HAUSDORFF DIMENSIONS OF TOPOLOGICALLY TRANSITIVE MARKOV HOM TREE-SHIFTS

Abstract. This paper features an analog of Sanov's theorem for finite-state Markov chains indexed by rooted d-trees, obtained via the method of types in the classical analysis of large deviations. Along with the theorem comes two applications: an almost-sure type convergence of sample mean and a formula for Hausdorff dimension of the symbolic space associated with the irreducible Markov chain.

1. Introduction

This paper presents a version of Sanov's theorem for finite-state Markov chains indexed by d-trees, accompanied by a number of folklore theorems regarding the distribution of canonical finite-state Markov chains as well as a touch of their interplay with the dimension theory.

The investigation of the tree-indexed Markov chains was initiated by Benjamini and Peres [4]. In the work, the authors studied the tree-indexed random walk on a (countable) graph in an attempt to understand the behavior of sample paths in a collective manner, and in particular, investigated the recurrence and ray-recurrence to give an insight into the behavior of random walks. Aside from that, this type of random models has connections with various mathematical fields, including percolation theory. To be precise, the percolation on the rooted d-tree can be modeled by a Markov chain involving only two states "open" and "closed". Considerations of such problems are taken in the work by [5]. Furthermore, another frequently explored aspect of the non-conventional Markov chains is their entropy, which is also the theme of the first part of this work. In this direction, researchers attempt to extend well-established theorems for canonical Markov chains, including the law of large numbers, Shannon-McMillan-Breiman theorem and ergodic theorems, to the tree-indexed systems. This line of research traces its origin back to [6], in which the authors studied stationary random fields on rooted/unrooted d-trees. Subsequent works, such as [17, 8], have also made progress in this field. Noteworthy contributions regarding deviation inequalities are also provided by [12], [13] and [16].

Let $(X_i)_{i\in\mathbb{N}}$ be i.i.d. random variables taking values in a finite subset of $\mathbb{R}$, say $\mathcal{A} = \{\log a_1, \cdots, \log a_k\}$, with the probabilities $p_1, p_2, \cdots p_k$, respectively. As a well-known result, the law of large numbers guarantees that the sample mean converges almost surely to the expectation of the random variables, namely,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}X_1 = \sum_{j=1}^{k} p_j \log a_j \text{ almost surely.}$$

Based upon the theorem, researchers later delve into investigations of a more intricate description of the distribution, e.g., the central limit theorems and the large deviation principles. In this direction, we propose the following analog of Sanov's theorem for tree-indexed Markov chains.

Theorem 1.1 (Sanov). *Let $\{X_g : g \in T\}$ be a Markov chain defined as in Definition 2.1 with the matrix $\mathbf{A}$ satisfying* (A0) *and* (A1)*, and A be a non-negative matrix of the same* *dimension as M such that $A_{a,b} > 0$ if and only if $\mathbf{A}_{a,b} > 0$. Then,*

$$\begin{aligned}
&\sup_{\alpha\in\mathring{S}} -\Lambda_j^*(\alpha) \\
&\leq \liminf_{n\to\infty} \frac{1}{|\Delta_{pn+j}|} \log \mathbb{P}\left(\frac{1}{|\Delta_{pn+j}|} \sum_{g\in\Delta_n\setminus\{\epsilon\}} \log A_{X_g, X_{\varsigma(g)}} \in S \,\middle|\, X_\epsilon = a_0 \right) \\
&\leq \limsup_{n\to\infty} \frac{1}{|\Delta_{pn+j}|} \log \mathbb{P}\left(\frac{1}{|\Delta_{pn+j}|} \sum_{g\in\Delta_n\setminus\{\epsilon\}} \log A_{X_g, X_{\varsigma(g)}} \in S \,\middle|\, X_\epsilon = a_0 \right) \\
&\leq \sup_{\alpha\in\overline{S}} -\Lambda_j^*(\alpha),
\end{aligned} \tag{1}$$

where

$$\begin{aligned}
\Lambda_j^*(\alpha) &= \sup_\mu \mu\alpha - P_j^{(\infty)}\left(d^{-1}, (A\wedge\mu)\odot M\right) \\
&:= \sup_\mu \mu\alpha - \sup_{(\mathsf{p},\mathsf{P})\in Z_j} \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)} | (A\wedge\mu)\odot M)^T \mathsf{p}^{(i+1)}.
\end{aligned}$$

We should note that results regarding large deviations of tree-indexed Markov chains are also considered in [7] for trees generated by a Galton-Watson process for which the probability of producing zero offspring is positive, which does not hold in our case.

The second part of the article is motivated by still another aspect of the tree-indexed Markov chains: the "size" of an outcome space of labeled trees. Such a space of outcomes, also known as Markov hom tree-shift, is a special type of symbolic systems introduced by Aubrun and Beál [1]. Subsequently, Petersen and Salama [14, 15] introduced the topological entropy for such spaces, as an analog of topological entropy for traditional shift spaces, to study their complexity. As it turns out, the topological entropy could also be interpreted as box-counting dimension under a proper metric (see (22)). Inspired by this observation, we aim to investigate the Hausdorff dimension under this specific metric and relate such quantity to the eigenvalues of some nonlinear transfer operators as follows. Drawing inspiration from the classical thermodynamic formalisms, we define the candidates of transfer operator $\mathcal{L}_{\mathbf{A},r} : \mathbb{R}^{\mathcal{A}}_{\geq 0} \to \mathbb{R}^{\mathcal{A}}_{\geq 0}$ for $r = (r_0, r_1, \cdots, r_{p-1}) \in \mathbb{R}^{p-1}_{>0}$ to be

$$\mathcal{L}_{\mathbf{A},r}(x) = \Psi_{\mathbf{A},r_{p-1}} \circ \Psi_{\mathbf{A},r_{p-2}} \circ \cdots \circ \Psi_{\mathbf{A},r_0}(x), \tag{2}$$

where $\Psi_{\mathbf{A},r}(x) = (A^T x)^s$. We then associate the Hausdorff dimension with the eigenvalues of this nonlinear operator, as is seen in the following main theorem of this section.

Theorem 1.2. *Suppose $\mathcal{T}_{\mathbf{A}}$ is a Markov hom tree-shift with an irreducible adjacency matrix $\mathbf{A}$ with period p. Then,*

$$\dim_H \mathcal{T}_{\mathbf{A}} = \min_{r\in\mathcal{R}_{p,d}} \min_{\mathcal{A}_j\in\mathcal{P}(\mathbf{A})} \left(\sum_{\ell=0}^{p-1} \prod_{i=0}^{\ell} r_i^{-1} \right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r}) \tag{3}$$

where

$$\mathcal{R}_{p,d} = \left\{ r\in(0,d]^p : \prod_{i=0}^{p-1} r_i = 1 \right\} \tag{4}$$

and

$$\rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r}) = \sup\{\alpha : \mathcal{L}_{\mathbf{A},r}(u) = \alpha u \in \mathcal{C}_j \setminus \{0\}\}. \tag{5}$$

Notably, if $\mathbf{A}$ is a primitive matrix, then $p = 1$ and $\mathcal{L}_{\mathbf{A},r} = A^T$ for all $r \in \mathcal{R}_{1,d} = \{1\}$, and thus $\dim_H \mathcal{T}_{\mathbf{A}} = \log\rho(\mathbf{A}^T)$ is the spectral radius of $\mathbf{A}$.

At this stage, two main concerns with the theorem arise: the existence of nonnegative eigenvectors and the attainment of $\min_{r\in\mathcal{R}_{p,d}}$. To address the former issue, we postpone the proof of the theorem after a brief introduction on the nonlinear Perron-Frobenius theory (Section 4.1), in which such existence will be resolved. For the latter, it will be discussed in Lemma 4.5, which forms the backbone of the theorem. We should note that the proposed formula of Hausdorff dimension can further be adapted (Proposition 23) to serve as an upper bound for any tree-shifts without assumptions (A0) and (A1), which gives rise to the following corollary.

Corollary 1.3. *Suppose $\mathcal{T}_{\mathbf{A}}$ is a Markov hom tree-shift. Then, $\dim_H \mathcal{T}_{\mathbf{A}} \leq \log \rho(\mathbf{A})$, where $\rho(\mathbf{A})$ denotes the spectral radius of the linear operator $\mathbf{A}$.*

To give the reader an overview of our proof strategy, it is worth mentioning that the upper bound of the Hausdorff dimension is derived via construction of efficient covers of $\mathcal{T}_{\mathbb{A}}$, while the lower bound is obtained by applying the classical mass distribution principle to the "optimal" Markov measure. Hence, the main contributions of this work lie in not only providing a formula for Hausdorff dimension, but also establishing an analog of variational principle for the tree-shifts.

The organization of the paper is as follows. In Section 2, we introduce preliminary definitions and lay down the general conventions of this paper. Section 3 is devoted to the proof of Theorem 1.1, accompanied by an analog of Cramér's theorem (Theorem 3.10) and a law of large numbers (3.11). Finally, Section 4 is devoted to the proposed varitional formula for Hausdorff dimensions (Theorem 1.2), which involves a recapitulation of the nonlinear Perron-Frobenius theory (Section 4.1), necessary lemmas regarding duality (Section 4.2), and lower and upper bounds of Hausdorff dimensions (Section 4.3 and Section 4.4). The section ends with a general upper bound for the Hausdorff dimension by dropping assumptions (A0) and (A1) along with an example illustrating the validity of our formula as an upper bound. Finally, the paper is concluded by a short discussion in Section 5.

2. Preliminaries

Let $\Sigma = \{1, 2, \cdots, d\}$ represent the set of generators for the semigroup

$$T = \cup_{i=0}^{\infty} \Sigma^i := \{\epsilon\} \cup \cup_{i=1}^{\infty} \{g_1 g_2 \cdots g_i : g_j \in \Sigma, \forall 1 \leq j \leq i\},$$

whose Cayley graph is the rooted d-tree (we adopt the convention that $\Sigma^0 = \{\epsilon\}$ to be the set containing the root of the tree). Denote by $\Xi_n = \Sigma^n$ the elements in the semigroup with length n, by $\Delta_n = \cup_{i=0}^{n} \Xi_i$ the initial n-subtree. In addition, let $\varsigma(g) = g_1 g_2 \cdots g_{n-1}$ be the parent of $g = g_1 g_2 \cdots g_n$ and $\sigma(g) = \{gi : i \in \Sigma\}$ be the sets of children of g. In this work, we consider a class of symbolic spaces upon the d-tree called Markov hom tree-shifts, each of which is defined via some adjacency matrix $\mathbf{A}$ as

$$\mathcal{T}_{\mathbf{A}} = \{t \in \mathcal{A}^T : \mathbf{A}_{t_{gi}, t_g} \neq 0 \text{ for } g \in T, i \in \Sigma\}.$$

We will assume throughout that

$$\text{(A0)} \qquad \textstyle\sum_a \mathbf{A}_{a,b} > 0 \text{ for all } b \in \mathcal{A}.$$

Otherwise, one can still find a submatrix $\mathbf{A}'$ of $\mathbf{A}$ such that $\mathcal{T}_{\mathbf{A}} = \mathcal{T}_{\mathbf{A}'}$. For conciseness, for every $k \in \mathbb{N}$ we denote $[k] = \{0, 1, \cdots, k-1\}$. For the scope of this paper, we would like to make the following standing assumption.

