ON THE TOPOLOGICAL PRESSURE OF AXIAL PRODUCT ON TREES

JUNG-CHAO BAN AND YU-LIANG WU

Abstract. This article investigates the topological pressure of isotropic axial products of Markov subshift on the d-tree. We show that the quantity increases with dimension d, and demonstrate that, with the introduction of surface pressure, the two types of pressure admit the same asymptotic value. To this end, the pattern distribution vectors and the associated transition matrices are introduced herein to partially transplant the large deviation theory to tree-shifts, and so the increasing property is proved via an almost standard argument. An application of the main result to a wider class of shift spaces is also provided in this paper, and numerical experiments are included for the purpose of verification.

1. Introduction

The present paper is devoted to the study of the topological pressure of the axial product of a Markov subshift on the d-tree. This is mainly motivated by the recent works on the the limiting entropy of the axial product of $\mathbb{N}^d$ [13, 14], and the asymptotic pressure of the axial product on d-tree [17]. Before presenting the main results, the motivation behind the study is set out below.

Let $\mathcal{A}$ be a finite set with $|\mathcal{A}| = k$ and $X_1, \dots, X_d \subseteq \mathcal{A}^{\mathbb{N}}$ be one-sided subshifts. The associated *axial product of subshifts* $X_1, \dots, X_d$ *on* $\mathbb{N}^d$, written as $\otimes_{i=1}^d X_i = X_1 \otimes \cdots \otimes X_d \subset \mathcal{A}^{\mathbb{N}^d}$, is defined as

$$\otimes_{i=1}^d X_i = \{x \in \mathcal{A}^{\mathbb{N}^d} : \forall g \in \mathbb{N}^d \ \forall i \in \{1, \dots, d\}, x_{g+\mathbb{Z}_+ \mathsf{e}_i} \in X_i\}, \tag{1}$$

where $x_{g+\mathbb{Z}_+ \mathsf{e}_i} \in \mathcal{A}^{\mathbb{N}}$ is the sequence obtained by shifting x by g and $\{\mathsf{e}_1, \dots, \mathsf{e}_d\}$ denotes the standard basis of $\mathbb{N}^d$. Suppose $\mathcal{T}^{(d)}$ is the conventional d-tree, that is, a free semigroup generated by the generators $\{f_1, \dots, f_d\}$ with an element of identity ϵ, which is the *root* of the tree. The *axial product of subshifts* $X_1, \dots, X_d \subseteq \mathcal{A}^{\mathbb{N}}$ *on* $\mathcal{T}^{(d)}$, denoted as $\times_{i=1}^d X_i = X_1 \times \cdots \times X_d$, is defined similarly. That is,

$$\times_{i=1}^d X_i = \{x \in \mathcal{A}^{\mathcal{T}^{(d)}} : \forall g \in \mathcal{T}^{(d)} \ \forall i \in \{1, \dots, d\}, (x_{gf_i^n})_{n \in \mathbb{Z}_+} \in X_i\}. \tag{2}$$

An axial product $\otimes_{i=1}^d X_i$ (or $\times_{i=1}^d X_i$) is called *isotropic* if $X_i = X_j$ for all $1 \leq i \neq j \leq d$, and is called *anisotropic* if it is not isotropic. Isotropic axial products of shifts on $\mathbb{N}^d$ are first introduced in [13], and many important physical systems, e.g. the hard square model in $\mathbb{N}^2$ or $\mathbb{Z}^2$, fall in this class. In particular, when it comes to tree-shifts, a notable subclass of isotropic axial products is the family of *hom tree-shifts*[1], whose X_i's are unanimously a one-sided Markov subshift. This family is widely considered in the literature [1, 2, 3, 8, 15, 16, 17] and has recently

Date: March 2022.

[1]Such shifts are also called the associated tree shifts in [15, 16].

attracted considerable attention, since $\mathcal{T}^{(d)}$ is not an amenable semigroup, and the shifts defined on it exhibit rich and diverse phenomena in terms of topological (cf. [2]) and statistical prospects (cf. [4, 5]).

For a dynamical system, the most significant quantity is the (topological) entropy, which measures the asymptotic growth in the number of admissible patterns on finite sets. More precisely, supposing $F \subseteq \mathbb{N}^d$ is a finite set, we denote by $\pi(F, X) : \mathcal{A}^{\mathbb{N}^d} \to \mathcal{A}^F$ the *canonical projection* of $X \subseteq \mathcal{A}^{\mathbb{N}^d}$ into $\mathcal{A}^F$, i.e., $\pi(F, X) = \{(x_g)_{g\in F} \in \mathcal{A}^F : x \in X\}$. If we denote $F_n := [0, n]^d$ and suppose $X \subset \mathcal{A}^{\mathbb{N}^d}$, the *entropy* of X is defined as

$$h(X) = \lim_{n\to\infty} \frac{\log |\pi(F_n, X)|}{|F_n|}, \tag{3}$$

where $|F|$ denotes the number of elements in F. The limit (3) exists since $\mathbb{N}^d$ is an amenable semigroup (cf. [10]), and F_n is a *Følner sequence*, i.e., $\lim_{n\to\infty} \frac{|gF_n \Delta F_n|}{|F_n|} = 0$. In fact, it is known that if $\{F'_n\}_{n=1}^{\infty}$ is any Følner sequence, the limit

$$h(X) = \lim_{n\to\infty} \frac{\log |\pi(F'_n, X)|}{|F'_n|}$$

exists and equals (3). For $T^{(d)} \subseteq \mathcal{A}^{\mathcal{T}^{(d)}}$, let

$$\Xi_n^{(d)} = \{f_1, f_2, \cdots, f_d\}^n \quad \text{and} \quad \Delta_n^{(d)} = \cup_{i=0}^m \Xi_i^{(d)}.$$

Then, the *topological entropy* of $T^{(d)} \subseteq \mathcal{A}^{\mathcal{T}^{(d)}}$ is defined similarly:

$$h(T^{(d)}) = \lim_{n\to\infty} \frac{\log |\pi(\Delta_n^{(d)}, T^{(d)})|}{\left|\Delta_n^{(d)}\right|}. \tag{4}$$

The existence of the limit (4) is proved in [15], and we emphasize that since the subadditive property does not hold for the shifts on $\mathcal{T}^{(d)}$, the proof of the existence of (4) is quite different from that of the shifts on $\mathbb{N}^d$. We refer the reader to [7] for the existence of the limit (4) for shifts defined in a large class of trees. It appears that the structures of $\{h(T) : T \text{ is a Markov tree-shift}\}$ and $\{h(X) : X \text{ is an } \mathbb{N}^d \text{ Markov shift}\}$ are quite different (cf. [4])

Let $X \subseteq \mathcal{A}^{\mathbb{Z}}$. Louidor, Marcus and Pavlov [13] introduced the concept of *limiting entropy*. Namely, $h^{(\infty)}(X) := \lim_{d\to\infty} h(X^{\otimes d})$, where $X^{\otimes d} = \overbrace{X \otimes \cdots \otimes X}^{d\text{-times}}$, and the limit exists since $h(X^{\otimes d})$ is non-increasing in d. Later, Meyerovitch and Pavlov [14] revealed that $h^{(\infty)}(X)$ coincides with the independence entropy (see the references for the definition of independence entropy). For $2 \leq d \in \mathbb{N}$, Petersen and Salama initiated the study of the limiting entropy of a hom tree change; that is, $h^{(\infty)}(X) := \lim_{d\to\infty} h(X^{\times d})$, whenever the limit exists. In [16], the authors consider X_G as a golden-mean[2] subshift and use the method of the *site strip approximation* to prove that $h(X_G^{\times d})$ is strictly increasing in d (Theorem 3.7 [16]). Thus, the limit $h^{(\infty)}(X_G)$ exists.

Recently, Petersen and Salama [17] generalized the notion of limiting entropy to *pressure*[3] (formally defined in Section 2.1) and a broad class of trees, namely, the

[2]A shift space with two symbols $0, 1$ that preclude the existence of neighboring 1's.

[3]They refer to it as *asymptotic pressure*.

generalized Fibonacci tree[4]. Precisely, let E be a $k \times k$ nonnegative matrix indexed by $\mathcal{A}$ and X_E be the associated one-sided subshift of finite type (SFT) defined as

$$X_E = \{x \in \mathcal{A}^{\mathbb{Z}_+} : E_{x_{n+1},x_n} > 0 \text{ for every } n \in \mathbb{Z}_+\}.$$

Suppose $X_E^{\times d}$ is the d-axial product of X_E on the generalized Fibonacci tree. The authors show in [17, Theorem 4.2] that if

$$\text{(A)} \qquad \sum_{a \in \mathcal{A}} E_{a,b} > 0 \quad \text{and} \quad \sum_{b \in \mathcal{A}} E_{a,b} > 0$$

the limit $\lim_{d\to\infty} \mathbf{P}(X_E^{\times d}, E)$ (see (8)) exists and coincides with $\log r_E$, where

$$r_E = \max\left\{\sum_{a \in \mathcal{A}} E_{a,b} : b \in \mathcal{A}\right\}. \tag{5}$$

The aim of this article is to explicate the finer structure of the function $d \mapsto \mathbf{P}(X_E^{\times d}, E)$. The main findings of this work are as follows. First, we introduce the notion of *pattern distribution* on a tree-shift, and propose a finite-dimensional optimization problem (Problem 1) for a hom tree-shift. Such a problem can be deemed as a special form of the large deviation problem, and it can be solved after a series of transformations. With two intermediate quantities $P^{(k)}(s, E)$ and $P^{(\infty)}(s, E)$ (maximizers of (Problem 4) and (Problem 5), respectively) defined in the process of the optimization problem, we are able to characterize the function $d \mapsto \mathbf{P}(X_E^{\times d}, E)$ and prove the main results of this paper, as stated below.

Theorem 1. *Suppose the adjacency matrix E satisfies* (A)*. Then, the following assertions hold true.*

(a) The function $P^{(\infty)}(s, E)$ is continuous in s on $(0, 1)$, and $P^{(k)}(s, E)$ converges uniformly to on all compact subintervals.

(b) $P^{(\infty)}(1/d, E) = \mathbf{P}(X_E^{\times d}, E)$.

(c) $P^{(\infty)}(s, E)$ is decreasing in s.

(d) We have

$$P^{(\infty)}(0+, E) = \lim_{d\to\infty} P^{(\infty)}(1/d, E) = \lim_{d\to\infty} \mathbf{P}_*(X_E^{\times d}, E) = \log r_E \tag{6}$$

and $P^{(\infty)}(1-, E) = \log \rho(E)$, where $\rho(E)$ is the spectral radius of E and r_E is defined as (7)*.*

We call $\mathbf{P}_*$ in (6) the surface pressure (defined in Section 2.1), of which the name traces back to the work of Berger and Ye [9], who were the first to define an analogous quantity called surface entropy for their study of entropy of Markov random fields on trees. As an application of the theorem, the following corollary partially generalizes the result discussed in [18], in which the discussion concentrates on the case of the golden-mean shift.

Corollary 2. *Let $\mathcal{G}^{(d)}$ be a free group with d generators, $E = E^T$, and $Y_E^{(d)}$ be a Markov shift space over $\mathcal{G}^{(d)}$. Then, the entropy $h(Y_E^{(d)})$ is increasing in d.*

This is a consequence of the fact that $h(Y_E^{(d)}) = h(X_E^{\times(2d-1)})$, as is proved in [7, Proposition 4.5].

A few remarks could be made at this point.

[4]We refer the reader to [17] for the formal definition of the generalized Fibonacci trees.

(1) The assumption (A) is reasonable in that (a) by dropping all such $b \in \mathcal{A}$ that $\sum_{a\in\mathcal{A}} E_{a,b} = 0$, we obtain a submatrix E' of E with $X_{E'} = X_E$, and thus $\mathbf{P}(X_{E'}^{\times d}, E') = \mathbf{P}(X_E^{\times d}, E)$; (b) by dropping such $b \in \sum_{b\in\mathcal{A}} E_{a,b} = 0$, there exist a subset $\mathcal{A}'$ of $\mathcal{A}$ and its associated submatrix E' of E such that $\mathbf{P}(X_{E'}^{\times d}, E') = \mathbf{P}(X_E^{\times d}, E)$ for all $d \geq 2$. Indeed, the equality follows from an essentially identical argument to that for a one-sided Markov shift.

(2) Combining Theorem 1 (b) and (c) yields the increasing property of the function

$$d \mapsto P^{(\infty)}(1/d, E) = \mathbf{P}(X_E^{\times d}, E). \tag{7}$$

We stress that if E is a binary matrix, then $\mathbf{P}(X_E^{\times d}, E) = h(X_E^{\times d})$. Thus, the increasing property of (7) provides another point of view for the increasing property of $d \mapsto h(X_E^{\times d})$ as we mentioned in the last paragraph and extends the property to any one-sided Markov subshift.

