, Floriani E and Lima R 1998 Relative entropy and identification of Gibbs measures in dynamical systems *J. Stat. Phys.* **90** 697–725

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

[6] Denker M 1990 Some new examples of Gibbs measures *Monat. Math.* **109** 49–62

[7] Gurevich B 1980 Stationary random sequences of maximal entropy *Multicomponent Random Systems* (New York: Dekker) pp 327–80, *Adv. Probab. Related Topics* **6**

[8] Haydn N T A and Ruelle D 1992 Equivalence of Gibbs and equilibrium states for homeomorphisms satisfying expansiveness and specification *Commun. Math. Phys.* **148** 155–67

[9] Katok A and Hasselblatt B 1995 Introduction to the modern theory of dynamical systems *Encyclopedia of Mathematics and its Applications* vol 54 (Cambridge: Cambridge University Press)

[10] Keller G 1998 *Equilibrium States in Ergodic Theory* (*London Mathematical Society Student Texts* vol 42) (Cambridge: Cambridge University Press)

[11] Kitchens B 1998 *Symbolic Dynamics* (Berlin: Springer)

[12] Le Borgne S 1995 Un codage sofique des automorphismes hyperboliques du tore *Séminaires de Probabilités de Rennes, Publ. Inst. Rech. Math. Rennes (Univ. Rennes I, Rennes, 1995)* p 35

[13] Ledrappier F 1974 Principe variationnel et systèmes dynamiques symboliques *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **30** 185–202

[14] Schmeling J 1997 Symbolic dynamics for β-shifts and self-normal numbers *Ergod. Theory Dynam. Syst.* **17** 675–94

[15] Seneta E 1981 *Non-Negative Matrices and Markov Chains (Springer Series in Statistics)* (Berlin: Springer)

[16] Shields P 1996 *The Ergodic Theory of Discrete Sample Paths* (*Graduate Studies in Mathematics* vol 13) (Providence, RI: American Mathematical Society)

[17] Walters P 1982 *An Introduction to Ergodic Theory* (Berlin: Springer)