(A1) There exists a symbol $a_0 \in \mathcal{A}$ such that for each $a \in \mathcal{A}$ there exists $n_a \in \mathbb{N}$ satisfying $(\mathbf{A}^{n_a})_{a,a_0} > 0$.

It is noted that even though such a_0 might not be unique, the choice of a_0 does not affect the discussions, and thus we will have a_0 fixed throughout. Let $p = \gcd\{n \in \mathbb{N} : (\mathbf{A}^n)_{a_0,a_0} >$

$0\}$. Assumption (A1) naturally induces a collection of sets $\mathcal{P}(\mathbf{A}) = \{\mathcal{A}_j : j \in [p]\}$ of $\mathcal{A}$ such that

$$a \in \mathcal{A}_j \text{ if and only if } (\mathbf{A}^n)_{a,a_0} > 0 \text{ for some } n = j (\mathrm{mod}\ p), \tag{6}$$

where we set $\mathcal{A}_{i+p} = \mathcal{A}_i$ for all $i \in \mathbb{Z}$. A matrix $\mathbf{A}$ is said to be *irreducible (with period p)* if all the elements of $\mathcal{A}$ play the role of a_0 in (A1).

Some frequently used symbols are borrowed from the author's previous work [2]. We let $\Gamma_{\mathcal{A}}$ be the set of all probability vectors indexed by $\mathcal{A}$, and $\Upsilon_{\mathcal{A}}$ be the set of stochastic matrices acting on $\Gamma_{\mathcal{A}}$. Upon $\Gamma_{\mathcal{A}}$ and $\Upsilon_{\mathcal{A}}$, we introduce the *variational distance* d_v^0 and d_V^0, respectively, defined as

$$\mathsf{d}_v^0(\mathsf{p},\mathsf{q}) = \max_{S \subseteq \mathcal{A}} |\sum_{a \in S} \mathsf{p}_a - \mathsf{q}_a| \text{ and } \mathsf{d}_V^0(\mathsf{P},\mathsf{Q}) = \max_{b \in \mathcal{A}} \mathsf{d}_v^0(\mathsf{P}_{a,b},\mathsf{Q}_{a,b}),$$

associated with which is a sup metric $\mathsf{d}_{v,V}^0((\mathsf{p},\mathsf{P}),(\mathsf{q},\mathsf{Q}))$ on the product space $\Gamma_{\mathcal{A}} \times \Upsilon_{\mathcal{A}}$ defined as

$$\mathsf{d}_{v,V}^0((\mathsf{p},\mathsf{P}),(\mathsf{q},\mathsf{Q})) = \max\{\mathsf{d}_v^0(\mathsf{p},\mathsf{q}), \mathsf{d}_V^0(\mathsf{P},\mathsf{Q})\}.$$

For a labeled tree $t \in \mathcal{T}_{\mathbf{A}}$, the *$n$-th level distribution* of t is denoted as

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

where χ_a is the characteristic function of the symbol a. In addition, the *n-th level transition* of t is written as

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

For simplicity, we write the distributions of t from level n to level m ($n \leq m$) by $\tau_{n:m}(t) = (\tau_n(t), \cdots, \tau_m(t))$, associated with which we put

$$D_{n:m}(\mathcal{T}_{\mathbf{A}}) = \{\tau_{n:m}(t) : t \in \mathcal{T}_{\mathbf{A}}\}$$

to be the set of set of all admissible distributions. Likewise, the transitions of t from level n to level m is denoted by $\eta_{n:m}(t) = (\eta_n(t), \cdots, \eta_{m-1}(t))$ and

$$S_{n:m}(\mathcal{T}_{\mathbf{A}}) = \{\eta_{n:m}(t) : t \in \mathcal{T}_{\mathbf{A}}\}$$

is the set of all admissible transitions. Also, let

$$W_{n:m}(\mathcal{T}_{\mathbf{A}}) = \{(\tau_{n:m}(t), \eta_{n:m}(t)) : t \in \mathcal{T}_{\mathbf{A}}\}$$

be the collective admissible sets of distributions and transitions. The set of admissible blocks $B_{n:m}(\mathcal{T}_{\mathbf{A}})$ by prescribed distributions and transitions is defined as

$$B_{n:m}(\mathcal{T}_{\mathbf{A}}; \mathsf{q}, \mathsf{Q}) = \{t|_{\cup_{i=n}^m \Xi_i} : t \in \mathcal{T}_{\mathbf{A}}, \tau_{m:n}(t) = \mathsf{q}, \eta_{n:m}(t) = \mathsf{Q}\},$$

so that $B_{n:m}(\mathcal{T}_{\mathbf{A}}) = \cup_{(\mathsf{q},\mathsf{Q}) \in W_{n:m}} B_{n:m}(\mathcal{T}_{\mathbf{A}}; \mathsf{q}, \mathsf{Q})$. It is noteworthy that a necessary condition for $(\mathsf{q},\mathsf{Q}) \in W_{n:m}(\mathcal{T}_{\mathbf{A}})$ is that $\mathsf{q}^{(i+1)} = \mathsf{Q}^{(i)}\mathsf{q}^{(i)}$ for all $0 \leq i < m-n$. For the convenience sake, we denote by $(\mathsf{p},\mathsf{P}) = (\overleftarrow{\mathsf{q}}, \overleftarrow{\mathsf{Q}})$ the reversed sequence of (q,Q), (i.e., $\mathsf{p}^{(i)} = \mathsf{q}^{(m-n-i)}$ and $\mathsf{P}^{(i)} = \mathsf{Q}^{(m-n-1-i)}$) and consider the following set of all extended reversed sequences:

$$Z := \{(\mathsf{p},\mathsf{P}) \in \Gamma^{\mathbb{Z}_+} \times \Upsilon^{\mathbb{Z}_+} : \mathsf{p}^{(i)} = \mathsf{P}^{(i)}\mathsf{p}^{(i+1)}, \mathsf{P}_{a,b}^{(i)} = 0 \text{ if } \mathbf{A}_{a,b} = 0\}, \tag{7}$$

together with which we assign $\mathcal{C}_j = \{x \in \mathbb{R}_{\geq 0}^{\mathcal{A}} : x_a = 0 \text{ if } a \notin \mathcal{A}_j\}$ and

$$Z_j = \{(\mathsf{p},\mathsf{P}) \in Z : \mathsf{p}^{(i)} \in \mathcal{C}_{j-i}\}.$$

For conciseness, we omit $\mathcal{T}_{\mathbf{A}}$ in these notations should no confusion occur, and suppress "n :" in the notation of $D_{n:m}$, $S_{n:m}$, $W_{n:m}$, and $B_{n:m}$ if $n = 0$. Finally, we would like to manifest our convention of matrices and vectors. In general, matrices are typed in uppercase while vectors are in lowercase, and stochastic matrices and probability vectors are in sans serif font, such as P, Q, p, and q. We denote by $A \odot B$ the Hadamard product of matrices A and B, and by $A \wedge s$ (respectively, $v \wedge s$) the entrywise exponentiation of nonnegative matrix A (respectively, nonnegative vector v) to the power s. In addition, for any $\mathsf{P} \in \Upsilon_{\mathcal{A}}$ and any non-negative matrix A of the same dimension satisfying $\mathsf{P}_{a,b} = 0$ if $A_{a,b} = 0$, we define

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

where $0 \log \frac{0}{0}$ is interpreted as 0. Also, when functions, such as division /, exponential function exp, and logarithm log, are acting on matrices or vectors, the actions are by default taken entrywise unless stated otherwise.

The Markov systems on a tree T can now be defined as follows.

Definition 2.1. *Let $\mathcal{A}$ be a finite alphabet. An* $\mathcal{A}$-valued Markov chains indexed by T *(associated with initial distribution π and with transition matrices M) is a collection of random variables $\{X_g : g \in T\}$ such that*

$$\mathbb{P}(X_{gi}|X_h : h \notin giT) = \mathbb{P}(X_{gi}|X_g) = M_{X_{gi},X_g} \text{ for } g \in T, i \in \Sigma, \tag{8}$$

and that $\mathbb{P}(X_\epsilon = a) = \pi_a$. For convenience, by putting

$$\mathbf{A}_{a,b} = \begin{cases} 1 & \text{if } M_{a,b} > 0; \\ 0 & \text{otherwise,} \end{cases}$$

we assume the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is chosen so that $\Omega = \mathcal{T}_{\mathbf{A}}$, that $\mathcal{F}$ is the Borel σ-algebra, and that $X_g(t) = t_g$ for $t \in \Omega$.

3. Large deviations of Markov chains on d-trees

The aim of this section is to carry out the proof of Theorem 1.1, which states the large deviation principle for the conditional sample mean.

To illustrate our proof strategy, we should present a heuristic proof of the following Cramér's theorem.

Theorem 3.1 (Cramér's theorem)**.** *Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of i.i.d. random variables taking values on a finite subset of $\mathbb{R}$, say $\{a_1, a_2, \cdots, a_n\}$. Then,*

$$\frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i \geq \alpha\right) \to \sup_{t\in\mathbb{R}} (t\alpha - \Lambda(t)) \text{ for } \alpha \geq \mathbb{E}X_1,$$

where

$$\Lambda(t) = \log \mathbb{E}[e^{tX_1}] = \log \sum_{j=1}^k p_j \cdot a_j^t.$$

Precisely, by virtue of Stirling's approximation, one can reformulate the probability in terms of an optimization theory:

$$\begin{aligned} &\frac{1}{n} \log \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n X_i \geq \alpha\right) \\ &= \sup_{\beta\geq\alpha} \left\{ \frac{1}{n} \log \mathbb{P}\left(\sum_{j=1}^k t_j \log a_j = \beta\right) : t_j = \frac{1}{n}\sum_{i=1}^n \delta_{a_j}(X_i) \right\} + O(n^{-1}\log n) \\ &= \sup_{\beta\geq\alpha} \left\{ \sum_{j=1}^k t_j \log\frac{p_j}{t_j} : \sum_{j=1}^k t_j \log a_j = \beta \right\} + o(1), \end{aligned}$$

where the first equality holds since the set

$$L_n = \left\{ (t_j)_{j=1}^k : t_j = \frac{1}{n}\sum_{i=1}^n \delta_{a_j}(X_i) \right\}$$

has its cardinality bounded from the above by $\binom{n}{k-1}$, and the second equality follows from that L_n is asymptotically dense in the simplex of probability vectors. The theorem is ultimately proved by deriving the *rate function* $\sup_{t\in\mathbb{R}} t\alpha - \Lambda(t)$, as a dual problem, via standard techniques in the optimization theory. Inspired by this technique, the author analogously reproduces in [2] the desired approximations and estimations above, based upon which the method of types can be adapted for Theorem 1.1.