(3) The quantity $\mathbf{P}_*(X_E^{\times d}, E)$ in Theorem 1 (d) is defined in [6] as well, which is an analogous concept to *maximal pattern entropy* of a topological dynamical system (cf. [6, 11, 12, 19]). Note that Theorem 1 (d) combined with Theorem 1 (b) reveals that

$$\lim_{d\to\infty} \mathbf{P}(X_E^{\times d}, E) = \lim_{d\to\infty} P^{(\infty)}(1/d, E) = \lim_{d\to\infty} \mathbf{P}_*(X_E^{\times d}, E) = \log r_E.$$

This also provides an alternative perspective for the result of Petersen and Salama ([17, Theorem 4.2]) on the d-axial product on the d-tree. Furthermore, (6) exposes the relationship between the topological pressure and the maximal pattern entropy and gives more information on the limiting behavior of $\lim_{d\to\infty} \mathbf{P}(X_E^{\times d}, E)$. Finally, $P^{(\infty)}(1-, E) = \log \rho(E)$ corresponds to the case of a degenerate tree-shift, or, in other words, the one-sided SFT X_E.

The contents of this paper are arranged in the following manner. Section 2.1 covers our conventions for notation as well as general backgrounds of tree-shifts, and Section 2.2 outlines the setup of pattern distribution for Theorem 1. Section 3 further delves into the optimization process of the topological pressure, which includes the statement of a technical theorem regarding the function $P^{(k)}(s, E)$ (Theorem 13). The proofs of Theorem 13 and 1 are left until Section 4 and Section 5, respectively. Finally, Section 6 contains two examples whose $P^{(\infty)}(s, E)$ are plotted for verification purposes (Figure 1 and 2).

2. Preliminaries

2.1. **Markov tree-shifts, topological pressure and surface pressure.** Throughout the paper, we let $E \in \mathbb{R}_{\geq 0}^{\mathcal{A}\times\mathcal{A}}$ satisfy assumption (A) and let $w \in \mathbb{R}_{>0}^{\mathcal{A}}$. We will focus on the class of isotropic d-axial products of a Markov shift $X_E^{\times d} \subseteq \mathcal{A}^{\mathcal{T}^{(d)}}$, which can be written as

$$X_E^{\times d} := \{t \in \mathcal{A}^{\mathcal{T}^{(d)}} : \forall g \in \mathcal{T}^{(d)}\ \forall i \in \{1, \ldots, d\},\ E_{x_{gf_i}, x_g} > 0\},$$

as mentioned in the previous section. We further write

$$\Delta_{n:m}^{(d)} = \cup_{i=n}^{m} \Xi_i^{(d)}$$

and express the set of *blocks*, in terms of the projection map, as

$$B_{n:m}(X_E^{\times d}) = \pi(\Delta_{n:m}^{(d)}, X_E^{\times d}) := \{(t_g)_{g\in\Delta_{n:m}^{(d)}} : t \in X_E^{\times d}\}.$$

The *weight* of $u \in B_{n:m}(X_E^{\times d})$, corresponding to the interaction matrix E, on $\Delta_{n:m}^{(d)}$ is defined as

$$Z_{n:m}[u,E] = w(u|_{\Xi_n^{(d)}}) \prod_{g\in\Delta_{n:m-1}^{(d)}} \prod_{i=1}^{d} E_{u_{gf_i},u_g},$$

where we slightly abuse the notation w to define

$$w(u|_{\Xi_n^{(d)}}) = \prod_{g\in\Xi_n^{(d)}} w_{u_g}.$$

We then call $|Z_{n:m}(X_E^{\times d}, E)| = \sum_{u\in B_{n:m}(X_E^{\times d})} Z_{n:m}[u,E]$ the *partition function* on $\Delta_{n:m}^{(d)}$. By denoting $|Z_n(X_E^{\times d}, E)| = |Z_{0:n}(X_E^{\times d}, E)|$, the *topological pressure* and *surface pressure* are respectively defined to be

$$\mathbf{P}(X_E^{\times d}, E) = \lim_{n\to\infty} \frac{\log |Z_n(X_E^{\times d}, E)|}{|\Delta_n|}, \tag{8}$$

$$\mathbf{P}_*(X_E^{\times d}, E) = \lim_{n\to\infty} \frac{\log |Z_{n:n}(X_E^{\times d}, E)|}{|\Xi_n|}, \tag{9}$$

where the limit of topological pressure is proved to exist in [15, Theorem 2.1] and [17, Theorem 2.1] while the existence of the other limit follows from a similar argument, and a proof for the case where E is a binary matrix and $w = \mathbb{1}$ (i.e., $|Z_{n:n}(X_E^{\times d}, E)| = |B_{n:n}(X_E^{\times d})|$) can be found in [6]. It is noteworthy that we suppress the dependency of $w \in \mathbb{R}$ in all of the notations above for the sake of conciseness, since the limits (8) and (9) do not depend on the vector. Also, this latter quantity (9) is shown in [6, Corollary 3.3] to coincide with the *maximal pattern entropy* of the hom tree-shifts, namely, the supremum of sequence entropy of tree-shifts over all possible sequences. For conciseness, we drop the notations $X_E^{\times d}$ and E should the meaning be clear from the context.

Our conventions for the notation of matrices and vectors are as follows. Let $\Gamma_{\mathcal{A}}$ be the set of all probability vectors on $\mathcal{A}$ and $\Upsilon_{\mathcal{A}}$ be the set of stochastic matrices acting on $\Gamma_{\mathcal{A}}$. Then, vectors in $\Gamma_{\mathcal{A}}$ are typed in lowercase of sans serif font, such as p, q while matrices in $\Upsilon_{\mathcal{A}}$ are typed in uppercase, such as P, Q. The sum of each column of a stochastic matrix P is 1, i.e., $\sum_{a\in\mathcal{A}} \mathsf{P}_{a,b} = 1$ for all $b \in \mathcal{A}$. The transpose of a matrix M (respectively, a vector v) is denoted by M^T (respectively, v^T). The product of two square matrices M and N (respectively, a matrix M and a vector v) of the same dimension is denoted by MN (respectively, Mv), while the Hadamard product of the two matrices is written as $M \circ N$. The notation $\circ$ should not be mistaken for the symbol of composition of functions, given the context. For conciseness, we denote $\prod_{i=1}^k M_i = M_1 M_2 \cdots M_k$ for any k square matrices of the same dimension. Finally, the standard vector of symbol a is denoted by e_a and the all-one vector and the all-one matrix is denoted by $\mathbb{1}$.

In this article, the space $\Gamma_{\mathcal{A}}$ and $\Upsilon_{\mathcal{A}}$ are implicitly endowed with the variational distance defined as follows.

Definition 3. *Let* $\mathsf{p},\mathsf{q}\in\Gamma_{\mathcal{A}}$ *and* $\mathsf{P},\mathsf{Q}\in\Upsilon_{\mathcal{A}}$. *The variational distance of the two vectors is defined to be*

$$\mathtt{d}_v(\mathsf{p},\mathsf{q})=\max_{S\subseteq\mathcal{A}}|\sum_{a\in S}\mathsf{p}_a-\mathsf{q}_a|=\frac{1}{2}\sum_{a\in\mathcal{A}}|\mathsf{p}_a-\mathsf{q}_a|$$

and the variational distance of the matrices is defined as

$$\mathtt{d}_V(\mathsf{P},\mathsf{Q})=\max_{b\in\mathcal{A}}\mathtt{d}_v(\mathsf{P}_{a,b},\mathsf{Q}_{a,b}).$$

In addition, for any pairs (p,P) *and* (q,Q)*, their variational distance is defined to be*

$$\mathtt{d}_{v,V}((\mathsf{p},\mathsf{P}),(\mathsf{q},\mathsf{Q}))=\max\{\mathtt{d}_v(\mathsf{p},\mathsf{q}),\mathtt{d}_V(\mathsf{P},\mathsf{Q})\}.$$

We note that $\mathtt{d}_v$ and $\mathtt{d}_V$ can be carried to the product spaces $\Gamma_{\mathcal{A}}^n$ and $\Upsilon_{\mathcal{A}}^n$ so they are assumed in what follows to be endowed with corresponding product metrics. This idea also extends to a metric $\mathtt{d}_{v,V}$ on the space $\Gamma_{\mathcal{A}}^n\times\Upsilon_{\mathcal{A}}^n$.

Proposition 4. *Suppose* $\mathsf{p},\mathsf{q}\in\Gamma_{\mathcal{A}}$ *and* $\mathsf{P},\mathsf{Q},\mathsf{R}\in\Upsilon_{\mathcal{A}}$. *Then,*

- $\mathtt{d}_v(\mathsf{Pp},\mathsf{Qp})\le\mathtt{d}_V(\mathsf{P},\mathsf{Q})$
- $\mathtt{d}_v(\mathsf{Pp},\mathsf{Pq})\le\mathtt{d}_v(\mathsf{p},\mathsf{q})$
- $\mathtt{d}_V(\mathsf{PR},\mathsf{QR}),\mathtt{d}_V(\mathsf{RP},\mathsf{RQ})\le\mathtt{d}_V(\mathsf{P},\mathsf{Q})$

Proof. It is clear that

$$\begin{aligned}\mathtt{d}_v(\mathsf{Pp},\mathsf{Qp})&=\frac{1}{2}\sum_{a\in\mathcal{A}}|\sum_{b\in\mathcal{A}}\mathsf{P}_{a,b}\mathsf{p}_b-\mathsf{Q}_{a,b}\mathsf{p}_b|\\&\le\frac{1}{2}\sum_{b\in\mathcal{A}}\mathsf{p}_b\sum_{a\in\mathcal{A}}|\mathsf{P}_{a,b}-\mathsf{Q}_{a,b}|\\&\le\mathtt{d}_V(\mathsf{P},\mathsf{Q}),\end{aligned}$$

that

$$\begin{aligned}\mathtt{d}_v(\mathsf{Pp},\mathsf{Pq})&=\frac{1}{2}\sum_{a\in\mathcal{A}}|\sum_{b\in\mathcal{A}}\mathsf{P}_{a,b}\mathsf{p}_b-\mathsf{P}_{a,b}\mathsf{q}_b|\\&\le\frac{1}{2}\sum_{b\in\mathcal{A}}|\mathsf{p}_b-\mathsf{q}_b|\sum_{a\in\mathcal{A}}\mathsf{P}_{a,b}\\&=\mathtt{d}_v(\mathsf{p},\mathsf{q})\end{aligned}$$

and that the last inequality is a consequence of the former. ☐

As part of our conventions, every sequence $\mathsf{q},\mathsf{p},\mathsf{Q},\mathsf{P}$ has initial index 0 unless mentioned otherwise. Furthermore, for $\mathsf{q}=(\mathsf{q}^{(0)},\mathsf{q}^{(1)},\cdots,\mathsf{q}^{(k)})\in\Gamma_{\mathcal{A}}^{k+1}$ and $\mathsf{Q}=(\mathsf{Q}^{(0)},\mathsf{Q}^{(1)},\cdots,\mathsf{Q}^{(k)})\in\Upsilon_{\mathcal{A}}^{k+1}$, we write

$$\overleftarrow{\mathsf{q}}=(\mathsf{q}^{(k)},\mathsf{q}^{(k-1)},\cdots,\mathsf{q}^{(0)})\text{ and }\overleftarrow{\mathsf{Q}}=(\mathsf{Q}^{(k)},\mathsf{Q}^{(k-1)},\cdots,\mathsf{Q}^{(0)}).$$

In the setting of hom tree-shift, we further define a typical pairs of distribution (q,Q) as follows.

Definition 5. *A pair* $(\mathsf{q},\mathsf{Q})\in\Gamma_{\mathcal{A}}\times\Upsilon_{\mathcal{A}}$ *is said to be canonically admissible by a non-negative matrix* M *of the same dimension if the following are satisfied.*

- $\mathsf{Q}_{a,b}>0$ *if and only if* $M_{a,b}>0$.
- $\mathsf{Q}_{a,b}=M_{a,b}\cdot(\sum_{c\in\mathcal{A}}M_{c,b})^{-1}$ *if* $\mathsf{q}_b=0$.

2.2. **Pattern distribution.** This subsection details the exposition of pattern distribution on tree-shifts. For a better conception of the topological pressure, here we introduce distribution vectors and the associated transition matrices to illustrate the structures of a hom tree-shift so as to prove $d \mapsto \mathbf{P}(X_E^{\times d}, E)$ is decreasing through reformulation of topological pressure as an optimization problem. In fact, the role played by the distribution vectors and transition matrices is similar to that played by the identical and independent variables in the large deviation theory to some degree. It is owing to this similarity that the techniques in the theory are able to be applied to the process of reformulation as well as to solving the optimization problems in this work.