To prove the theorem, our starting point would be an observation asserting that the topological space $\Gamma^{\mathbb{Z}_+} \times \Upsilon^{\mathbb{Z}_+}$ is metrizable, and that a natural choice of the metric would be induced by the variational distance on $\Gamma^{\mathbb{Z}_+} \times \Upsilon^{\mathbb{Z}_+}$:

$$\mathsf{d}_{v,V}((\mathsf{p},\mathsf{P}),(\mathsf{q},\mathsf{Q})) := \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \mathsf{d}^0_{v,V}((\mathsf{p}^{(i)},\mathsf{P}^{(i)}),(\mathsf{q}^{(i)},\mathsf{Q}^{(i)})). \tag{9}$$

We note that this metric is compatible with the product topology.

Lemma 3.2. *The Tychonoff topology of* $\Gamma^{\mathbb{Z}_+} \times \Upsilon^{\mathbb{Z}_+}$ *is compatible with* $d_{v,V}$. *In addition, the functions*

$$(\mathsf{p},\mathsf{P}) \mapsto \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)}|M)^T \mathsf{p}^{(i+1)} \text{ and}$$

$$(\mathsf{p},\mathsf{P}) \mapsto \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \mathbb{1}^T(\mathsf{P}^{(i)} \odot \log A)\mathsf{p}^{(i+1)}$$

are uniformly convergent and thus continuous on Z (defined in (7)) with respect to $\mathsf{d}_{v,V}$.

Proof. Since the argument of the metrizability of the product metric space via (9) is standard, it remains to show the uniform convergence and continuity. Indeed,

$$(\mathsf{p},\mathsf{P}) \mapsto \Phi(\mathsf{P}^{(i)}|M)^T \mathsf{p}^{(i+1)} \text{ and } (\mathsf{p},\mathsf{P}) \mapsto \mathbb{1}^T(\mathsf{P}^{(i)} \odot \log A)\mathsf{p}^{(i+1)}$$

are continuous and uniformly bounded, and that the coefficients $(\frac{d-1}{d^{i+1}})_{i\in\mathbb{Z}_+}$ are absolutely summable. From these, the uniform convergence as well as the continuity follow naturally. □

A corollary of the above lemma implies that $\Gamma^{\mathbb{Z}_+} \times \Upsilon^{\mathbb{Z}_+}$ is a compact metrizable space, and hence so is the subspace Z. The continuity and the compactness recur frequently in the rest of our discussion. Our next step is to discuss combinatorial approximations of empirical distribution of patterns, which are manifested in the following series of lemmas.

Lemma 3.3 (Proposition 6, [2])**.** *For any $n \le m$,*

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

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

The lemma essentially indicates that the size of admissible distributions and transitions grows subexponentially with respect to $\Delta_{n:m}$, which is substantially smaller than exponential growth rate of the cardinality of the set $B_{n:m}$. The next two lemmas, on the hand, demonstrates that the set $W_{n:m}$ is, in some sense, asymptotically domain in Z, and the

Stirling's approximations can be applied as in our heuristic proof of Cramér's theorem. Indeed, if we define

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

we have the following lemma.

Lemma 3.4 (Lemma 8, [2]). $\lim_{n\to\infty}\sup_{(\mathsf{p},\mathsf{P})\in W_k'} \mathsf{d}_{v,V}(W_{n:n+k}, (\overleftarrow{\mathsf{p}}, \overleftarrow{\mathsf{P}})) = 0$.

Lemma 3.5 (Proposition 7, [2]). *Let* $(\mathsf{q}(n), \mathsf{Q}(n)) \in W_n$ *be any feasible sequence. Then,*

$$\begin{aligned}\frac{\log|B_n(\mathsf{q},\mathsf{Q})|}{|\Delta_n|} &= \sum_{i=0}^{n-1}\frac{|\Xi_{n-i}|}{|\Delta_n|}\Phi(\mathsf{Q}^{(i)}(n)|\mathbf{A})^T\mathsf{q}^{(i)}(n) + O\left(\frac{n\log|\Delta_n|}{|\Delta_n|}\right)\\ &= \sum_{i=0}^{n-1}\frac{d-1}{d^{n-i}}\Phi(\mathsf{Q}^{(i)}(n)|\mathbf{A})^T\mathsf{q}^{(i)}(n) + O\left(\frac{n\log|\Delta_n|}{|\Delta_n|}\right)\end{aligned}$$

where O *is the big oh notation.*

Proof. A proof can be found in [2] for the first equality, and thus is omitted here. As for the second equality, we note that it follows from that $\frac{d-1}{d^n}/\frac{|\Xi_{n-i}|}{|\Delta_n|} = 1 - d^{-n}$ and that $\Phi(\mathsf{Q}^{(i)}(n)|\mathbf{A})^T\mathsf{q}^{(i)}(n) \le \log|\mathcal{A}|$. □

Now, similar to the classical arguments using method of types, the theory of convex optimization plays a key role. In this work, we resort to the Sion's minimax theorem (see for example [9] for a proof). To apply it, we need to verify a priori the concavity/convexity of the objective function in advance, and it is the moment that the following lemma steps in.

Lemma 3.6. *Suppose* V *is a convex set in a vector space,* $(E^{(i)})_{i=0}^{n-1}$ *are constant matrices in* $\mathbb{R}_{\ge 0}^{k\times k}$, $(q_i(x))_{i=0}^{n-1}$ *are* $\mathbb{R}_{>0}$*-valued affine functions on* V *and* $(\lambda^{(i),N}(x))_{i=0}^{n-1}$ *are* $\mathbb{R}_{\ge 0}^k$*-valued functions on* V *recursively defined by*

$$\lambda^{(i+1)}(x) = q_i(x)\log({E^{(i)}}^T e^{q_i(x)^{-1}\lambda^{(i),N}(x)}) \quad \textit{for } i \in [n].$$

Then, $(\lambda^{(i),N}(x))_{i=0}^{n}$ *are all convex if* $\lambda^{(0)}(x)$ *is convex.*

Proof. We prove the lemma by induction on i of $\lambda^{(i)}$. The case $i = 0$ is included as part of the assumption of the lemma. Now we can suppose the convexity of $\lambda^{(i)}$ and prove that for $\lambda^{(i+1)}$. Indeed, for nonnegative α, α' satisfying $\alpha + \alpha' = 1$, we have

$$\lambda^{(i+1)}(\alpha x + \alpha' x') = \log\left({E^{(i)}}^T e^{\frac{\lambda^{(k)}(\alpha x+\alpha' x')}{q_k(\alpha x+\alpha' x')}}\right)^{q_k(\alpha x+\alpha' x')}$$
$$= \log\left({E^{(i)}}^T e^{\frac{\lambda^{(k)}(\alpha x+\alpha' x')}{\alpha q_k(x)+\alpha' q_k(x')}}\right)^{\alpha q_k(x)+\alpha' q_k(x')}$$
$$\le \log\left({E^{(i)}}^T e^{\frac{\alpha q_k(x)}{\alpha q_k(x)+\alpha' q_k(x')}\frac{\lambda^{(k)}(x)}{q_k(x)} + \frac{\alpha' q_k(x')}{\alpha q_k(x)+\alpha' q_k(x')}\frac{\lambda^{(k)}(x')}{q_k(x')}}\right)^{\alpha q_k(x)+\alpha' q_k(x')}$$
$$\le \log\left({E^{(i)}}^T e^{\frac{\lambda^{(k)}(x)}{q_k(x)}}\right)^{\alpha q_k(x)} + \log\left({E^{(i)}}^T e^{\frac{\lambda^{(k)}(x')}{q_k(x')}}\right)^{\alpha' q_k(x')}$$
$$= \alpha\lambda^{(k+1)}(x) + \alpha'\lambda^{(k+1)}(x'),$$

where the first inequality is the induction hypothesis in convexity, while the second follows from the Hölder's inequality. The theorem is then proved by induction. □

Lemma 3.7. *Suppose A and M are nonnegative matrices of the same dimension as $\mathbf{A}$ such that*

$$A_{a,b}, M_{a,b} > 0 \text{ if and only if } \mathbf{A}_{a,b} > 0$$

and that $\mathbf{A}$ satisfies assumption (A1). Let $E = M \odot (A \wedge \mu)$ ($\mu \in \mathbb{R}$) and define a sequence $\lambda^{(0)} = 0$ and

$$\lambda^{(i+1)} = \frac{d-1}{d^{i+1}} \log\left(E^T e^{\frac{d^{i+1}}{d-1}\lambda_b^{(i)}}\right) \qquad \text{for } i \in \mathbb{N}. \tag{10}$$

For every $n \in \mathbb{N} \cup \{\infty\}$ and $\alpha \in \mathbb{R}$, denote by $F_{n,j}(\alpha)$ and $\Omega_{n,j}$ the maximum and the feasible domain, respectively, of the optimization problem

$$\begin{cases} \text{maximize} & \sum_{i=0}^{n-1} \frac{d-1}{d^{i+1}} \mathbb{1}^T \Phi(\mathsf{P}^{(i)}|M)^T \mathsf{p}^{(i+1)} \\ \text{subject to} & \sum_{i=0}^{n-1} \frac{d-1}{d^{i+1}} \mathbb{1}^T (\mathsf{P}^{(i)} \odot \log A)\mathsf{p}^{(i+1)} = \alpha, (\mathsf{p}, \mathsf{P}) \in Z_j \end{cases} \tag{11}$$

Then,

1. *The functions $\lambda_a^{(n)}$ can be expressed as*

$$\lambda_a^{(n)} = \max_{\substack{\mathsf{p},\mathsf{P}: \\ \mathsf{p}^{(n)} = e_a}} \sum_{i=0}^{n-1} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)}|E)\mathsf{p}^{(i+1)} \quad (i \geq 0, a \in \mathcal{A})$$

and are convex in μ. Hence, $P_j^{(\infty)}(d^{-1}, E) = \lim_{n\to\infty} \sup_{a \in \mathcal{A}_{n+p-j}} \lambda_a^{(n)}$ is also convex in μ.

2. *The strong duality holds for the optimization for all $n \in \mathbb{N} \cup \{\infty\}$. Moreover,*

$$F_{n,j}(\alpha) = \begin{cases} \inf_\mu -\mu\alpha + \sup_{\mathsf{p}^{(n)} \in \mathcal{C}_{j-n}} \lambda^{(n)^T} \mathsf{p}^{(n)} & \text{if } n \in \mathbb{N}; \\ \inf_\mu -\mu\alpha + P_j^{(\infty)}(d^{-1}, E) & \text{if } n = \infty. \end{cases}$$

3. *The function $F_\infty(\alpha)$ is concave, upper semicontinuous, and not identically $-\infty$.*

Proof. The minimax theorem forms the backbone of the proof.

(1) The equality is proved in Ban-Wu, while the convexity is secured according to Lemma 3.6.

(2) We first show the case $n \in \mathbb{N}$. Recall that

$$Z_j = \{(\mathsf{p}, \mathsf{P}) \in \Gamma^{\mathbb{Z}_+} \times \Upsilon^{\mathbb{Z}_+} : \mathsf{p}^{(i)} = \mathsf{P}^{(i)}\mathsf{p}^{(i+1)}, \mathsf{p}^{(0)} \in \mathcal{C}_j\},$$

and $\mathsf{p}^{(0)} \in \mathcal{C}_{j-n}$ if and only if $\mathsf{p}^{(n)} \in \mathcal{C}_j$ due to (6) under the assumption of (A1). Note that

$$\sup_{(\mathsf{p},\mathsf{P}) \in Z_j} \left\{ \sum_{i=0}^{n-1} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)}|M)^T \mathsf{p}^{(i+1)} : \sum_{i=0}^{n-1} \frac{d-1}{d^{i+1}} \mathbb{1}^T (\mathsf{P}^{(i)} \odot \log A)\mathsf{p}^{(i+1)} = \beta \right\}$$
$$= \sup_{\mathsf{p}^{(n)} \in \mathcal{C}_{j-n}} \sup_{\mathsf{P}^{(0:n-1)}} \inf_\mu -\mu\beta + \sum_{i=0}^{n-1} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)}|(A \wedge \mu) \odot M)^T \left(\prod_{\ell=i+1}^{n-1} \mathsf{P}^{(\ell)}\right) \mathsf{p}^{(n)}.$$

By fixing $\mathsf{P}^{(1:n-1)}$ and $\mathsf{p}^{(n)}$, we may apply the minimax theorem to swap "$\sup_{\mathsf{P}^{(0)}}$" and "$\inf_\mu$" and simplify the expression:

$$\begin{aligned}
&\sup_{\mathsf{p}^{(n)}\in\mathcal{C}_{j-n}} \sup_{\mathsf{P}^{(0:n-1)}} \inf_\mu -\mu\beta + \sum_{i=0}^{n-1} \frac{d-1}{d^{i+1}} \mathbb{1}^T \Phi(\mathsf{P}^{(i)}|E)^T \left(\prod_{\ell=i+1}^{n-1} \mathsf{P}^{(\ell)}\right) \mathsf{p}^{(n)} \\
=& \sup_{\mathsf{p}^{(n)}\in\mathcal{C}_{j-n}} \sup_{\mathsf{P}^{(1:n-1)}} \inf_\mu -\mu\beta + \lambda^{(1)^T} \left(\prod_{\ell=1}^{n-1} \mathsf{P}^{(\ell)}\right) \mathsf{p}^{(n)} \\
&\qquad + \sum_{i=1}^{n-1} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)}|E)^T \left(\prod_{j=i+1}^{n-1} \mathsf{P}^{(\ell)}\right) \mathsf{p}^{(n)}.
\end{aligned}$$

We then continue, by applying the minimax theorem recursively, to swap $\sup_{\mathsf{P}^{(i)}}$ and $\inf_\mu$ as before. Finally, we apply (1) and the minimax theorem to swap $\sup_{\mathsf{p}^{(n)}\in\mathcal{C}_0}$ and $\inf_\mu$. Combining these yields the the case $n \neq \infty$. As for $n = \infty$, we let $\alpha_1 = \min\{\alpha : F_{\infty,j}(\alpha) > -\infty\}$ and that $\alpha_2 = \max\{\alpha : F_{\infty,j}(\alpha) > -\infty\}$, which are well-defined and finite by compactness and continuity (Lemma 3.2). The goal of this lemma is to verify the following expression of $F_{\infty,j}(\alpha)$:

$$\begin{aligned}
&\sup_{(\mathsf{p},\mathsf{P})\in Z_j} \left\{ \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)}|M)^T \mathsf{p}^{(i+1)} : \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \mathbb{1}^T (\mathsf{P}^{(i)} \odot \log A) \mathsf{p}^{(i+1)} = \alpha \right\} \\
&= \inf_\mu \sup_{(\mathsf{p},\mathsf{P})\in Z_j} -\mu\alpha + \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)}|(A\wedge\mu)\odot M) \left(\prod_{j=i+1}^{n-1} \mathsf{P}^{(j)}\right) \mathsf{p}^{(n)} \\
&=: \inf_\mu -\mu\alpha + P_j^{(\infty)}(d^{-1}, E)
\end{aligned}$$

In fact, it suffices to show the "$\leq$" part, since the other is the so-called weak duality and holds automatically. To show this, we make use of the following readily checked identity: Suppose $\beta \in \mathbb{R}^{\mathcal{A}}_{>0}$, then

$$\max_{\mathsf{q}\in\Gamma_{\mathcal{A}}} \left| \sum_{a\in\mathcal{A}} \mathsf{q}_a \log \frac{\beta_a}{\mathsf{q}_a} \right| \leq \max\left\{ \max_a |\log\beta_a|, \left|\log \sum_{a\in\mathcal{A}} \beta_a\right| \right\} \leq \log|\mathcal{A}| + \max_a |\log\beta_a|.$$

Hence, under the assumption that $M_{a,b}, A_{a,b} > 0$ whenever $\mathbf{A}_{a,b} > 0$, we may set

$$C = \max\left\{ \max_{a,b:\mathbf{A}_{a,b}} |\log A_{a,b}|, \max_{a,b:\mathbf{A}_{a,b}} |\log M_{a,b}|, \log|\mathcal{A}| \right\},$$

so that

$$|\Phi(\mathsf{P}^{(i)}|(A\wedge\mu)\odot M)| \leq C\cdot(\mu+2).$$

As a consequence,

$$|P^{(\infty)}(d^{-1}, E) - \sup_{\mathsf{p}^{(n)}\in\mathcal{C}_{n-j}} \lambda^{(n)^T} \mathsf{p}^{(n)}| \leq Cd^{-n}(\mu+2) \quad \text{for all } n \in \mathbb{N}.$$

Now if $(\mathsf{p}^*, \mathsf{P}^*)$ is a maximizer of $\sup_{(\mathsf{p},\mathsf{P})\in Z_j} \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)}|(A\wedge\mu)\odot M) \left(\prod_{j=i+1}^{n-1} \mathsf{P}^{(j)}\right) \mathsf{p}^{(n)}$, then by putting

$$\epsilon_n = \sum_{i=0}^{n-1} \frac{d-1}{d^{i+1}} \mathbb{1}^T (\mathsf{P}^{(i)^*} \odot \log A) \mathsf{p}^{(i+1)^*} - \alpha \quad (\to 0 \text{ as } n\to\infty),$$

we have, for $\alpha \in (\alpha_1, \alpha_2)$, that

$$\begin{aligned} F_{\infty,j}(\alpha) &\leq \liminf_{n\to\infty} F_{n,j}(\alpha+\epsilon_n) + 2Cd^{-n} \\ &\leq \liminf_{n\to\infty} \inf_{\mu} -\mu(\alpha+\epsilon_n) + P_j^{(\infty)}(d^{-1}, E) + Cd^{-n}\mu \\ &\leq \liminf_{n\to\infty} \inf_{\mu} -\mu(\alpha+\epsilon_n - Cd^{-n}) + P_j^{(\infty)}(d^{-1}, E) \\ &\leq \inf_{\mu} -\mu\alpha + P_j^{(\infty)}(d^{-1}, E), \end{aligned}$$

where the last inequality follows from the continuity at $\alpha \in (\alpha_1, \alpha_2)$, resulting from the convexity of the conjugate function. As for the boundary points α_1 (similar for α_2), we note that by compactness (Lemma 3.2) and affinity of the objective function, there exists $(\mathsf{q}, \mathsf{Q}) \in Z_0$ such that $\mathsf{q}^{(i)} = e_{b_{i+1}} = \mathsf{Q}^{(i)}_{b_i, b_{i+1}} = 1$ for some sequence $(b_i)_{i\geq 0} \in \mathcal{A}^{\mathbb{Z}_+}$ so that

$$\alpha_1 = \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \log A_{b_i,b_{i+1}}, \text{ and}$$

$$P_j^{(\infty)}(d^{-1}, E) = \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \left[\log M_{b_i,b_{i+1}} + \mu \log A_{b_i,b_{i+1}}\right].$$

Thus,

$$F_{\infty,j}(\alpha_1) = \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \log M_{b_i,b_{i+1}} = \inf_{\mu} -\mu\alpha_1 + P_j^{(\infty)}(d^{-1}, E),$$

and the lemma is proved.

(3) The concavity and the upper semimcontinuity are consequences of (2), while the compactness (Lemma 3.2) implies that $F_{\infty,j}(\mathcal{A})$ is not identically $-\infty$. $\square$

Lemma 3.8. *Suppose the assumption* (A1) *is satisfied. Then, there exists* $n \in \mathbb{N}$ *such that* $(\mathbf{A})^{pn}_{a,a_0} > 0$ *for every* $a \in \mathcal{A}_0$*. In addition, for every* $\mathsf{q} \in \mathcal{C}_0 \cap \Gamma_{\mathcal{A}}$ *there exists* $(\mathsf{p}, \mathsf{P}) \in Z_0$ *such that* $\mathsf{p}^{(pn)} = \mathsf{e}_{a_0}$ *and that* $\mathsf{p}^{(0)} = \mathsf{q}$.