For $t \in X_E^{\times d}$, the *distribution vector of* t *on level* n is defined as

$$\tau_n(t) = \left(\frac{\sum_{g \in \Xi_n^{(d)}} \chi_a(t_g)}{|\Xi_n^{(d)}|} \right)_{a \in \mathcal{A}} \in \Gamma_{\mathcal{A}} \tag{10}$$

and $D_n(X_E^{\times d})$ denotes the set of all distribution vectors on level n, i.e.,

$$D_n(X_E^{\times d}) = \left\{ \tau_n(t) : t \in X_E^{\times d} \right\}. \tag{11}$$

In addition, by writing $\sigma(g)$ the set of all children of g, the *transition matrix of* t *from level* n *to* $n+1$ is defined to be

$$\eta_n(t)_{a,b} = \begin{cases} \left(\frac{\sum_{g \in \Xi_n^{(d)}, t_g = b} \sum_{h \in \sigma(g)} \chi_a(t_h)}{\sum_{g \in \Xi_n^{(d)}, t_g = b} |\sigma(g)|} \right) & \text{if } \sum_{g \in \Xi_n^{(d)}, t_g = b} |\sigma(g)| > 0 \\ \frac{E_{a,b}}{\sum_{a \in \mathcal{A}} E_{c,b}} & \text{if otherwise,} \end{cases} \tag{12}$$

and $S_n(X_E^{\times d})$ stands for the set of all transition matrices:

$$S_n(X_E^{\times d}) = \left\{ \eta_n(t) : t \in X_E^{\times d} \right\}. \tag{13}$$

For the sake of convenience, we further generalize the notations in the same way that

$$\tau_{n:m}(t) = (\tau_n(t), \cdots, \tau_m(t)) \text{ and } D_{n:m}(X_E^{\times d}) = \left\{ \tau_{n:m}(t) : t \in X_E^{\times d} \right\},$$

$$\eta_{n:m}(t) = (\eta_n(t), \cdots, \eta_{m-1}(t)) \text{ and } S_{n:m}(X_E^{\times d}) = \left\{ \eta_{n:m}(t) : t \in X_E^{\times d} \right\},$$

and denote by $W_{n:m}(X_E^{\times d})$ the set of all compatible pairs $(\mathsf{q}, \mathsf{Q}) \in D_{n:m}(X_E^{\times d}) \times S_{n:m}(X_E^{\times d})$, i.e.,

$$\begin{aligned} W_{n:m}(X_E^{\times d}) &= \{ (\tau_{n:m}(t), \eta_{n:m}(t)) : t \in X_E^{\times d} \} \\ &= \{ (\mathsf{q}, \mathsf{Q}) \in D_{n:m}(X_E^{\times d}) \times S_{n:m}(X_E^{\times d}) : \mathsf{q}^{(i+1)} = \mathsf{Q}^{(i)} \mathsf{q}^{(i)} \}. \end{aligned}$$

Finally, the sets of blocks with given distribution vectors and transition matrices are defined as

$$B_{n:m}(X_E^{\times d}; \mathsf{q}) = \{ (t_g)_{g \in \Delta_{n:m}^{(d)}} : t \in X_E^{\times d}, \tau_{n:m}(t) = \mathsf{q} \},$$

$$B_{n:m}(X_E^{\times d}; \mathsf{q}, \mathsf{Q}) = \{ (t_g)_{g \in \Delta_{n:m}^{(d)}} : t \in X_E^{\times d}, \tau_{n:m}(t) = \mathsf{q}, \eta_{n:m}(t) = \mathsf{Q} \}.$$

For conciseness, we suppress the notation $X_E^{\times d}$ and d whenever no ambiguity should occur. It turns out that, by a standard argument in large deviation theory, the size of these sets merely has a sub-exponential growth rate with respect to $(m-n)$, as is shown in the following proposition.

Proposition 6. *For any $n \leq m$,*

$$1 \leq |D_{n:m}| \leq \prod_{i=n}^{m} (|\Xi_i|+1)^{|\mathcal{A}|} \leq \left(\frac{|\Delta_{n:m}|}{m-n+1} + 1 \right)^{(m-n+1)\cdot|\mathcal{A}|},$$

$$1 \leq |S_{n:m}| \leq \prod_{i=n}^{m} (|\Xi_i|+1)^{|\mathcal{A}|(|\mathcal{A}|+1)} \leq \left(\frac{|\Delta_{n:m}|}{m-n+1} + 1 \right)^{2(m-n+1)\cdot|\mathcal{A}|^2}.$$

In addition, $\lim_{|\Delta_{n:m}|\to\infty}(m-n)/|\Delta_{n:m}| = 0$.

Proof. The inequalities follow from a simple fact that given any $k \in \mathbb{N}$, the size of the set $I_{k,\ell} := \{(\frac{a_i}{k})_{1\leq i\leq \ell} : a_i \in \mathbb{Z}_+, \sum_{i=1}^{\ell} a_i = k\}$ is no more than $(k+1)^{\ell}$.

For the set $D_{n:m}$, the first inequality is trivial, and the second follows from that $D_i \subset I_{|\Xi_i|,|\mathcal{A}|}$ and that $|D_{n:m}| \leq \prod_{i=n}^{m} |D_i|$. The third inequality is a consequence of arithmetic and geometric means.

As for the set $S_{n:m}$, the first inequality is again trivial. On the other hand, the set of vectors $\Lambda_i = \{(\sum_{g\in\Xi_i, t_g=b} |\sigma(g)|)_{b\in\mathcal{A}} : t \in X_E^{\times d}\}$ is contained in the set $|\Xi_{i+1}| \cdot I_{|\Xi_{i+1}|,|\mathcal{A}|}$ and

$$\begin{aligned} |S_i| &\leq \sum_{(\sum_{g\in\Xi_i,t_g=b}|\sigma(g)|)_{b\in\mathcal{A}}\in\Lambda_i} \prod_{b\in\mathcal{A}} \max\left\{ |I_{\sum_{g\in\Xi_i,t_g=b}|\sigma(g)|,|\mathcal{A}|}|, 1 \right\} \\ &\leq \sum_{(\sum_{g\in\Xi_i,t_g=b}|\sigma(g)|)_{b\in\mathcal{A}}\in\Lambda_i} (|\Xi_{i+1}|+1)^{|\mathcal{A}|^2} \\ &\leq (|\Xi_{i+1}|+1)^{|\mathcal{A}|(|\mathcal{A}|+1)}. \end{aligned}$$

The rest of the argument is similar to the one stated above.

Finally, the remaining limit of the ratio holds true as a consequence that $|\Xi_n^{(d)}|$ grows exponentially. □

3. Combinatorial optimization

This section deals with the optimization problem of partition functions. Noting that $B_{n:m}$ is a disjoint union of $B_{n:m}(\mathsf{q})$ ($\mathsf{q} \in D_{n:m}$), we may estimate the partition function by

$$\max_{\mathsf{q}\in D_{n:m}} |Z_{n:m}(\mathsf{q})| \leq |Z_{n:m}| \leq |D_{n:m}| \cdot \max_{\mathsf{q}\in D_{n:m}} |Z_{n:m}(\mathsf{q})|. \tag{14}$$

Therefore, for any unbounded increasing sequence $(\Delta_{n_i:m_i})_{i\in\mathbb{N}}$, i.e., a sequence satisfying that $|\Delta_{n_i:m_i}| \leq |\Delta_{n_{i+1}:m_{i+1}}|$ and that $\lim_{i\to\infty} |\Delta_{n_i:m_i}| = \infty$, we have

$$\liminf_{i\to\infty} \frac{\log|Z_{n_i:m_i}|}{|\Delta_{n_i:m_i}|} = \liminf_{i\to\infty} \max_{\mathsf{q}\in D_{n_i:m_i}} \frac{\log|Z_{n_i:m_i}(\mathsf{q})|}{|\Delta_{n_i:m_i}|} \tag{15}$$

$$\limsup_{i\to\infty} \frac{\log|Z_{n_i:m_i}|}{|\Delta_{n_i:m_i}|} = \limsup_{i\to\infty} \max_{\mathsf{q}\in D_{n_i:m_i}} \frac{\log|Z_{n_i:m_i}(\mathsf{q})|}{|\Delta_{n_i:m_i}|} \tag{16}$$

due to Proposition 6. These two quantities are related to $\mathbf{P}$ and $\mathbf{P}_*$ in the way that

$$\mathbf{P} \leq \liminf_{i\to\infty} \frac{\log|Z_{n_i:m_i}|}{|\Delta_{n_i:m_i}|} \leq \limsup_{i\to\infty} \frac{\log|Z_{n_i:m_i}|}{|\Delta_{n_i:m_i}|} \leq \mathbf{P}_*, \tag{17}$$

which can be verified by a similar argument to [15, Theorem 2.1]. In particular, if $\lim_{i\to\infty} m_i - n_i = \infty$,

$$\lim_{i\to\infty} \frac{\log|Z_{n_i:m_i}|}{|\Delta_{n_i:m_i}|} = \lim_{i\to\infty} \max_{\mathsf{q}\in D_{i:i}} \frac{\log|Z_{n_i:m_i}(\mathsf{q})|}{|\Delta_{n_i:m_i}|} = \mathbf{P}, \tag{18}$$

which again follows from a similar argument as [15, Theorem 2.1]. We note that the above discussions (in particular, equations (14)-(18)) remain valid if we replace $B_{n:m}(\mathsf{q})$ by $B_{n:m}(\mathsf{q},\mathsf{Q})$ and $D_{n:m}$ by $W_{n:m}$. Under this latter setting, (18) turns out to be useful for our expositions, which leads to the investigation of the following finite-dimensional optimization problem:

$$\text{(Problem 1)} \qquad \begin{cases} \text{maximize} & \frac{\log|Z_{n:n+k}(\mathsf{q},\mathsf{Q})|}{|\Delta_{n:n+k}|} \\ \text{subject to} & (\mathsf{q},\mathsf{Q}) \in W_{n:n+k} \end{cases}.$$

In view of the difficulty of solving a combinatorial optimization problem, we shall take the following steps to transform the problem into a regular optimization problem without changing the limiting behavior of the maximum:

(1) Rephrase the objective function.
(2) Extend the feasible domain to a convex set of Euclidean space.

For the first step, a heuristic treatment for the aforementioned optimal solutions is to apply Stirling's approximation to obtain an explicit expression of the objective function of (Problem 1). Before we demonstrate this, we first give the following definition. Let $\mathsf{P} \in \Upsilon_{\mathcal{A}}$ and M be a non-negative matrix of the same dimension and $\mathsf{P}_{a,b} = 0$ if $M_{a,b} = 0$. We define the vector

$$\Phi(\mathsf{Q}|M)_b := \sum_{a\in\mathcal{A}} -(\mathsf{Q}_{a,b}) \log(\frac{\mathsf{Q}_{a,b}}{M_{a,b}}),$$

where $0\log\frac{0}{0}$ is interpreted as 0. This is essentially the relative entropy of Q from M in each column, except that M is not necessarily a stochastic matrix and so $\Phi(\mathsf{Q}|M)$ need not be negative.