Proof. The existence of n satisfying $(\mathbf{A})^{pn}_{a,a_0} > 0$ follows straightforwardly from (A1). To see the additional properties are also satisfied, we first construct for every $a \in \mathcal{A}_0$ a sequence $(b^{a,0})_{i\in\mathbb{Z}_+}$ satisfying that $b^{a,0} = a$, that $b^{a,pn} = a_0$, and that $\mathbf{A}_{b^{a,i+1}, b^{a,i}} = 1$ for all $i \in \mathbb{Z}_+$. In addition, we may assume that for every $i = 0, 1, \cdots, pn$, $b^{a,i} = b^{a',i}$ implies $(b^{a,j})_{j\geq i} = (b^{a',j})_{j\geq i}$. Indeed, if $b^{a,1} = b^{a',1}$ but $(b^{a,j})_{j\geq 1} \neq (b^{a',j})_{j\geq 1}$, then we may replace our choice of $(b^{a',j})_{j\geq 1}$ by $(b^{a,j})_{j\geq 1}$ and repeat the process until the property is satisfied for for all $b^{a,1}$, $a \in \mathcal{A}$. Then, we can argue in a similar manner so that the property is satisfied for all $b^{a,i}$, $i = 1, \cdots, pn-1$ and $a \in \mathcal{A}$. Then, the lemma is proved by choosing

$$\mathsf{p}^{(0)} = \mathsf{q} \text{ and } \mathsf{p}^{(i)}_b = \sum_{a: b^{a,i-1} = b} \mathsf{p}^{(i+1)}_{b^{a,i-1}} \quad (i \in \mathbb{Z}_+)$$

and

$$\mathsf{P}^{(i)}_{b',b} = \begin{cases} \frac{\mathsf{p}^{(i)}_{b'}}{\mathsf{p}^{(i+1)}_b} & \text{if } \mathsf{p}^{(i+1)}_b > 0; \\ \frac{\mathbf{A}_{b',b}}{\sum_c \mathbf{A}_{c,b}} & \text{otherwise.} \end{cases} \quad (i \in \mathbb{Z}_+)$$

$\square$

Proof of Theorem 1.1. We borrow ideas from Ban and Wu [2]. For conciseness, we prove only the case $j = 0$.

Similar to the heuristic proof of Theorem 3.1, we first infer from Lemma 3.3 and Lemma 3.5 that

$$\begin{aligned}
&\frac{1}{|\Delta_{pn}|}\log\mathbb{P}\left(\frac{1}{|\Delta_{pn}|}\sum_{g\in\Delta_{pn}\setminus\{\epsilon\}}\log A_{X_g,X_{\varsigma(g)}}\in S\,\middle|\,X_\epsilon=a_0\right)\\
=&\sup_{\beta\in S}\Bigg\{\frac{1}{|\Delta_{pn}|}\log\mathbb{P}\left(\sum_{i=0}^{pn-1}\frac{|\Xi_{pn-i}|}{|\Delta_{pn}|}\mathbb{1}^T(\mathsf{P}^{(i)}\odot\log A)\mathsf{p}^{(i+1)}=\beta\right):\mathsf{p}^{(pn)}=\mathsf{e}_{a_0},\\
&\qquad\mathsf{p}^{(i)}=\tau_{pn-i}(X),\mathsf{P}^{(i)}=\eta_{pn-i}(X)\Bigg\}+O\left((pn+1)\frac{\log|\Delta_{pn}|}{|\Delta_{pn}|}\right)\\
=&\sup_{\beta\in S}\Bigg\{\sum_{i=0}^{pn-1}\frac{|\Xi_{pn-i}|}{|\Delta_{pn}|}\Phi(\mathsf{P}^{(i)}|M)^T\mathsf{p}^{(i+1)}:\mathsf{p}^{(pn)}=\mathsf{e}_{a_0},\mathsf{p}^{(i)}=\tau_{pn-i}(X),\mathsf{P}^{(i)}=\eta_{pn-i}(X)\\
&\qquad\sum_{i=0}^{pn-1}\frac{|\Xi_{pn-i}|}{|\Delta_{pn}|}\mathbb{1}^T(\mathsf{P}^{(i)}\odot\log A)\mathsf{p}^{(i+1)}=\beta\Bigg\}+O\left((pn+1)\frac{\log|\Delta_{pn}|}{|\Delta_{pn}|}\right)
\end{aligned}$$

To prove the upper bound, we may choose $(\overleftarrow{\mathsf{p}}(pn_\ell),\overleftarrow{\mathsf{P}}(pn_\ell))\in W_{pn_\ell}$ such that

$$\sum_{i=0}^{pn_\ell-1}\frac{|\Xi_{pn_\ell-i}|}{|\Delta_{pn_\ell}|}\mathbb{1}^T(\mathsf{P}^{(i)}(pn_\ell)\odot\log A)\mathsf{p}^{(i+1)}(pn_\ell)\in S$$

and that

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{|\Delta_{pn}|}\log\mathbb{P}\left(\frac{1}{|\Delta_{pn}|}\sum_{g\in\Delta_n\setminus\{\epsilon\}}\log A_{X_g,X_{\varsigma(g)}}\in S\,\middle|\,X_\epsilon=a_0\right)\\
\le&\lim_{\ell\to\infty}\sum_{i=0}^{pn_\ell-1}\frac{|\Xi_{pn_\ell-i}|}{|\Delta_{pn_\ell}|}\Phi(\mathsf{P}^{(i)}(pn_\ell)|M)^T\mathsf{p}^{(i+1)}(pn_\ell)
\end{aligned}\tag{12}$$

Without loss of generality, we may in addition assume $(\mathsf{p}^{(i)}(pn_\ell),\mathsf{P}^{(i)}(pn_\ell))$ converges to a $(\mathsf{p}^{(i)^*},\mathsf{P}^{(i)^*})\in\Gamma_{\mathcal{A}}\times\Upsilon_{\mathcal{A}}$ due to compactness. Under the circumstances, it is readily checked that $(\mathsf{p}^*,\mathsf{P}^*)\in Z_0$ by assumption (A1), that

$$\sum_{i=0}^{\infty}\frac{d-1}{d^{i+1}}\mathbb{1}^T(\mathsf{P}^{(i)^*}\odot\log A)\mathsf{p}^{(i+1)^*}\in\overline{S}\tag{13}$$

and that

$$\lim_{\ell\to\infty}\sum_{i=0}^{pn_\ell-1}\frac{|\Xi_{pn_\ell-i}|}{|\Delta_{pn_\ell}|}\Phi(\mathsf{P}^{(i)}|M)^T\mathsf{p}^{(i+1)}=\sum_{i=0}^{\infty}\frac{d-1}{d^{i+1}}\Phi(\mathsf{P}^{(i)^*}|M)^T\mathsf{p}^{(i+1)^*}.\tag{14}$$

In particular, according to Lemma 3.7, the relations (12)(13)(14) imply

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{|\Delta_{pn}|}\log\mathbb{P}\left(\frac{1}{|\Delta_{pn}|}\sum_{g\in\Delta_n\setminus\{\epsilon\}}\log A_{X_g,X_{\varsigma(g)}}\in S\,\middle|\,X_\epsilon=a_0\right)\\
=&\sum_{i=0}^{\infty}\frac{d-1}{d^{i+1}}\Phi(\mathsf{P}^{(i)^*}|M)^T\mathsf{p}^{(i+1)^*}\le\sup_{\beta\in\overline{S}}-\Lambda_0^*(\beta).
\end{aligned}$$

To prove the lower bound, we note that for every $\epsilon > 0$ we may find $(\mathsf{p}^*, \mathsf{P}^*) \in Z_0$ such that

$$\sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \mathbb{1}^T (\mathsf{P}^{(i)^*} \odot \log A) \mathsf{p}^{(i+1)^*} \in \mathring{S} \tag{15}$$

and that

$$\sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)^*} | M)^T \mathsf{p}^{(i+1)^*} > \sup_{\beta \in \mathring{S}} -\Lambda_0^*(\beta) - \epsilon. \tag{16}$$

Due to assumption (A1), we may further assume that there exists an n_0 such that

$$\mathsf{p}^{(n_0)} = \mathsf{e}_{a_0}. \tag{17}$$

Otherwise, according to Lemma 3.8, for all $0 < \delta$, we can always find $(\mathsf{p}', \mathsf{P}') \in Z_0$ such that $\mathsf{p}^{(i)'} = \mathsf{p}^{(i)^*}$ for $i = 0, \cdots, \lceil -\log_d \delta \rceil$ with $\mathsf{p}^{(-\lceil \log_d \frac{\epsilon}{2|\mathcal{A}|} \rceil + pn_1)} = \mathsf{e}_{a_0}$, that $\mathsf{P}^{(i)'} = \mathsf{P}^{(i)^*}$ for $i = 0, \cdots, \lceil -\log_d \delta \rceil - 1$, that

$$\left| \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \mathbb{1}^T (\mathsf{P}^{(i)^*} \odot \log A) \mathsf{p}^{(i+1)^*} - \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \mathbb{1}^T (\mathsf{P}^{(i)'} \odot \log A) \mathsf{p}^{(i+1)'} \right| \le \delta \max_{a,b: \mathbf{A}_{a,b}=1} |\log A_{a,b}|,$$

and that

$$\sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)'} | M)^T \mathsf{p}^{(i+1)'} > \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)^*} | M)^T \mathsf{p}^{(i+1)^*} - \delta \max_{a,b: \mathbf{A}_{a,b}=1} |\log M_{a,b}|$$

In other words, we may replace (p, P) by $(\mathsf{p}', \mathsf{P}')$ and the very same properties are still satisfied by choosing proper δ. Under the assumption of (17) and (15), we may deduce from Lemma (3.4) that for all sufficiently large n, there exists $(\overleftarrow{\mathsf{p}}(pn), \overleftarrow{\mathsf{P}}(pn)) \in W_{pn}$ such that

$$\sum_{i=0}^{pn-1} \frac{|\Xi_{pn-i}|}{|\Delta_{pn}|} \mathbb{1}^T (\mathsf{P}^{(i)}(pn) \odot \log A) \mathsf{p}^{(i+1)}(pn) \in \mathring{S} \tag{18}$$

and that

$$\sum_{i=0}^{pn-1} \frac{|\Xi_{pn-i}|}{|\Delta_{pn}|} \Phi(\mathsf{P}^{(i)}(pn) | M)^T \mathsf{p}^{(i+1)}(pn) \ge \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)^*} | M)^T \mathsf{p}^{(i+1)^*} - \epsilon \tag{19}$$

Hence, combining (16), (18) and (19) we prove the lower bound. □

Remark 3.9. *The optimization problem* (11) *in Lemma 3.7 can further be adapted as*

$$\begin{cases} \textit{maximize} & \sum_{i=0}^{n-1} \frac{d-1}{d^{i+1}} \mathbb{1}^T \Phi(\mathsf{P}^{(i)} | M)^T \mathsf{p}^{(i+1)} \\ \textit{subject to} & \sum_{i=0}^{n-1} \frac{d-1}{d^{i+1}} \mathbb{1}^T (\mathsf{P}^{(i)} \odot \log A^{[a,b]}) \mathsf{p}^{(i+1)} = \alpha_{a,b}, \forall a, b \textit{ satisfying } \mathbf{A}_{a,b} = 1 \\ & (\mathsf{p}, \mathsf{P}) \in Z_j \end{cases}$$

where

$$\log A_{a',b'}^{[a,b]} = \begin{cases} 1 & \textit{if } (a', b') = (a, b); \\ -\infty & \textit{if } A_{a',b'}^{[a,b]} = 0; \\ 0 & \textit{otherwise.} \end{cases}$$

For the adapted problem, we may write $E = (\odot_{a,b:\mathbf{A}_{a,b}=1} (A^{[a,b]}) \wedge \mu^{[a,b]}) \odot M$, *and so its dual solution becomes*

$$\Lambda_j^{\dagger}(\alpha) = \inf_{\mu} \sup_{(\mathsf{p}, \mathsf{P}) \in Z_j} -\mu^T \alpha + \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \Phi(\mathsf{P}^{(i)} | E) \left(\prod_{j=i+1}^{n-1} \mathsf{P}^{(j)} \right) \mathsf{p}^{(n)},$$

for which a proof is almost identical, and it is in this form that we see the theorem is indeed an analog of Sanov's theorem. Under the circumstances, Theorem 1.1 immediately follows with respect to the rate function $\Lambda_j^{\dagger}(\alpha)$, where the interior and the closure are interpreted as in the relative sense with respect to the probability simplex.

Corollary 3.10 (Cramér)**.** *Suppose A and M satisfy the assumption of Theorem 1.1. Then,*

$$\lim_{\epsilon\to 0^+}\lim_{n\to\infty}\frac{1}{|\Delta_{pn+j}|}\log\mathbb{P}\left(\frac{1}{|\Delta_{pn+j}|}\sum_{g\in\Delta_{pn+j}\setminus\{\epsilon\}}\log A_{X_g,X_{\varsigma(g)}}\in(\alpha-\epsilon,\alpha+\epsilon)\,\middle|\,X_\epsilon=a_0\right)=-\Lambda_j^*(\alpha).$$

Proof. The corollary is an consequence of Theorem 1.1. More precisely, the corollary follows by letting $S=(\alpha-\epsilon,\alpha+\epsilon)$ and exploiting the concavity and upper semicontinuity of $-\Lambda_j^*(\alpha)$. □

Theorem 3.11. *Suppose A and M satisfy the assumption of Theorem 1.1 and M is irreducible. Then, conditioned on $X_\epsilon=a_0$,*

$$\begin{aligned}&\lim_{n\to\infty}\left(\frac{1}{|\Delta_{pn+j}|}\sum_{g\in\Delta_{pn+j}\setminus\{\epsilon\}}\log A_{X_g,X_{\varsigma(g)}}\,\middle|\,X_\epsilon=a_0\right)\\&=\sum_{i=0}^{p-1}\frac{d^{-i}}{\sum_{\ell=0}^{p-1}d^{-\ell}}\sum_{a,b}\pi_b^{(i-j)}M_{a,b}\log A_{a,b},\end{aligned}$$

where $\pi^{(j)}=M\pi^{(j+1)}$ are eigenvectors of M^p with $\pi^{(0)}\in\mathcal{C}_0$.

Proof of Theorem 3.11. For simplicity, we only give the proof of the case $j=0$.

We first show that for the optimization problem

$$\begin{cases}\text{maximize} & \sum_{i=0}^{\infty}\frac{d-1}{d^{i+1}}\mathbb{1}^T(\mathsf{P}^{(i)}\odot\log\frac{M}{\mathsf{P}^{(i)}})\mathsf{p}^{(i+1)}\\ \text{subject to} & (\mathsf{p},\mathsf{P})\in Z_0\end{cases}$$

any maximizer (p,P) satisfies

$$\sum_{i=0}^{\infty}\frac{d-1}{d^{i+1}}\mathbb{1}^T(\mathsf{P}\odot\log A)\mathsf{p}^{(i+1)}=\sum_{i=0}^{\infty}\frac{d-1}{d^{i+1}}\mathbb{1}^T(M\odot\log A)\pi^{(i+1)}.\tag{20}$$

To begin with, we note that

$$(\mathsf{p},\mathsf{P})=((\pi^{(0)},\pi^{(1)},\cdots,\pi^{(p-1)})^{\mathbb{Z}_+},M^{\mathbb{Z}_+})\in Z_j$$

is a maximizer. Indeed, for every $(\mathsf{p},\mathsf{P})\in Z_0$,

$$\sum_{a\in\mathcal{A}}\mathsf{P}_{a,b}^{(i)}\odot\log\frac{M_{a,b}}{\mathsf{P}_{a,b}^{(i)}}\le 0\quad\text{and}\quad\sum_{a\in\mathcal{A}}\mathsf{P}_{a,b}^{(i)}\odot\log\frac{M_{a,b}}{\mathsf{P}_{a,b}^{(i)}}=0\text{ iff }\mathsf{P}_{ab}^{(i)}=M_{ab}\text{ for all }a,\tag{21}$$

from which the maximality follows. As for property (20), we apply (21) to derive that for any optimal solution (p,P),

$$\mathsf{p}^{(pN+i)}=M^p\mathsf{p}^{(p(N+1)+i)}\text{ for all }i\ge 0,n\ge 0.$$

A corollary of Perron-Frobenius (for primitive matrices) asserts that

$$\mathsf{p}^{(pN+i)}=\lim_{n\to\infty}M^n\mathsf{p}^{(pN+i+pn)}=\pi^{(i)}.$$

Since $\mathsf{p}_a^{(pn+i)}=\pi_a^{(i)}$ is positive if and only if $a\in\mathcal{A}_{-i-1}$ and $n,i\ge 0$, once again we can apply (21) to derive $\mathsf{P}_{a,b}^{(pn+i)}=M_{a,b}$ if $\mathcal{A}_{-i}\times\mathcal{A}_{-i-1}$. Hence, $\Lambda_0^*(\alpha)$ admits a unique maximum point at

$$\alpha = \alpha^* := \sum_{i=0}^{p-1} \frac{d^{-i}}{\sum_{\ell=0}^{p-1} d^{-\ell}} \sum_{i=0}^{p-1} \mathbb{1}^T (M \odot \log A) v^{(i+1)},$$

at which Λ_0^* reaches its maximum 0.

The rest of the proof is merely an application of Borel-Cantelli lemma. To be more precise, it suffices to show that for all $\epsilon > 0$,

$$\mathbb{P}\left(\limsup_{n\to\infty} \left\{ \left| \frac{1}{|\Delta_{pn}|} \sum_{g \in \Delta_{pn} \setminus \{\epsilon\}} \log A_{X_g, X_{\varsigma(g)}} - \alpha^* \right| > \epsilon \right\} \middle| X_\epsilon = a_0 \right) = 0.$$

Now that $\delta(\epsilon) = \sup_{|\alpha - \alpha^*| \geq \epsilon} -\Lambda_0^*(\alpha) < 0$ for all $\epsilon > 0$ due to the uniqueness of α^*. This implies the existence of $N \in \mathbb{N}$ such that for all $n \geq N$,

$$\mathbb{P}\left(\left\{ \left| \frac{1}{|\Delta_{pn}|} \sum_{g \in \Delta_{pn} \setminus \{\epsilon\}} \log A_{X_g, X_{\varsigma(g)}} - \alpha^* \right| > \epsilon \right\} \middle| X_\epsilon = a_0 \right) \leq e^{\delta(\epsilon)|\Delta_{pn}|}$$

Consequently,

$$\sum_{i=N}^{\infty} \mathbb{P}\left(\left\{ \left| \frac{1}{|\Delta_{pn}|} \sum_{g \in \Delta_{pn} \setminus \{\epsilon\}} \log A_{X_g, X_{\varsigma(g)}} - \alpha^* \right| > \epsilon \right\} \middle| X_\epsilon = a_0 \right) \leq \sum_{i=N}^{\infty} e^{\delta(\epsilon)|\Delta_{pn}|} < \infty,$$

and our hypothesis holds as a consequence of Borel-Cantelli lemma. □

Example 3.12. *In this example, let $d = 2$, A and M be matrices defined by*

$$M = \begin{bmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{bmatrix}, A = \begin{bmatrix} 1 & 2 \\ 1 & 0 \end{bmatrix}.$$

We note that M is primitive, and thus in this case Theorem 1.1 states that

$$\frac{1}{|\Delta_n|} \sum_{g \in \Delta_n \setminus \{\epsilon\}} \log A_{X_g, X_{\varsigma(g)}} \to \sum_{a,b} \pi_b M_{a,b} \log A_{a,b} \text{ almost surely.}$$

and the distribution satisfies the large deviation principle given in Theorem 1.1. Numerically, $-\Lambda_0^*(\alpha)$ can be derived as Figure 1. We note that $-\Lambda_0^*(\alpha)$ is finite if and only if $\alpha \in [0, \frac{2}{3}\log 2]$, where $\frac{2}{3}\log 2 \approx 0.4621$. Furthermore, the maximum point is $\alpha = \frac{1}{3}\log 2 \approx 0.2310$.