Proposition 7. *Let $(\Delta_{n_i:m_i})_{i\in\mathbb{N}}$ be an unbounded increasing sequence, and $(\mathsf{q}(i),\mathsf{Q}(i)) \in D_{n_i:m_i} \times S_{n_i:m_i}$ be any feasible sequence. Then, $\frac{\log|Z_{n_i:m_i}(\mathsf{q},\mathsf{Q})|}{|\Delta_{n_i:m_i}|}$ shares the common upper limit and lower limit with the sequence*

$$\frac{\log|Z_{n_i:n_i}(\mathsf{q}^{(0)}(i))|}{|\Delta_{n_i:m_i}|} + \sum_{j=0}^{m_i-n_i-1} \frac{|\Xi_{n_i+j+1}|}{|\Delta_{n_i:m_i}|} \Phi(\mathsf{Q}^{(j)}(i)|E)^T \mathsf{q}^{(j)}(i).$$

Proof. As a corollary of Stirling's approximation, we note that for any positive integers n,

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}.$$

A combinatorial argument shows that

$$\begin{aligned} |Z_{n_i:m_i}(\mathsf{q}(i),\mathsf{Q}(i))| =& |Z_{n_i:n_i}(\mathsf{q}^{(0)}(i))| \prod_{j=0}^{m_i-n_i-1} \prod_{b\in\mathcal{A}} \Bigg[\binom{|\Xi_{n_i+j+1}|\cdot \mathsf{q}_b^{(j)}(i)}{|\Xi_{n_i+j+1}|\cdot \mathsf{q}_b^{(j)}(i)\cdot \mathsf{Q}_{a,b}^{(j)}(i)} \\ & \cdot \prod_{a\in\mathcal{A}} E_{a,b}^{|\Xi_{n_i+j+1}|\cdot \mathsf{q}_b^{(j)}(i)\cdot \mathsf{Q}_{a,b}^{(j)}(i)} \Bigg], \end{aligned}$$

where $\binom{a}{b_1, b_2, \cdots, b_k} = \frac{a!}{b_1! b_2! \cdots b_k!}$ is the multinomial coefficient. Therefore,

$$\begin{aligned}
&\frac{\log |Z_{n_i:m_i}(\mathsf{q}(i), \mathsf{Q}(i))|}{|\Delta_{n_i:m_i}|} \\
\leq &\frac{\log |Z_{n_i}(\mathsf{q}^{(0)}(i))|}{|\Delta_{n_i:m_i}|} + \sum_{j=0}^{m_i-n_i-1} \frac{|\Xi_{n_i+j+1}|}{|\Delta_{n_i:m_i}|} \sum_{b\in\mathcal{A}} \Bigg[(\mathsf{q}_b^{(j)}(i)) \log(\mathsf{q}_b^{(j)}(i)) \\
&\qquad - \sum_{a\in\mathcal{A}} (\mathsf{q}_b^{(j)}(i) \mathsf{Q}_{a,b}^{(j)}(i)) \log\left(\mathsf{q}_b^{(j)}(i) \frac{\mathsf{Q}_{a,b}^{(j)}(i)}{E_{a,b}} \right) \Bigg] \\
&+ \sum_{j=0}^{m_i-n_i-1} \frac{1}{|\Delta_{n_i:m_i}|} \left[\log \sqrt{2\pi |\Xi_{n_i+j+1}|} + \frac{1}{12} \right] \\
\leq &\frac{\log |Z_{n_i}(\mathsf{q}^{(0)}(i))|}{|\Delta_{n_i:m_i}|} + \sum_{j=0}^{m_i-n_i-1} \frac{|\Xi_{n_i+j+1}|}{|\Delta_{n_i:m_i}|} \Phi(\mathsf{Q}^{(j)}(i)|E)^T \mathsf{q}^{(j)}(i) \\
&\qquad + (1+C) \cdot \log |\Delta_{n_i:m_i}| \cdot \frac{m_i - n_i}{|\Delta_{n_i:m_i}|}
\end{aligned}$$

for some constant $C = o(1)$ with respect to $|\Delta_{n:m}|$. Similarly,

$$\begin{aligned}
&\frac{\log |Z_{n_i:m_i}(\mathsf{q}(i), \mathsf{Q}(i))|}{|\Delta_{n_i:m_i}|} \\
\geq &\frac{\log |Z_{n_i}(\mathsf{q}^{(0)}(i))|}{|\Delta_{n_i:m_i}|} |\Delta_{n_i:m_i}| + \sum_{j=0}^{m_i-n_i-1} \frac{|\Xi_{n_i+j+1}|}{|\Delta_{n_i:m_i}|} \Phi(\mathsf{Q}^{(j)}(i)|E)^T \mathsf{q}^{(j)}(i) \\
&\qquad - \sum_{j=0}^{m_i-n_i-1} \frac{|\mathcal{A}|}{|\Delta_{n_i:m_i}|} \left[\log \sqrt{2\pi |\Xi_{n_i+j+1}|} + \frac{1}{12} \right] \\
\geq &\frac{\log |Z_{n_i}(\mathsf{q}^{(0)}(i))|}{|\Delta_{n_i:m_i}|} + \sum_{j=0}^{m_i-n_i-1} \frac{|\Xi_{n_i+j+1}|}{|\Delta_{n_i:m_i}|} \Phi(\mathsf{Q}^{(j)}(i)|E)^T \mathsf{q}^{(j)}(i) \\
&\qquad + C \cdot \log |\Delta_{n_i:m_i}| \cdot \frac{m_i - n_i}{|\Delta_{n_i:m_i}|}
\end{aligned}$$

for some constant $C = o(1)$. The proof is complete by noting that $\Delta_{n_i:m_i}$ is an unbounded increasing sequence. $\square$

As a corollary of the proposition above, we transform (Problem 1) into the following.

$$\text{(Problem 2)} \qquad \begin{cases} \text{maximize} & \frac{\log |Z_{n:n}(\mathsf{q}^{(0)}(i))|}{|\Delta_{n:n+k}|} + \sum_{j=0}^{k-1} \frac{|\Xi_{n+j+1}|}{|\Delta_{n:n+k}|} \Phi(\mathsf{Q}^{(j)}|E)^T \mathsf{q}^{(j)} \\ \text{subject to} & (\mathsf{q}, \mathsf{Q}) \in W_{n:n+k} \end{cases}$$

The trickiest part of (Problem 2) lies in the first term involving $|Z_{n:n}(\mathsf{q}^{(0)}(i))|$, for which, to the authors' knowledge, no good estimate is available. Nevertheless, our main interest lies in the case $k \to \infty$, for which it is not hard to see that the term tends to zero, and thus we shall continue our discussion by dropping it for the time being. We then reformulate the problem, by writing $\mathsf{p} = \overleftarrow{\mathsf{q}}$ and $\mathsf{P} = \overleftarrow{\mathsf{Q}}$ and replacing the feasible domain, as

$$\text{(Problem 3)} \qquad \begin{cases} \text{maximize} & \sum_{j=0}^{k-1} \frac{|\Xi_{n+k-j}|}{|\Delta_{n:n+k}|} \Phi(\mathsf{P}^{(j)}|E)^T \mathsf{p}^{(j+1)} \\ \text{subject to} & (\mathsf{p}, \mathsf{P}) \in \Omega_k \end{cases}$$

where

$$\Omega_k = \{(\mathsf{p}, \mathsf{P}) \in \Gamma_{\mathcal{A}}^{k+1} \times \Upsilon_{\mathcal{A}}^k : \mathsf{P}_{a,b}^{(j)} = 0 \text{ if } E_{a,b} = 0, \mathsf{p}^{(j)} = \mathsf{P}^{(j)}\mathsf{p}^{(j+1)}, 0 \le j < k\}.$$

We remark that the bottom-up convention of (Problem 3) turns out to be more convenient for our later discussion. In what follows, we show that the topological pressure remains unchanged. A quick glance at (Problem 3) reveals that for any fixed P, the objective function is just a linear function of $\mathsf{p}^{(0)}$, and therefore we can assume without loss of generality that there exists an optimal solution (p, P) with $\mathsf{p}^{(0)}$ of the form $\mathsf{p}^{(0)} = \mathsf{e}_a$ $(a \in \mathcal{A})$ which lies within

$$\begin{aligned} \Omega_k' = \{(\mathsf{p}, \mathsf{P}) \in \Gamma_{(\{\mathsf{e}_a\}_{a\in\mathcal{A}} \times \mathcal{A}^k)} \times \Upsilon_{\mathcal{A}}^k : \mathsf{P}_{a,b}^{(j)} = 0 \text{ if } E_{a,b} = 0 \\ \mathsf{p}^{(j+1)} = \mathsf{P}^{(j)}\mathsf{p}^{(j)}, \mathsf{P}_{a,b}^{(i)} = \frac{E_{a,b}}{\sum_{c\in\mathcal{A}} E_{c,b}} \text{ if } \mathsf{p}_b^{(j+1)} = 0, 0 \le j < k\}. \end{aligned}$$

Hence, it suffices to show the set $W_{n:n+k}$ is relatively dense in $\Omega_{n:n+k}$.

Lemma 8. *Let $(\mathsf{q}', \mathsf{Q}) \in D_n \times \Upsilon_{\mathcal{A}}$ be canonically admissible by E. Then, there exists $\mathsf{Q}' \in \Upsilon_{\mathcal{A}}$ such that $(\mathsf{q}', \mathsf{Q}') \in D_{n:n+1}$ and that*

$$\max_{a,b\in\mathcal{A}} |\mathsf{Q}_{a,b} - \mathsf{Q}'_{a,b}| \le \min_{\mathsf{q}'_a > 0} (|\Xi_{n+1}|\mathsf{q}'_a)^{-1}.$$

Moreover, ${\mathsf{Q}_{a,b}}' = 0$ if $\mathsf{Q}_{a,b} = 0$.

Proof. This is a consequence of the fact that $(\mathsf{q}', \mathsf{Q}') \in D_{n:n+1}$ if and only if $\mathsf{Q}' \in \Upsilon_{\mathcal{A}}$ and that the following hold.

- $\mathsf{Q}'_{a,b} = \frac{c_{a,b}}{|\Xi_{n+1}|\mathsf{q}'_b}$ for some integer $c_{a,b}$ if $\mathsf{q}'_b \neq 0$;
- $\mathsf{Q}'_{a,b} = E_{a,b} \cdot (\sum_{a\in\mathcal{A}} E_{c,b})^{-1}$ if $\mathsf{q}_b = 0$.

The lemma follows by defining Q' column by column. □

Proposition 9. $\lim_{n\to\infty} \sup_{(\mathsf{p},\mathsf{P})\in\Omega_k'} \mathsf{d}_{v,V}(W_{n:n+k}, (\overleftarrow{\mathsf{p}}, \overleftarrow{\mathsf{P}})) = 0.$

Proof. Let $\epsilon > 0$ and $(\mathsf{q}, \mathsf{Q}) = (\overleftarrow{\mathsf{p}}, \overleftarrow{\mathsf{P}}) \in \Omega_k$ be fixed, and denote

$$\zeta = \min \left\{ \min_{\mathsf{q}_a^{(i)} > 0} \mathsf{q}_a^{(i)}, \min_{\mathsf{Q}_{a,b}^{(i)} > 0} \mathsf{Q}_{a,b}^{(i)} \right\}.$$

We take n sufficiently large such that $\delta := \max_{\mathsf{q}_a^{(i)} > 0} (|\Xi_{i+1}|\mathsf{q}_a^{(i)})^{-1}$ satisfies $2|\mathcal{A}|k\frac{\delta}{\zeta} < \min\{\epsilon, \zeta, \frac{1}{2}\}$. We then construct a sequence $(\mathsf{q}', \mathsf{Q}') \in \Omega_k$ iteratively and show that this sequence also lies within $W_{n:n+k}$. For the initial distribution, take ${\mathsf{q}^{(0)}}' = \mathsf{q}^{(0)}$. Now if $(({\mathsf{q}^{(j)}}')_{0\le j\le i}, ({\mathsf{Q}^{(j)}}')_{0\le j\le i-1}) \in W_{n:n+i-1}$ is defined, one can choose ${\mathsf{q}^{(i+1)}}', {\mathsf{Q}^{(i)}}'$ such that $(({\mathsf{q}^{(j)}}')_{0\le j\le i+1}, ({\mathsf{Q}^{(j)}}')_{0\le j\le i}) \in W_{n:n+i}$ and ${\mathsf{Q}^{(i)}}'$ is a closest element to $\mathsf{Q}^{(i)}$. We then show that the following properties hold for all $i = 0, \cdots, k$.

(a) ${\mathsf{q}_a^{(i)}}' = 0$ if and only if $\mathsf{q}_a^{(i)} = 0$, and $\mathsf{d}_v({\mathsf{q}^{(i)}}', \mathsf{q}^{(i)}) \le i|\mathcal{A}|\delta$.

(b) $\max_{{\mathsf{q}_a^{(i)}}' > 0} (|\Xi_{i+1}|2{\mathsf{q}_a^{(i)}}')^{-1} \le 2\delta$

(c) ${\mathsf{Q}_{a,b}^{(i)}}' = 0$ if and only if $\mathsf{Q}_{a,b}^{(i)} = 0$, and $\mathsf{d}_V({\mathsf{Q}^{(i)}}', \mathsf{Q}^{(i)}) \le |\mathcal{A}|\delta$

We prove the foregoing by induction on i. When $i = 0$, since $\mathsf{q}^{(0)} = \mathsf{q}^{(0)\prime}$, property (a) and (b) are automatic and $\max_{a,b\in\mathcal{A}} |\mathsf{Q}^{(0)}_{a,b}| \le \delta$ follows immediately by Lemma 8. As a consequence, if $\mathsf{q}^{(1)}_{a,b} \neq 0$, then

$$\mathsf{Q}^{(1)\prime}_{a,b} \ge \mathsf{Q}^{(1)}_{a,b} - 2\cdot \mathsf{d}_V(\mathsf{Q}^{(1)\prime}, \mathsf{Q}^{(1)})\zeta \ge \zeta - |\mathcal{A}|\delta > 0,$$

and $\mathsf{Q}^{(1)\prime}_{a,b} = \mathsf{Q}^{(1)}_{a,b} = E_{a,b}\cdot(\sum_{a\in\mathcal{A}} E_{c,b})^{-1}$ if $\mathsf{q}^{(1)}_{a,b} = 0$ by definition. This finishes the proof of property (c).