Example 3.13. *We provide here an example of irreducible M in order to (a) demonstrate the* lim sup *and* lim inf *of the conditional sample mean might differ, and (b) compare the conditional sample, unconditional sample, and the expectation of the sample mean. We consider $d = 2$ and*

$$A = M = \begin{bmatrix} 0 & 1 & 1 \\ \frac{1}{3} & 0 & 0 \\ \frac{2}{3} & 0 & 0 \end{bmatrix}, \pi = \begin{bmatrix} \frac{1}{6} \\ \frac{1}{3} \\ \frac{1}{2} \end{bmatrix}.$$

for which the sample mean conditioned associated with $\mathcal{A}_0 := \{0\}$ and on $\mathcal{A}_1 := \{1, 2\}$ are plotted in Figure 2, respectively. To see (a), if we denote by α_0^* and by α_1^* the maximum points of λ_0^* and λ_1^*, respectively, we infer from Theorem (1.1) that almost surely

$$\alpha^- := \liminf_{n\to\infty} \left(\frac{1}{|\Delta_n|} \sum_{g \in \Delta_n \setminus \{\epsilon\}} \log A_{X_g, X_{\varsigma(g)}} \right) = \alpha_1^*$$

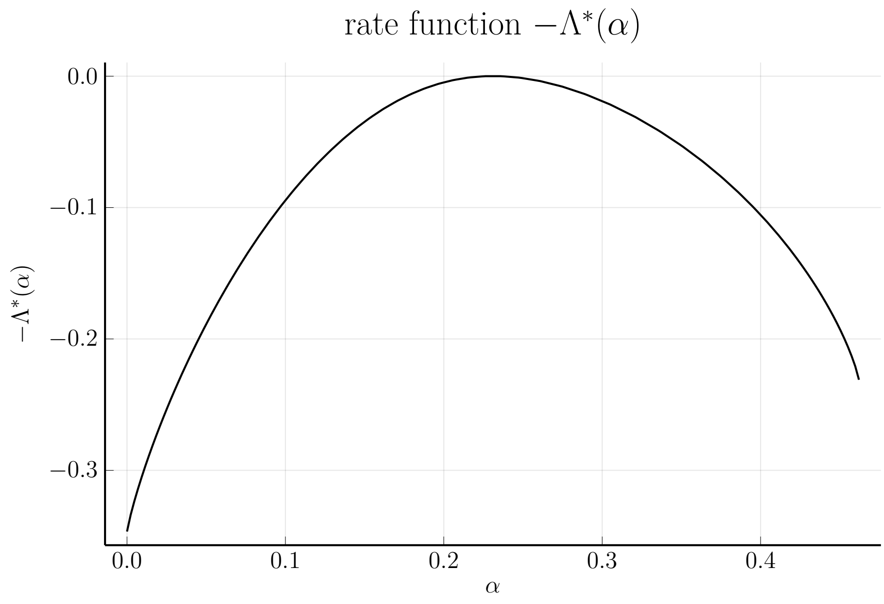

FIGURE 1. rate function Λ_0^*

and

$$\alpha^+ := \limsup_{n\to\infty} \left(\frac{1}{|\Delta_n|} \sum_{g \in \Delta_n \setminus \{\epsilon\}} \log A_{X_g, X_{\varsigma(g)}} \right) = \alpha_0^*.$$

As for (b), the rate function λ^* of the unconditional sample mean is

$$\lambda^*(\alpha) = \max\{\lambda_j^*(\alpha) : \pi|_{\mathcal{A}_j} \neq 0\} = \max\{\lambda_0(\alpha), \lambda_1(\alpha)\} =,$$

which admits two maximal points α_0^* and α_1^*. As for the expectation of sample mean, we may apply the dominated convergence theorem to deduce that

$$\beta^- := \liminf_{n\to\infty} \mathbb{E} \left(\frac{1}{|\Delta_n|} \sum_{g \in \Delta_n \setminus \{\epsilon\}} \log A_{X_g(t), X_{\varsigma(g)}}(t) \right) = \min_{0 \le i \le 1} \sum_{j=0}^{1} \sum_{a \in \mathcal{A}_j} \pi_a \alpha_{i+j}^*.$$

$$\beta^+ := \limsup_{n\to\infty} \mathbb{E} \left(\frac{1}{|\Delta_n|} \sum_{g \in \Delta_n \setminus \{\epsilon\}} \log A_{X_g(t), X_{\varsigma(g)}}(t) \right) = \max_{0 \le i \le 1} \sum_{j=0}^{1} \sum_{a \in \mathcal{A}_j} \pi_a \alpha_{i+j}^*.$$

In particular, in contrast to the case where $\mathbf{A}$ and M are primitive, we have

$$\alpha^- < \beta^- < \beta^+ < \alpha^+,$$

while the four numbers coincide in the primitive case. On account of this, the conditional sample mean provides more information than the unconditional one and the expectation. In fact, in some extreme cases, say

$$A = M = \begin{bmatrix} 0 & 1 & 1 \\ \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 \end{bmatrix}, \pi = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \\ \frac{1}{4} \end{bmatrix}$$

we have that

$$\alpha^- = \frac{1}{3}\log 2, \qquad \alpha^+ = \frac{2}{3}\log 2, \qquad \beta^- = \beta^+ = \frac{1}{2}\log 2$$

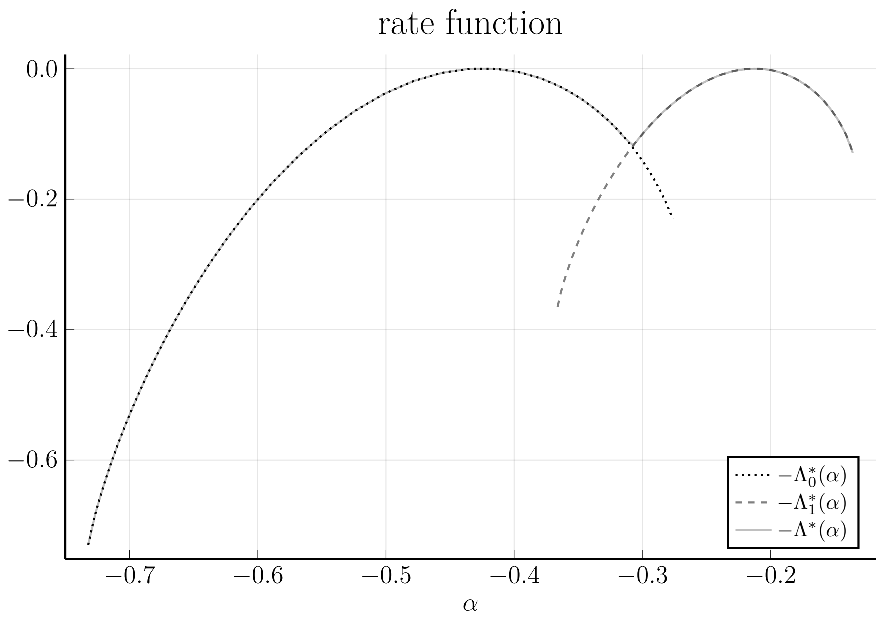

FIGURE 2. rate function Λ_0^*, Λ_1^*, and Λ^*

we may see that over every $t \in \Omega$, the sample mean can be explicitly calculated as

$$\left(\frac{1}{|\Delta_n|} \sum_{g \in \Delta_n \setminus \{\epsilon\}} \log A_{X_{g}(t), X_{\varsigma(g)}}(t) \right) \in \left\{ \alpha^- + o(1), \alpha^+ + o(1) \right\}.$$

However, not a single labeled tree $t \in \Omega$ has sample mean approaching $\beta^+ = \beta^-$.

4. Hausdorff dimensions of transitive Markov hom tree-shifts

The section is devoted to the exposition of the Hausdorff dimension formula of $\mathcal{T}_{\mathbf{A}} \subset \mathcal{A}^T$ when $\mathbf{A}$ is an irreducible matrix. For this purpose, our settings and notations should first be introduced. In this work, we consider the following metric of the tree-shift $\mathcal{T}_{\mathbf{A}}$:

$$D(x, y) = e^{-\sup\{|\Delta_n| : x|_{\Delta_n} = y|_{\Delta_n}\}}, \tag{22}$$

The rationale behind adopting this particular metric lies in its intimate connection with the topological entropy introduced by Petersen and Salama [14]. More specifically, it could be readily checked that, according to this definition, the box counting dimension of $\mathcal{T}_{\mathbf{A}}$ is simply the topological entropy in the sense that $\overline{\dim}_B \mathcal{T}_A = d \cdot h_{top}(\mathcal{T}_A)$ and $\underline{\dim}_B \mathcal{T}_A = h_{top}(\mathcal{T}_A)$, where $h_{top}(\mathcal{T}_A)$ is defined as

$$h_{top}(\mathcal{T}_A) = \lim_{n \to \infty} \frac{\log |B_n(\mathcal{T}_A)|}{|\Delta_n|},$$

with the existence of the limit addressed in the aforementioned work.

The section is organized as follows. A nonlinear Perron-Frobenius theorem will be carried out in 4.1. In Section 4.2, we will present several essential lemmas regarding the duality and discuss the "variational principle" related to the operator $\mathcal{L}_{\mathbf{A},r}$, based on which we could carry out the computation of lower bound and upper bound for $\dim_H \mathcal{T}_{\mathbf{A}}$ in Section 4.3 and 4.4, respectively. Finally, in Section 4.5, we will show that our dimension formula essentially serves as an upper bound for all $\mathcal{T}_{\mathbf{A},r}$ even with (A1) dropped.

4.1. **Nonlinear Perron-Frobenius theory.** The nonlinear Perron-Frobenius theory, as its name suggests, studies the eigenspace of a class of (not necessarily linear) functions and recovers several classical results regarding nonnegative primitive/irreducible matrix transformations. In this work, we will consider such analysis under the framework of [10], of which the setting can be described as follows. Let $f : \mathbb{R}^{\mathcal{A}}_{\geq 0} \to \mathbb{R}^{\mathcal{A}}_{\geq 0}$ be a continuous function. It is said to be *order-preserving* if $f(x) \leq f(y)$ for every $x \leq y$, and it is called *homogeneous* if $f(\alpha x) = \alpha f(x)$ for all $\alpha \in \mathbb{R}_{\geq 0}$, where the comparison of two vectors via $\leq$ is taken entrywise. In addition, we say that f is multiplicatively convex if $\log \circ f \circ \exp$ is a convex function on $\mathbb{R}^{\mathcal{A}}_{>0}$. Within this framework, we establish that the operator $\mathcal{L}_{\mathbf{A},r}$ possesses all the aforementioned characteristics, and that its eigenspace can be characterized as follows. We note that throughout the section, $\|\cdot\|$ denotes the 1-norm.

Proposition 4.1. *The following properties hold for the operator $\mathcal{L}_{\mathbf{A},r}$ when $r \in \mathcal{R}_{p,d}$.*

(1) $\mathcal{L}_{\mathbf{A},r} : \mathbb{R}^{\mathcal{A}}_{\geq 0} \to \mathbb{R}^{\mathcal{A}}_{\geq 0}$ is continuous, order-preserving, homogeneous, analytic on $\mathbb{R}^{\mathcal{A}}_{>0}$, and multiplicatively convex.