As for the induction step, we assume the hypotheses hold for i. Under the circumstances, we have $\mathsf{d}_v(\mathsf{q}^{(i)\prime}, \mathsf{q}^{(i)}) \le i|\mathcal{A}|\delta$ and $\mathsf{d}_v(\mathsf{Q}^{(i)\prime}, \mathsf{Q}^{(i)}) \le |\mathcal{A}|\delta$ according to the induction hypothesis, which result in

$$\begin{aligned}\mathsf{d}_v(\mathsf{q}^{(i+1)\prime}, \mathsf{q}^{(i+1)}) &= \mathsf{d}_v(\mathsf{Q}^{(i)\prime}\mathsf{q}^{(i)\prime}, \mathsf{Q}^{(i)}\mathsf{q}^{(i)}) \\ &\le \mathsf{d}_V(\mathsf{Q}^{(i)\prime}, \mathsf{Q}^{(i)}) + \mathsf{d}_v(\mathsf{q}^{(i)\prime}, \mathsf{q}^{(i)}) \le (i+1)|\mathcal{A}|\delta\end{aligned}$$

and $\mathsf{q}^{(i+1)\prime}_a = (\mathsf{Q}^{(i)\prime}\mathsf{q}^{(i)\prime})_a > 0$ if and only if $\mathsf{q}^{(i+1)}_a = (\mathsf{Q}^{(i)}\mathsf{q}^{(i)})_a > 0$. This proves (a). As for (b), when $\mathsf{q}^{(i+1)\prime}_a \neq 0$,

$$\begin{aligned}(|\Xi_{i+2}|\mathsf{q}^{(i+1)\prime}_a)^{-1} &= \left(|\Xi_{i+2}|\mathsf{q}^{(i+1)}_a\left(1 - \frac{\mathsf{q}^{(i+1)\prime}_a - \mathsf{q}^{(i+1)}_a)}{\mathsf{q}^{(i+1)}_a}\right)\right)^{-1} \\ &\le \left(|\Xi_{i+2}|\mathsf{q}^{(i+1)}_a\left(1 - \frac{2\mathsf{d}_v(\mathsf{q}^{(i+1)\prime}, \mathsf{q}^{(i+1)})}{\mathsf{q}^{(i+1)}_a})\right)\right)^{-1} \\ &\le \left(|\Xi_{i+2}|\mathsf{q}^{(i+1)}_a\left(1 - \frac{|\mathcal{A}|(i+1)\delta}{\zeta})\right)\right)^{-1} \le 2\delta.\end{aligned}$$

Finally, we apply Lemma 8 again to obtain $\mathsf{d}_V(\mathsf{Q}^{(i+1)\prime}, \mathsf{Q}^{(i+1)}) \le |\mathcal{A}|\delta$. Furthermore, $(\mathsf{Q}^{(i+1)\prime}_{a,b}) = \mathsf{Q}^{(i+1)\prime}_{a,b}$ if $\mathsf{q}^{(i+1)}_{a,b} = 0$, and if $\mathsf{q}^{(i+1)}_{a,b} \neq 0$, then

$$\mathsf{Q}^{(i+1)\prime}_{a,b} \ge \mathsf{Q}^{(i+1)}_{a,b} - 2\cdot \mathsf{d}_V(\mathsf{Q}^{(i+1)\prime}, \mathsf{Q}^{(i+1)}) \ge \zeta - 2|\mathcal{A}|\delta > 0.$$

The property (c) is then proved, and the proof of induction is therefore completed. □

Corollary 10. (Problem 4) *shares the same upper and lower limits as* (Problem 2).

Proof. This is a consequence of Proposition 9 and the fact that the objective function of (Problem 3) is continuous with respect to the metric $\mathsf{d}_{v,V}$. □

Finally, we note that for $(\mathsf{p}, \mathsf{P}) \in \Omega_k$, $\Phi(\mathsf{P}^{(j)}|E)^T\mathsf{p}^{(j+1)}$ in (Problem 3) are uniformly bounded (independent of j, k and a), due to the continuity of the function on the compact set Ω_k. Furthermore, the coefficients $\frac{|\Xi_{n+k-j}|}{|\Delta_{n:n+k}|}$ in (Problem 3) add up to 1 with each of them admitting a limit

$$\lim_{k\to\infty} \frac{|\Xi_{n+k-j}|}{|\Delta_{n:n+k}|} = \lim_{k\to\infty} \frac{d-1}{d^{j+1}} \cdot \frac{d^k}{d^k - 1} = \frac{d-1}{d^{j+1}}$$

Therefore, we may rephrase (Problem 3) as the following optimization problem without changing the limiting behavior:

$$\text{(Problem 4)} \qquad \begin{cases} \text{maximize} & \sum_{j=0}^{k-1} \frac{d-1}{d^{j+1}} \Phi(\mathsf{P}^{(j)}|E)^T \mathsf{p}^{(j+1)} \\ \text{subject to} & (\mathsf{p}, \mathsf{P}) \in \Omega_k \end{cases}$$

which is related to the maximum of the infinite-dimensional optimization problem

$$\text{(Problem 5)} \qquad \begin{cases} \text{maximize} & \sum_{j=0}^{\infty} \frac{d-1}{d^{j+1}} \Phi(\mathsf{P}^{(j)}|E)^T \mathsf{p}^{(j+1)} \\ \text{subject to} & \mathsf{p} \in \Gamma_{\mathcal{A}}^{\mathbb{Z}_+}, \mathsf{P} \in \Upsilon_{\mathcal{A}}^{\mathbb{Z}_+} \\ & \mathsf{P}_{a,b}^{(j)} = 0 \text{ if } E_{a,b} = 0 \\ & \mathsf{p}^{(j)} = \mathsf{P}^{(j)} \mathsf{p}^{(j+1)}, 0 \le j < \infty \end{cases}.$$

We should note that the space $(\Gamma_{\mathcal{A}}^{\mathbb{Z}_+} \times \Upsilon_{\mathcal{A}}^{\mathbb{Z}_+}, \mathsf{d}_{v,V}^{\infty})$ is compact with respect to the metric $\mathsf{d}_{v,V}^{\infty}$, which is defined by

$$\mathsf{d}_{v,V}^{\infty}((\mathsf{p}, \mathsf{P}), (\mathsf{p}', \mathsf{P}')) = \sum_{j=0}^{\infty} \frac{d-1}{d^{j+1}} \mathsf{d}_{v,V}((\mathsf{p}^{(j)}, \mathsf{P}^{(j)}), (\mathsf{p}^{(j)\prime}, \mathsf{P}^{(j)\prime})),$$

It is not hard to see that the objective function is continuous with respect to this metric and that the feasible domain is compact, which ensure the existence of maximum points. Furthermore, if we denote $P^{(k)}(\cdot, E)$ and $P^{(\infty)}(\cdot, E)$ as the maxima of (Problem 4) and (Problem 5) respectively, then $P^{(\infty)}(\cdot, E)$ is continuous on $(0, 1)$. However, the proof for this requires some knowledge about the solutions of (Problem 4) and is on hold for the time being. At this point, we begin with the investigation of the maximizers of (Problem 4).

As has been pointed out, the objective function (Problem 4) is linear in $\mathsf{p}^{(k)}$ and thus $\mathsf{p} = \mathsf{e}_a$ can be assumed since $\Gamma_{\mathcal{A}}$ is the convex hull of $\{\mathsf{e}_a\}$. Therefore, we consider the optimization problem

$$\text{(19)} \qquad \begin{cases} \text{maximize} & \sum_{j=0}^{k-1} \frac{d-1}{d^{j+1}} \Phi(\mathsf{P}^{(j)}|E)^T \prod_{\ell=j+1}^{k-1} \mathsf{P}^{(\ell)} \mathsf{p} \\ \text{subject to} & \mathsf{p} = \mathsf{e}_a, \mathsf{P} \in \Upsilon_{\mathcal{A}}^{k} \\ & \mathsf{P}_{a,b}^{(j)} = 0 \text{ if } E_{a,b} = 0, 0 \le j < k \end{cases}$$

We denote by $F_k(\mathsf{p}, \mathsf{P}; 1/d, E)$ the objective function above and by $F_{\infty}^{(d)}(\mathsf{p}, \mathsf{P}; 1/d, E)$ the objective function of (Problem 5), and we deduce a maximizer as follows. For conciseness, we suppress E and d if doing so should not occasion any confusion.

Proposition 11. *The optimization problem* (19) *admits an optimal transition* $\mathsf{P}^{(i)}$, $i = 0, \cdots, k-1$, *independent of* p, *defined by*

$$\mathsf{P}_{a,b}^{(i)} = \begin{cases} \dfrac{e^{\frac{d^{i+1}}{d-1}\lambda_a^{(i)}} E_{a,b}}{\sum_{c:E_{c,b}>0} e^{\frac{d^{i+1}}{d-1}\lambda_c^{(i)}} E_{c,b}} & \textit{if } E_{a,b} > 0; \\ 0 & \textit{otherwise,} \end{cases},$$

where $\lambda^{(0)} = 0$ *and*

$$\lambda_a^{(i)} = \frac{d-1}{d^i} \log \sum_{b:E_{b,a}>0} e^{\frac{d^i}{d-1}\lambda_b^{(i-1)}} E_{b,a} \qquad \textit{for } i = 1, \cdots, k.$$

Moreover, the maximum of $F_k(\mathsf{p}, \mathsf{P})$ *is* ${\lambda^{(k)}}^T \mathsf{p}$.

Proof. The idea of proving the optimality is as follows. We construct a sequence P, independent of p and k, such that given any $i = 0, \cdots, k-1$ and $\mathsf{Q}^{(0)}, \cdots, \mathsf{Q}^{(i-1)} \in \Upsilon_{\mathcal{A}}$, $\mathsf{Z} = \mathsf{P}^{(i)}$ is a maximizer of the function

$$\mathsf{Z} \xmapsto{F_{k;\mathsf{Q}^{(i+1)},\cdots,\mathsf{Q}^{(k-1)}}} F_k(\mathsf{p}, \mathsf{P}^{(0)}, \cdots, \mathsf{P}^{(i-1)}, \mathsf{Z}, \mathsf{Q}^{(i+1)}, \cdots, \mathsf{Q}^{(k-1)}).$$

We now prove our claim. To this end, we introduce an auxiliary sequence $\lambda^{(i)}$ in our construction. By writing $\lambda^{(0)} = 0 \in \mathbb{R}^{\mathcal{A}}$, we note that $F_{k;\mathsf{Q}^{(1)},\cdots,\mathsf{Q}^{(k-1)}}$ is an entropy maximization problem in each column of Z and thus admits a maximizer (independent of p, Q and k) such that

$$\mathsf{P}^{(0)}_{a,b} = \begin{cases} \dfrac{e^{\frac{d}{d-1}\lambda^{(0)}_a} E_{a,b}}{\sum_{c:E_{c,b}>0} e^{\frac{d}{d-1}\lambda^{(0)}_c} E_{c,b}} = \dfrac{E_{a,b}}{\sum_{c\in\mathcal{A}} E_{c,b}} & \text{if } E_{c,b} > 0; \\ 0 & \text{otherwise,} \end{cases}$$

Now if we suppose a maximizer $\mathsf{P}^{(i-1)}$ satisfying our hypothesis is found, we let

$$\begin{aligned} \lambda^{(i)} :=& \frac{d-1}{d^i}\Phi(\mathsf{P}^{(i-1)}|E) + \mathsf{P}^{(i-1)^T}\lambda^{(i-1)} \\ =& \sum_{j=0}^{i-1} \frac{d-1}{d^{j+1}} \left(\prod_{\ell=j+1}^{i-1} \mathsf{P}^{(\ell)} \right)^T \Phi(\mathsf{P}^{(j)}|E) \end{aligned} \tag{20}$$

and $F_{k;\mathsf{Q}^{(i+1)},\cdots,\mathsf{Q}^{(k-1)}}$ can be written as

$$\begin{aligned} & F_{k;\mathsf{Q}^{(i+1)},\cdots,\mathsf{Q}^{(k-1)}}(\mathsf{Z}) \\ &= \sum_{j=0}^{i-1} \frac{d-1}{d^{j+1}} \Phi(\mathsf{P}^{(j)}|E)^T \left(\prod_{\ell=j+1}^{i-1} \mathsf{P}^{(\ell)} \right) Z \left(\prod_{\ell=i+1}^{k-1} \mathsf{Q}^{(\ell)} \right) \mathsf{p} \\ &\quad + \frac{d-1}{d^{i+1}} \Phi(\mathsf{Z}|E)^T \prod_{\ell=i+1}^{k-1} \mathsf{Q}^{(\ell)}\mathsf{p} + \sum_{j=i+1}^{k-1} \frac{d-1}{d^{j+1}} \Phi(\mathsf{Q}^{(j)}|E)^T \prod_{\ell=j+1}^{k-1} \mathsf{Q}^{(\ell)}\mathsf{p} \\ &= \left(\lambda^{(i)^T}\mathsf{Z} + \frac{d-1}{d^{i+1}}\Phi(\mathsf{Z}|E)^T \right) \prod_{\ell=i+1}^{k-1} \mathsf{Q}^{(\ell)}\mathsf{p} \\ &\quad + \sum_{j=i+1}^{k-1} \frac{d-1}{d^{j+1}} \Phi(\mathsf{Q}^{(j)}|E)^T \prod_{\ell=j+1}^{k-1} \mathsf{Q}^{(\ell)}\mathsf{p}, \end{aligned} \tag{21}$$

which also results in a classical entropy maximization problem in each column of Z, and an optimal solution independent of p, Q and k is

$$\mathsf{P}^{(i)}_{a,b} = \begin{cases} \dfrac{e^{\frac{d^{i+1}}{d-1}\lambda^{(i)}_a} E_{a,b}}{\sum_{c:E_{c,b}>0} e^{\frac{d^{i+1}}{d-1}\lambda^{(i)}_c} E_{c,b}} & \text{if } E_{a,b} > 0; \\ 0 & \text{otherwise,} \end{cases}, \tag{22}$$

In this manner, we successfully construct the desired sequence and prove the optimality. Moreover, if we plug (22) into (20), we deduce the recursive relation of λ, and the proof is complete. $\square$

Remark 12. *From Proposition 11, we note that the vector* $e^{d^{i+1}/(d-1)\lambda^{(i)}}$ *satisfies* $e^{d/(d-1)\lambda^{(0)}} = \mathbb{1}$ *and that*

$$e^{\frac{d^{i+1}}{d-1}\lambda^{(i)}} = (E^T e^{\frac{d^i}{d-1}\lambda^{(i-1)}})^d \text{ for } i \geq 1,$$

which is essentially a generalization of the formula used in [7, Algorithm 1] for the computation of entropy.

The maximizer P^* satisfies the following properties.

Theorem 13. *Let*

$$\mathcal{A}_\infty = \{a \in \mathcal{A} : \exists n \in \mathbb{N} \text{ such that } (E^n)_{a,a} > 0\},$$

and

$$L = \min\{n \in \mathbb{N} : (E^n)_{a,a} > 0, \forall a \in \mathcal{A}_\infty\}.$$

Then, the maximizer found in Proposition 11 satisfies the following properties.

(a) For every $\mathsf{p} \in \Gamma_\mathcal{A}$ *and* $k \geq 1$, $F_k(\mathsf{p}, \mathsf{P}^{(0:k-1)^*}) = {\lambda^{(k)}}^T \mathsf{p}$ *takes values in* $[(1 - d^{-k})\alpha, (1 - d^{-k})\beta]$, *where*

$$\alpha = \log \min_b \sum_a E_{a,b} \text{ and } \beta = \log \max_b \sum_a E_{a,b}.$$

(b) If $a, b \in \mathcal{A}$ *with* $(E^i)_{a,b} > 0$, *then for all* $j \geq 0$ *and* $k \geq 1$,

$$F_{k+i}(\mathsf{e}_b, \mathsf{P}^{(0:k+i-1)^*}) - \sum_{\ell=0}^{k+i-1} \frac{d-1}{d^{\ell+1}} \cdot \gamma \geq F_k(\mathsf{e}_a, \mathsf{P}^{(j:j+k-1)^*}) - \sum_{\ell=0}^{k-1} \frac{d-1}{d^{\ell+1}} \cdot \gamma,$$

where $\gamma = \log \min_{E_{a,b}>0} E_{a,b}$. *In particular, for all* $b \in \mathcal{A}_\infty$,

$$F_{Ln}(\mathsf{e}_b, \mathsf{P}^{(0:Ln-1)^*}) - \sum_{\ell=0}^{Ln-1} \frac{d-1}{d^{\ell+1}} \cdot \gamma$$

is non-negative and increasing.

(c) There exists a $\mathsf{p}^* \in \Gamma_\mathcal{A}^{\mathbb{Z}_+}$ *such that* $(\mathsf{p}^*, \mathsf{P}^*)$ *is a maximizer of* (Problem 5). *In particular,*

$$P^{(\infty)}(s, E) = F_\infty(\mathsf{p}^*, \mathsf{P}^*) = \max_{a \in \mathcal{A}_\infty} \lim_{n \to \infty} F_{Ln}(\mathsf{e}_a, \mathsf{P}^{(0:Ln-1)^*})$$
$$= \lim_{k \to \infty} \max_{a \in \mathcal{A}_\infty} F_k(\mathsf{e}_a, \mathsf{P}^{(0:k-1)^*}) = \lim_{k \to \infty} P^{(k)}(s, E).$$

We should note that under the assumption (A), $\mathcal{A}_\infty$ is nonempty, and thus L is well-defined.

4. Proof of Theorem 13

(a) We show by induction that the map $\lambda^{(i)} \xmapsto{f_i} \lambda^{(i+1)}$ in Proposition 11 satisfies that

$$f_k \circ \cdots \circ f_0(\lambda^{(0)})_a \in [(1 - d^{-k-1})\alpha, (1 - d^{-k-1})\beta] \quad \text{for all } k \geq 0.$$

The backbone of the proof is the monotonicity of f_i: $f_i(\lambda') \geq f_i(\lambda'')$ if $\lambda' \geq \lambda''$. When $i = 1$,

$$f_0(\lambda^{(1)})_a = \frac{d-1}{d} \log \sum_{b: E_{b,a}>0} E_{b,a} \in [(1 - d^{-1})\alpha, (1 - d^{-1})\beta].$$

Now if the hypothesis holds for $k-1$, then

$$\begin{aligned} f_k(f_{k-1}\circ\cdots\circ f_0(\lambda^{(0)}))_a &\le \frac{d-1}{d^{k+1}}\log\sum_{b:E_{b,a}>0}\beta^{(1-d^{-k})\cdot\frac{d^{k+1}}{d-1}}E_{b,a} \\ &\le (1-d^{-k-1})\beta, \end{aligned}$$

and a similar argument also applies to the lower bound $f_k(f_{k-1}\circ\cdots\circ f_0(\lambda^{(0)}))_a \ge (1-d^{-k-1})\alpha$. The first item then holds by induction.

(b) For each $a,b\in\mathcal{A}$, one can choose a sequence $(\xi_\ell)_{\ell\ge 0}$ as follows:

$$(\xi_\ell)_{\ell\ge0} = (\xi_\ell^{a,b,i})_{\ell\ge0} \text{ such that } \xi_0=a, \xi_i=b, E_{\xi_\ell,\xi_{\ell+1}}>0, \forall \ell\ge 0. \tag{23}$$

Then, we take transition matrices $\mathsf{P}^{(k+\ell)'}$ such that ${\mathsf{P}^{(k+\ell)'}}_{\xi_\ell,\xi_{\ell+1}}=1$. Under the circumstances, we have $(\prod_{\ell=0}^{k+1}\mathsf{P}^{(k+j)'})_{a,b}\mathsf{e}_b=\mathsf{e}_a$. Therefore, according to (19),

$$\begin{aligned} &F_k(\mathsf{e}_a,\mathsf{P}^{(j:j+k-1)^*})+\sum_{\ell=k}^{k+i-1}\frac{d-1}{d^{\ell+1}}\cdot\gamma \\ =&\sum_{\ell=0}^{k-1}\frac{d-1}{d^{\ell+1}}\Phi(\mathsf{P}^{(j+\ell)^*}|E)^T\prod_{\ell'=j+\ell+1}^{k-1}\mathsf{P}^{(\ell')^*}\mathsf{e}_a+\sum_{\ell=k}^{k+i-1}\frac{d-1}{d^{\ell+1}}\cdot\gamma \\ \le& F_{k+i}(\mathsf{e}_b,(\mathsf{P}^{(j:j+k-1)^*},\mathsf{P}^{(j+k:j+k+i-1)'})) \\ \le& F_{k+i}(\mathsf{e}_b,\mathsf{P}^{(0:k+i-1)^*}). \end{aligned}$$

This finishes the proof.

(c) To begin with, we note the existence of optimizer $(\mathsf{p}',\mathsf{P}')$ is guaranteed by the compactness of the feasible domain and the continuity of the objective function. Next, by recursively replacing P' by P^* according to Proposition 11, the compactness again asserts that there exists a maximizer of the form $(\mathsf{p}^*,\mathsf{P}^*)$ (and thus the first equality). To prove the second equality, for each $a\in\mathcal{A}$ and $n\ge0$ we may extend, in the manner of (23), the admissible pair $((\prod_{\ell=j}^{Ln-1}\mathsf{P}^{(\ell)^*}\mathsf{e}_a)_{j=0}^{Ln-1},\mathsf{P}^{(0:Ln-1)^*})$ for (Problem 4) to a pair of infinite sequence $(\mathsf{p}(a,n),\mathsf{P}(a,n))$ so that

$$F_{Ln}(\mathsf{e}_a,\mathsf{P}^{(0:Ln-1)^*})+\sum_{\ell=Ln}^{\infty}\frac{d-1}{d^{\ell+1}}\cdot\gamma\le F_\infty(\mathsf{p}(a,n),\mathsf{P}(a,n)).$$

(More specifically, we take $(\xi_\ell)=(\xi^{a,a,0})$.) Then, $\lim_{n\ge0}\max_{a\in\mathcal{A}_\infty}=\sup_{n\ge0}\max_{a\in\mathcal{A}_\infty}=\max_{a\in\mathcal{A}_\infty}\sup_{n\ge0}=\max_{a\in\mathcal{A}_\infty}\lim_{n\to\infty}$ all coincide on the sequence $F_{Ln}(\mathsf{e}_a,\mathsf{P}^{(0:Ln-1)^*})-\sum_{\ell=0}^{Ln-1}\frac{d-1}{d^{\ell+1}}\cdot\gamma$ and

$$\begin{aligned} &\max_{a\in\mathcal{A}_\infty}\lim_{n\to\infty}F_{Ln}(\mathsf{e}_a,\mathsf{P}^{(0:Ln-1)^*})-\sum_{\ell=0}^{Ln-1}\frac{d-1}{d^{\ell+1}}\cdot\gamma \\ \le&\max_{a\in\mathcal{A}_\infty}F_\infty(\mathsf{p}(a,n),\mathsf{P}(a,n))-\sum_{\ell=0}^{\infty}\frac{d-1}{d^{\ell+1}}\cdot\gamma \\ \le& F_\infty(\mathsf{p}^*,\mathsf{P}^*)-\sum_{\ell=0}^{\infty}\frac{d-1}{d^{\ell+1}}\cdot\gamma, \end{aligned}$$

which implies $F_\infty(\mathsf{p}^*,\mathsf{P}^*) \geq \max_{a\in\mathcal{A}_\infty}\lim_{n\to\infty} F_{Ln}(\mathsf{e}_a,\mathsf{P}^{(0:Ln-1)^*})$. On the other hand, due to assumption (A), by writing

$$\mathcal{A}'_\infty = \{a\in\mathcal{A} : \exists b\in\mathcal{A}_\infty, \exists i\in\mathbb{N} \text{ such that } (E^i)_{a,b}>0\},$$

we have that for each $a\in\mathcal{A}'_\infty$ and each $n\geq|\mathcal{A}|$, there exists $b\in\mathcal{A}_\infty$ such that $(E^n)_{a,b}>0$. In addition, $\mathsf{p}_a^{(n)^*}>0$ implies $a\in\mathcal{A}'_\infty$. Hence, by (b) and the fact that $\Phi(\mathsf{p}^{(\ell)}|E)\leq\beta\cdot\mathbb{1}$,

$$\begin{aligned}
&F_\infty(\mathsf{p}^*,\mathsf{P}^*) - \sum_{\ell=Ln}^{\infty}\frac{d-1}{d^{\ell+1}}\cdot\beta \\
&=\sum_{\ell=0}^{\infty}\frac{d-1}{d^{\ell+1}}\Phi(\mathsf{P}^{(\ell)^*}|E)^T\mathsf{p}^{(\ell+1)^*} - \sum_{\ell=Ln}^{\infty}\frac{d-1}{d^{\ell+1}}\cdot\beta \\
&\leq \max_{a\in\mathcal{A}'_\infty} F_{Ln}(\mathsf{e}_a,\mathsf{P}^{(0:Ln-1)^*}) \\
&\leq \max_{a\in\mathcal{A}_\infty} F_{L(n+|\mathcal{A}|)}(\mathsf{e}_a,\mathsf{P}^{(0:L(n+|\mathcal{A}|)-1)^*}) - \sum_{\ell=Ln}^{L(n+|\mathcal{A}|)-1}\frac{d-1}{d^{\ell+1}}\cdot\gamma
\end{aligned}\tag{24}$$

and thus

$$F_\infty(\mathsf{p}^*,\mathsf{P}^*) \leq \lim_{n\to\infty}\max_{a\in\mathcal{A}_\infty} F_{Ln}(\mathsf{e}_a,\mathsf{P}^{(0:Ln-1)^*}) = \max_{a\in\mathcal{A}_\infty}\lim_{n\to\infty} F_{Ln}(\mathsf{e}_a,\mathsf{P}^{(0:Ln-1)^*}),$$

which proves the second equality. As for the third equality, it suffices to justify the existence of $\lim_{k\to\infty}\max_{a\in\mathcal{A}_\infty} F_k(\mathsf{e}_a,\mathsf{P}^{(0:k-1)^*})$, so the equality follows from the argument above for the subsequence $\max_{a\in\mathcal{A}_\infty} F_{Ln}(\mathsf{e}_a,\mathsf{P}^{(0:Ln-1)^*})$. Indeed, for every $b\in\mathcal{A}_\infty$, there exists $a\in\mathcal{A}_\infty$ such that $E_{a,b}>0$, and thus, by (b),

$$\max_{a\in\mathcal{A}_\infty} F_k(\mathsf{e}_a,\mathsf{P}^{(0:k-1)^*}) - \sum_{\ell=0}^{k-1}\frac{d-1}{d^{\ell+1}}\cdot\gamma \leq \max_{a\in\mathcal{A}_\infty} F_{k+1}(\mathsf{e}_a,\mathsf{P}^{(0:k)^*}) - \sum_{\ell=0}^{k}\frac{d-1}{d^{\ell+1}}\cdot\gamma$$

as desired. Finally, to prove the last equality, it suffices to demonstrate that

$$\lim_{k\to\infty}\max_{a\in\mathcal{A}_\infty} F_k(\mathsf{e}_a,\mathsf{P}^{(0:k-1)^*}) = \lim_{k\to\infty}\max_{a\in\mathcal{A}} F_k(\mathsf{e}_a,\mathsf{P}^{(0:k-1)^*}) \left(= \lim_{k\to\infty} P^{(k)}(s,E)\right).$$

To this end, we note that there for $i\geq|\mathcal{A}|$, $(E^i)_{a,b}>0$ implies $a\in\mathcal{A}'_\infty$, and thus we can adopt a similar argument as (24) to deduce

$$\begin{aligned}
&\max_{a\in\mathcal{A}_\infty} F_k(\mathsf{e}_a,\mathsf{P}^{(0:k-1)^*}) \leq \max_{a\in\mathcal{A}} F_k(\mathsf{e}_a,\mathsf{P}^{(0:k-1)^*}) \\
&\leq \max_{a\in\mathcal{A}'} F_{k-|\mathcal{A}|}(\mathsf{e}_a,\mathsf{P}^{(0:k-|\mathcal{A}|-1)^*}) + \sum_{\ell=k-|\mathcal{A}|}^{k-1}\frac{d-1}{d^{\ell+1}}\cdot\beta \\
&\leq \max_{a\in\mathcal{A}_\infty} F_k(\mathsf{e}_a,\mathsf{P}^{(0:k)^*}) - \sum_{\ell=k-|\mathcal{A}|}^{k-1}\frac{d-1}{d^{\ell+1}}\cdot\gamma + \sum_{\ell=k-|\mathcal{A}|}^{k-1}\frac{d-1}{d^{\ell+1}}\cdot\beta.
\end{aligned}$$

by applying the squeeze theorem.

5. Proof of Theorem 1

Let β and γ be the constants defined in Theorem 13, which are independent of s. We show that $P^{(k)}$ converges to $P^{(\infty)}$ in a strong sense.

Proof of Theorem 1. We note that ${\mathsf{P}^{(i)}}^* = {\mathsf{P}^{(i)}}^*(s)$ is a C^1 function of s on the interval $(0,1)$. Hence, for any fixed p and $n, m \in \mathbb{N}$, the maximum

$$F_{m-n}(\mathsf{p}, {\mathsf{P}^{(n:m-1)}}^*(s); s, E) = {\lambda^{(m-n)}}^*(s)^T \mathsf{p}$$

of (19) is also C^1 on the interval $(0,1)$. For convenience, we denote

$$G_{n:m}(\mathsf{p}; s, E) = F_{m-n}(\mathsf{p}, {\mathsf{P}^{(n:m-1)}}^*(s); s, E) - \sum_{\ell=0}^{m-n-1} (s^\ell - s^{\ell+1}) \cdot \gamma.$$

Also, by denoting $q_k(s) = s^k - s^{k+1}$, we have

$${\lambda^{(k)}}^*(s) = \sum_{j=0}^{k-1} q_j(s) \left(\prod_{\ell=j+1}^{k-1} {\mathsf{P}^{(\ell)}}^*(s) \right)^T \Phi({\mathsf{P}^{(j)}}^*(s)|E).$$

and we can compute the derivative of the vector ${\lambda^{(k)}}^*$

$$\begin{aligned}
&\frac{d}{ds}{\lambda^{(k)}}^*(s) = \frac{d}{ds} q_k(s) \log\left(E^T e^{q_k(s)^{-1}{\lambda^{(k-1)}}^*(s)}\right) \\
=& q_k'(s) \log\left(E^T e^{q_k(s)^{-1}{\lambda^{(k-1)}}^*(s)}\right) + \frac{q_k(s)}{\left(E^T e^{q_k(s)^{-1}{\lambda^{(k-1)}}^*(s)}\right)} \cdot \\
&\left[E^T \mathrm{diag}\left(-\frac{q_k'(s)}{q_k(s)^2} {\lambda^{(k-1)}}^*(s) + \frac{1}{q_k(s)} \frac{d}{ds} {\lambda^{(k-1)}}^*(s) \right) e^{q_k(s)^{-1}{\lambda^{(k-1)}}^*(s)} \right] \\
=& q_k'(s) \Phi({\mathsf{P}^{(k-1)}}^*(s)|E) + {\mathsf{P}^{(k-1)}}^*(s)^T \frac{d}{ds} {\lambda^{(k-1)}}^*(s) \qquad (25) \\
=& \sum_{j=0}^{k-1} q_k'(s) \left(\prod_{\ell=j+1}^{k-1} {\mathsf{P}^{(\ell)}}^*(s) \right)^T \Phi({\mathsf{P}^{(j)}}^*(s)|E) \\
=& \sum_{j=0}^{k-1} (j s^{j-1} - (j+1)s^j) \left(\prod_{\ell=j+1}^{k-1} {\mathsf{P}^{(\ell)}}^*(s) \right)^T \Phi({\mathsf{P}^{(j)}}^*(s)|E).
\end{aligned}$$

(a) For convergence of ${\lambda_a^{(Ln)}}^* = F_{Ln}(\mathsf{e}_a, {\mathsf{P}^{(0:Ln-1)}}^*)$ for $a \in \mathcal{A}_\infty$, we note that, by Theorem 13, $G_{0:Ln}(\mathsf{e}_a, {\mathsf{P}^{(0:Ln-1)}}^*)$ is increasing in n. In addition,

$$\left| \frac{d}{ds} {\lambda^{(Ln)}}^*(s) \right| \le \sum_{j=0}^{\infty} (j s^{j-1} - (j+1)s^j)(|\beta| + |\gamma|),$$

which assures the equicontinuity of $G_{0:Ln}(\mathsf{e}_a, {\mathsf{P}^{(0:Ln-1)}}^*(s))$ on any compact sub-interval of $(0,1)$. Hence, by the Arzelà-Ascoli theorem we know the convergence of $G_{0:Ln}(\mathsf{e}_a, {\mathsf{P}^{(0:Ln-1)}}^*(s))$ is uniform on any compact sub-interval of $(0,1)$. This implies

$$\lim_{n\to\infty} F_{Ln}(\mathsf{e}_a, {\mathsf{P}^{(0:Ln-1)}}^*) = \sup_{n\ge 0} G_{0:Ln}(\mathsf{e}_a) + \gamma$$

exists and is continuous, and so is

$$P^{(\infty)}(s,E)=\max_{a\in\mathcal{A}_\infty}\lim_{n\to\infty}F_{Ln}(\mathsf{e}_a,\mathsf{P}^{(0:Ln-1)^*}(s);s,E).$$

The item is thus proved.

(b) Since $P^{(\infty)}(s,E)$ is the limiting function of $P^{(k)}(s,E)$ as shown in Theorem 13 and $P^{(k)}(s,E)$ shares the same limiting behavior as the maximizer of (Problem 1) as $k\to\infty$, the claim is proved.

(c) From the argument above, we claim that the interval $(0,1)$ contains disjoint open sets I_a, $a\in\mathcal{A}_\infty$, on which

$$\sup_{n\geq 0}G_{0:Ln}(\mathsf{e}_a;s,E)\geq\sup_{n\geq 0}G_{0:Ln}(\mathsf{e}_b;s,E)\quad\text{for all } b\in\mathcal{A}_\infty, s\in I_a$$

such that $\cup_{a\in\mathcal{A}_\infty}I_a$ is dense in $(0,1)$. This follows inductively from the fact that for any continuous functions $f_1,f_2:U\subset\mathbb{R}^n\to\mathbb{R}$ defined on some open set U,

$$\begin{aligned}\{x\in U:f_1(x)>f_2(x)\}\cup\{x\in U:f_2(x)>f_1(x)\}&\\ \cup\, int(\{x\in U:f_2(x)=f_1(x)\})&\end{aligned}$$

is dense in U. Therefore, for all $s\in I_a$,

$$P^{(\infty)}(s,E)=\lim_{n\to\infty}\left[F_{Ln}(\mathsf{e}_a,\mathsf{P}^{(0:Ln-1)^*}(s);s,E)-\sum_{j=0}^{Ln-1}(s^j-s^{j+1})\gamma\right],$$

and it is sufficient to prove the monotonicity on each I_a. Now, consider an alternative expression of (25):

$$\begin{aligned}\frac{d}{ds}\mathsf{e}_a^T\lambda^{(k)^*}(s)&=\sum_{j=0}^{k-1}(js^{j-1}-(j+1)s^j)\mathsf{e}_a^T\left(\prod_{\ell=j+1}^{k-1}\mathsf{P}^{(\ell)^*}(s)\right)^T\Phi(\mathsf{P}^{(j)^*}(s)|E)\\ &=\sum_{j=1}^{k-1}\sum_{i=1}^{j}(s^{j-1}-s^j)\mathsf{e}_a^T\left(\prod_{\ell=j+1}^{k-1}\mathsf{P}^{(\ell)^*}(s)\right)^T\Phi(\mathsf{P}^{(j)^*}(s)(s)|E)\\ &\qquad-\sum_{j=0}^{k-1}s^j\mathsf{e}_a^T\left(\prod_{\ell=j+1}^{k-1}\mathsf{P}^{(\ell)^*}(s)\right)^T\Phi(\mathsf{P}^{(j)^*}(s)|E)\\ &=\sum_{i=1}^{k-1}s^{i-1}F_{k-i}(\mathsf{e}_a,\mathsf{P}^{(i:k-1)^*}(s);s,E)-\frac{1}{1-s}\lambda_a^{(k)^*}(s),\end{aligned}\tag{26}$$

where $\frac{1}{1-s}\lambda_a^{(Ln)^*}\to P^{(\infty)}(s,E)$ for each $s\in I_a$ as $n\to\infty$. Expressing (26) in terms of function G, one derives, by noting $(1-s)^{-1}=\sum_{i=1}^\infty s^{i-1}$, that

$$\begin{aligned}\frac{d}{ds}\mathsf{e}_a^T\lambda^{(Ln)^*}(s)=&\sum_{i=1}^{Ln-1}s^{i-1}\left[G_{i:Ln}(\mathsf{e}_a;s,E)-G_{0:Ln}(\mathsf{e}_a;s,E)\right]\\ &-\sum_{i=Ln}^{\infty}s^{i-1}\lambda^{(Ln)^*}(s)_a-Lns^{Ln-1}\gamma,\end{aligned}$$

in which the first sum is non-positive while the rest altogether are bounded above by $\frac{s^{Ln-1}(1-s^{Ln-1})}{1-s}(|\alpha|+|\beta|)-Lns^{Ln-1}\gamma$. By the mean value theorem, this provides the following bound for any $[s', s''] \subset I_a$:

$$\begin{aligned} & P^{(\infty)}(s'', E) - P^{(\infty)}(s', E) \\ &= \lim_{n\to\infty} F_{Ln}(\mathsf{e}_a, \mathsf{P}^{(0:Ln-1)^*}; s'', E) - F_{Ln}(\mathsf{e}_a, \mathsf{P}^{(0:Ln-1)^*}; s', E) \\ &\leq \limsup_{n\to\infty} \max_{s\in[s',s'']} \left[\frac{s^{Ln-1}(1-s^{Ln-1})}{1-s}(|\alpha|+|\beta|) - Lns^{Ln-1}\gamma \right](s''-s') = 0. \end{aligned}$$

(d) We first prove (6). The first equality follows from (c). As for the third equality, we should note that $\mathbf{P}_*(1/d, E) \geq \beta = \log r_E$ for all $d \geq 2$ since E satisfies $\sum_{a\in\mathcal{A}} E_{a,b} > 0$ (which is equivalent to $\sum_{a\in\mathcal{A}} E_{a,b} > 0$ in this case) for all $b \in \mathcal{A}$. More specifically, we can find $a \in \mathcal{A}$ satisfying $\sum_{b\in\mathcal{A}} E_{a,b} = r_E$ and a sequence $a_0 a_1 \cdots a_{n-1} = a$ satisfying $E_{a_i, a_{i+1}} = 1$ such that

$$|Z_{n:n}| \geq |\{u \in B_{0:n} : u|_{\Delta_{n-1}} \in B_{0:n-1}(\{\mathsf{e}_{a_i}\}_{i=0}^{n-1})\}| = e^{\beta|\Xi_n|}.$$

To prove the other inequality, we observe that

$$\lim_{n\to\infty} \frac{\log|Z_{n:n}|}{|\Xi_n|} \leq \lim_{n\to\infty} \frac{\log|B_{n-1:n-1}|e^{\beta|\Xi_n|}}{|\Xi_n|} \leq \frac{1}{d}\cdot\log|\mathcal{A}| + \beta,$$

which approaches β as $d \to \infty$. It now remains to prove the second equality in (6). To this end, Theorem 13 guarantees $P^{(\infty)}(0+, E) \leq \beta$, and according to (Problem 5) we have

$$\begin{aligned} P^{(\infty)}(0+, E) &= \lim_{s\to 0^+} P^{(\infty)}(s, E) \\ &\geq \lim_{s\to 0^+} P^{(1)}(s, E) = \lim_{s\to 0^+}(1-s)\log\max_{b\in\mathcal{A}} \sum_{a\in\mathcal{A}} E_{a,b} = \beta. \end{aligned}$$

As for $P^{(\infty)}(1-, E)$, we plug the Parry measure

$$\mathsf{P}^{(i)}_{a,b} = \frac{E_{b,a} w_a}{\rho(E) w_b} \quad \text{and} \quad \mathsf{p}_a = v_a w_a,$$

into (Problem 4) to deduce $P^{(\infty)}(1-, E) \geq \log\rho(E)$. To prove the other inequality, for every $\epsilon > 0$ we first construct the interaction matrix

$$E' = \begin{bmatrix} \rho(E)+\epsilon & 0 \\ \mathbb{1} & E \end{bmatrix}.$$

Since $P^{(\infty)}(s, E) \leq P^{(\infty)}(s, E')$ for all $k \geq 0$ and $d \geq 2$ according to (Problem 5), it is sufficient to show that $P^{(\infty)}(1-, E') \leq \log\rho(E') = \log\rho(E) + \epsilon$. Let v and w be the left and right eigenvector, respectively, associated with $\rho(E')$ such that $w^T v = 1$. According to the spectral decomposition of E', we know $\lim_{n\to\infty}\rho(E')^{-n}(E')^n = vw^T$ is non-negative, and thus we can assume without loss of generality that $v = \rho(E')^{-1}E'v$ is a positive probability vector. We see from Remark 12 that for the system defined by E', there exists $C_d = \max_{a\in\mathcal{A}'} v_a^{1-d} \geq 1$ such that

$$\begin{aligned} & v^T e^{\frac{d^{k+1}}{d-1}\lambda^{(k)}} = v^T((E')^T e^{\frac{d^k}{d-1}\lambda^{(k-1)}})^d \\ \leq & C_d (v^T (E')^T e^{\frac{d^k}{d-1}\lambda^{(k-1)}})^d = C_d \cdot \rho(E')^d (v^T e^{\frac{d^k}{d-1}\lambda^{(k-1)}})^d. \end{aligned} \tag{27}$$

Indeed, we note that for all non-negative numbers x_a,

$$1 \leq \frac{\sum_{a\in\mathcal{A}'} v_a x_a^d}{(\sum_{a\in\mathcal{A}'} v_a x_a)^d} \leq \max_{a\in\mathcal{A}'} \frac{v_a x_a^d}{v_a^d x_a^d} = \max_{a\in\mathcal{A}'} v_a^{1-d}.$$

This provides a uniform bound for $P^{(\infty)}(1-, E')$, and the claim is proved by estimating the pressure using (27) and letting $d \to 1$. More specifically,

$$\begin{aligned} v^T e^{\frac{d^{k+1}}{d-1}\lambda^{(k)}} &\leq C_d^{1+d+\cdots+d^{k-1}} \rho(E')^{d+d^2+\cdots+d^k} (v^T e^{\frac{d^k}{d-1}\lambda^{(0)}})^d \\ &= C_d^{1+d+\cdots+d^{k-1}} \rho(E')^{d+d^2+\cdots+d^k}, \end{aligned}$$

and thus $P^{(\infty)}(1-, E) \leq \log(\rho(E) + \epsilon)$. □

6. Experiments

In this section, we give two examples related to Theorem 1. Consider the golden-mean tree-shifts $X_G^{\times d}$ with $G = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. We claim that the function $P^{(\infty)}(s, E)$ can be approximated by $P^{(k)}(s, E)$ with a crude estimate of error

$$(1-s)^{-1} s^{\ell} \cdot \gamma \leq P^{(\infty)}(s, E) - P^{(k)}(s, E) \leq (1-s)^{-1} s^{\ell} \cdot \beta.$$

Indeed, the first inequality follows from Theorem 13 (b), and the second follows, by comparing (Problem 4) and (Problem 5), from the fact that each term in (Problem 5) admits an upper bound $\frac{d-1}{d^{j+1}} \Phi(\mathsf{P}^{(j)}|E)\mathsf{p}^{(j+1)} \leq \frac{d-1}{d^{j+1}}\beta$. The figure of the topological entropy is given in Figure 1. For the purpose of demonstration of Theorem 1 for general interaction matrices, we include Figure 2 to show the increasing property of pressure when

$$w = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} 2 & 2 \\ 1 & 0 \end{bmatrix}.$$

Both of the figures turn out to be consistent with Theorem 1 in the sense that both functions are continuous, $P^{(\infty)}(0+, G) = \log_{10} 2 \approx 0.3010$, $P^{(\infty)}(0+, E) = \log_{10} 3 \approx 0.4771$, $P^{(\infty)}(1-, G) = \log_{10} \frac{1+\sqrt{5}}{2} \approx 0.2090$, and $P^{(\infty)}(1-, E) = \log_{10}(1+\sqrt{3}) \approx 0.4365$.

References

[1] N. Aubrun and M.-P. Beal, *Tree-shifts of finite type*, Theoret. Comput. Sci **459** (2012), 16–25.

[2] J.-C. Ban and C.-H. Chang, *Tree-shifts: Irreducibility, mixing, and chaos of tree-shifts*, Trans. Am. Math. Soc. **369** (2017), 8389–8407.

[3] ———, *Tree-shifts: The entropy of tree-shifts of finite type*, Nonlinearity **30** (2017), 2785–2804.

[4] J.-C. Ban, C.-H. Chang, W.-G. Hu, and Y.-L. Wu, *On structure of topological entropy for tree-shift of finite type*, J. Differential Equations **292** (2021), 325–353.

[5] ———, *Topological entropy for shifts of finite type over $\mathbb{Z}$ and trees*, Theoret. Comput. Sci **930** (2022), 24–32.

[6] J.-C. Ban, C.-H. Chang, W.-G. Hu, G.-Y. Lai, and Y.-L. Wu, *An analogue of topological sequence entropy for Markov hom tree-shifts*, Studia Math., accepted.

[7] J.-C. Ban, C.-H. Chang, Y.-L. Wu, and Y.-Y. Wu, *Stem and topological entropy on Cayley trees*, Math. Phys. Anal. Geom **25** (2022), no. 1, 1.

[8] I. Benjamini and Y. Peres, *Markov chains indexed by trees*, Ann. Probab. (1994), 219–243.

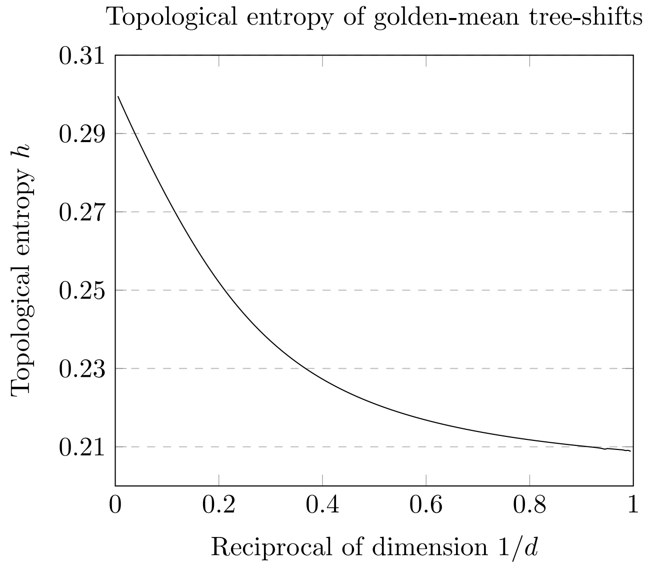

FIGURE 1. Topological entropy of golden-mean tree-shifts

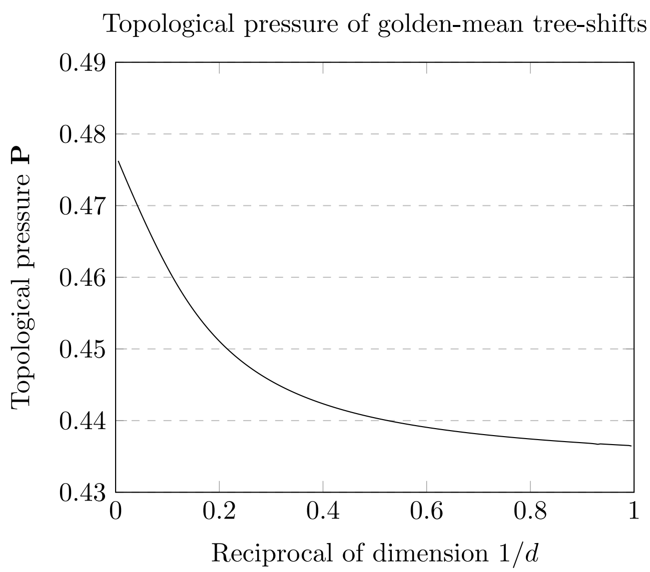

FIGURE 2. Topological pressure of golden-mean tree-shifts

[9] T. Berger and Z. Ye, *Entropic aspects of random fields on trees*, IEEE Trans. Inf. Theory **36** (1990), no. 5, 1006–1018.
[10] T. Ceccherini-Silberstein and M. Coornaert, *Cellular automata and groups*, Springer Science & Business Media, 2010.

[11] W. Huang, S. Shao, and X.-D. Ye, *Mixing via sequence entropy*, Contemp. Math. **385** (2005), 101–122.

[12] W. Huang and X.-D. Ye, *Combinatorial lemmas and applications to dynamics*, Adv. Math. **220** (2009), no. 6, 1689–1716.

[13] E. Louidor, B. Marcus, and R. Pavlov, *Independence entropy of $\mathbb{Z}^d$-shift spaces*, Acta Appl. Math. **126** (2013), no. 1, 297–317.

[14] T. Meyerovitch and R. Pavlov, *On independence and entropy for high-dimensional isotropic subshifts*, Proc. Lond. Math. Soc. **109** (2014), no. 4, 921–945.

[15] K. Petersen and I. Salama, *Tree shift topological entropy*, Theoret. Comput. Sci **743** (2018), 64–71.

[16] ———, *Entropy on regular trees*, Discrete Contin. Dyn. Syst. **40** (2020), no. 7, 4453–4477.

[17] ———, *Asymptotic pressure on some self-similar trees*, Stoch. Dyn. (2022).

[18] S. T. Piantadosi, *Symbolic Dynamics on Free Groups*, Discrete Contin. Dyn. Syst. **20** (2008), no. 3, 725–738.

[19] F. Tan, X. Ye, and R.-F. Zhang, *The set of sequence entropies for a given space*, Nonlinearity **23** (2009), no. 1, 159.