(2) For each $\mathcal{A}_j \in \mathcal{P}(\mathbf{A})$, there exist an eigenvector $v^{(j)} \in \mathcal{C}_j \setminus \{0\}$ associated with eigenvalue $\rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r})$ such that

$$\limsup_{n\to\infty} \left\|\mathcal{L}^n_{\mathbf{A},r}(w)\right\|^{1/n} \leq \rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r}) \text{ for all } w \in \mathcal{C}_j.$$

In particular, $v^{(0)}_{a_0} > 0$. If $\mathbf{A}$ is irreducible, then $v^{(j)}$ is unique up to scaling.

(3) Let $\mathcal{C}_j^+ = \{x \in \mathcal{C}_j : x_a > 0 \text{ if } v^{(j)}_a > 0\}$. Then, for every $x \in \mathcal{C}_j^+$ exists an eigenvector $w \in \mathcal{C}_j^+$ and $0 < \theta < 1$ such that

$$\limsup_{n\to\infty} \left\| \frac{\mathcal{L}^n_{\mathbf{A},r}(x)}{\left\|\mathcal{L}^n_{\mathbf{A},r}(x)\right\|} - \frac{w^{(j)}}{\left\|w^{(j)}\right\|} \right\|^{1/n} = \theta.$$

In particular, $\mathcal{L}_{\mathbf{A},r}(w) = \rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r})w$

Proof. (1) The first four conditions follows from a routine check of definition, while the multiplicative convexity follows from Lemma 3.6 by taking $E^{(i)} = \mathbf{A}$, $q_{i+1}(x) = s_i q_i(x)$ with $q_0(x) \equiv 1$ for $i \in [p]$ so that $\log \circ \mathcal{L}_{\mathbf{A},r} \circ \exp(x) = r_{p-1}\lambda^{(p)}(x)$ is convex given that $\lambda^{(0)}(x) = x$ is convex.

(2) We should first note that, for $s > 0$ and $x \in \mathbb{R}^{\mathcal{A}}_{\geq 0}$, $\Psi_{\mathbf{A},s}(x)_a > 0$ if and only if $(\mathbf{A}^T x)_a > 0$. Recursively applying this property one deduces that $\mathcal{L}_{\mathbf{A},r}(x)_a > 0$ if and only if $((\mathbf{A}^T)^p x)_a > 0$. In particular, $\mathcal{L}_{\mathbf{A},r}$ maps $\mathcal{C}_j$ into $\mathcal{C}_j$. Hence, according to [10, Corollary 5.4.2], $\mathcal{L}_{\mathbf{A},r}$ admits an eigenvector in $v^{(j)} \in \mathcal{C}_j$ such that $\mathcal{L}_{\mathbf{A},r}(v^{(j)}) = \rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r}) \cdot v^{(j)}$ and $\|f(w)\| \leq \rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r}) \cdot \|w\|$ for all $w \in \mathcal{C}_j$. Now that if $v^{(0)}$ is an eigenvector as found, then $((A^T)^{pn} v^{(0)})_{a_0} > 0$ for all sufficiently large n according to (A1). Hence, $v^{(0)}_{a_0} = (\rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r}))^{-n} \mathcal{L}_{\mathbf{A},r}(v^{(0)})_{a_0} > 0$. If in addition that $\mathbf{A}$ is irreducible, then by $v^{(j)}_a > 0$ for all $a \in \mathcal{A}_j$. Now by writing $\iota_j : \mathcal{C}_j \to \mathbb{R}^{\mathcal{A}}_{\geq 0}$ the natural inclusion, we note that $\iota_j^{-1} \circ \mathcal{L}_{\mathbf{A},r} \circ \iota_j$ maps $\mathbb{R}^{\mathcal{A}_j}_{\geq 0}$ into $\mathbb{R}^{\mathcal{A}_j}_{\geq 0}$ and the derivative of $\iota_j^{-1} \circ \mathcal{L}_{\mathbf{A},r} \circ \iota_j$ at $\iota_j^{-1}(v^{(j)})$ is primitive. Therefore, by [10, Corollary 6.5.8], $\iota_j^{-1}(v^{(j)})$ is unique up to scaling, and thus so is $v^{(j)}$.

(3) We note that all the properties of $\mathcal{L}_{\mathbf{A},r}$ in (1) are inherited by $\iota_j^{-1} \circ \mathcal{L}_{\mathbf{A},r} \circ \iota_j$. Hence, the existence of w and θ follows as a consequence of [11, Theorem 4.7]. Now that $w \in \mathcal{C}_j^+$ and is associated with some eigenvalue μ, there exists some $C > 0$ such that $w \geq Cv^{(j)}$.

Hence,

$$\mu = \lim_{n\to\infty} \left(\frac{\left\| \mathcal{L}^n_{\mathbf{A},r}(w) \right\|}{\|w\|} \right)^{1/n} \geq \lim_{n\to\infty} \left(\frac{\left\| \mathcal{L}^n_{\mathbf{A},r}(v^{(j)}) \right\|}{\left\|v^{(j)}\right\|} \right)^{1/n} = \rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r}).$$

The proposition is hence proved. □

4.2. **Some lemmas.** The following lemmas are crucial in the later discussions.

Lemma 4.2. *The following equality holds.*

$$\max_{(\mathsf{p},\mathsf{P})\in Z_0} \min_{t\in\Gamma_{[N]}} \sum_{n=0}^{N-1} t_n \cdot f_n(\mathsf{p},\mathsf{P}) = \min_{t\in\Gamma_{[N]}} \max_{(\mathsf{p},\mathsf{P})\in Z_0} \sum_{n=0}^{N-1} t_n \cdot f_n(\mathsf{p},\mathsf{P}).$$

Lemma 4.3. *The following minimax property holds.*

$$\max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ p\text{-periodic}}} \min_{s\in\Gamma_{[p]}} \sum_{m=0}^{p-1} s_m f_m(\mathsf{p},\mathsf{P}) = \min_{s\in\Gamma_{[p]}} \max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ p\text{-periodic}}} \sum_{m=0}^{p-1} s_m f_m(\mathsf{p},\mathsf{P}). \tag{23}$$

Remark 4.4. *We note that in Lemma 4.3, the restriction $(\mathsf{p},\mathsf{P}) \in Z_0$ means that $\mathsf{p}^{(i)}_a > 0$ implies $a \in \mathcal{A}_{-i}$. Therefore, $\mathsf{p}^{(i)} = \mathsf{P}^{(i)}\mathsf{p}^{(i+1)}$ implies $\mathsf{p}^{(i)}_a$ depends only on $\mathsf{P}^{(i)}_{a,b}$ and $\mathsf{p}^{(i+1)}_b$ for $b \in \mathcal{A}_{-i-1}$. Hence, if (p,P) is a p-periodic optimizer, then so is $(\mathsf{p},\mathsf{P}^*)$, where*

$${\mathsf{P}^{(i)}}^*_{a,b} = {\mathsf{P}^{(0)}}^*_{a,b} := \mathsf{P}^{(j)}_{a,b} \quad \text{if} \quad b \in \mathcal{A}_{-j-1}.$$

Therefore, the value of (32) remains unchanged even with every appearance of "$\max_{\substack{\mathsf{P}\in\Upsilon^p_{\mathcal{A}}\\ \mathsf{p}^{(p)}\in\mathcal{C}_0}}$" replaced by "$\max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ \mathsf{P}\ \text{is 1-periodic}\\ \mathsf{p}\ \text{is } p\text{-periodic}}}$".

For convenience, we introduce the following notation before we present our quintessential lemma.

$$t^*_{n,N} = \begin{cases} \left(N + \frac{1}{d^p-1}\right)^{-1} \cdot \frac{d^p}{d^p-1} & \text{if } n = 0; \\ \left(N + \frac{1}{d^p-1}\right)^{-1} & \text{if } p|n \text{ and } 0 < n < pN; \\ 0 & \text{otherwise.} \end{cases}$$

Lemma 4.5. *The following minimax properties hold.*

$$\min_{r\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1} \prod_{i=0}^{\ell} r_i^{-1} \right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r}) \tag{24}$$

$$= \lim_{N\to\infty} \min_{s\in\Gamma_{[p]}} \max_{(\mathsf{p},\mathsf{P})\in Z_0} \sum_{m=0}^{p-1} s_m \sum_{n=0}^{pN-1} t^*_{n,N} f_{n+m}(\mathsf{p},\mathsf{P}) \tag{25}$$

If in addition $\mathbf{A}$ is irreducible, then the above also concide with

$$\min_{s\in\Gamma_{[p]}} \max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ \mathsf{P}\ \text{is 1-periodic}\\ \mathsf{p}\ \text{is } p\text{-periodic}}} \sum_{m=0}^{p-1} s_m f_m(\mathsf{p},\mathsf{P}) \tag{26}$$

$$= \max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ \mathsf{P}\ \text{is 1-periodic}\\ \mathsf{p}\ \text{is } p\text{-periodic}}} \min_{s\in\Gamma_{[p]}} \sum_{m=0}^{p-1} s_m f_m(\mathsf{p},\mathsf{P}) \tag{27}$$

Furthermore, any maximizer P^* *of* (27) *can be chosen to be irreducible and* $\mathsf{p}^{(0)^*} \in \mathcal{C}_0$ *to be a right eigenvector of* $(\mathsf{P}^{(0)^*})^p$.

Remark 4.6. *It is seen from the symmetry of* s *and the symmetry of* t^* *in* (25) *that*

$$\min_{r\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r})$$

is actually independent of j *as long as assumption* (A1) *is satisfied.*

As a consequence of Lemma 4.5, we have the following corollary.

Corollary 4.7. *There exists a Markov measure* $\mathbb{P}$ *associated with a transition matrix* M *and an initial distribution* π *such that it is supported on* $\mathcal{T}_{\mathbf{A}}$ *and that, almost surely,*

$$\liminf_{n\to\infty} \frac{\log \mathbb{P}(\mathsf{C}_n(X))}{|\Delta_n|} = \min_{r\in\mathcal{R}_{p,d}} \min_{\mathcal{A}_j\in\mathcal{P}(\mathbf{A})} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r}), \tag{28}$$

where $\mathsf{C}_n(\cdot)$ *denotes* n*-th level cylinder set associated with* $t \in \Omega$.

$$\mathsf{C}_n(t) = \{t' \in \Omega : X_g(t) = X_g(t'), \forall g \in \Delta_n\}. \tag{29}$$

Proof. The corollary follows from Lemma 4.5 and Corollary 3.10. More precisely, we may choose the initial distribution as in 4.5, so that initial distribution of a symbol a is positive if and only if $a \in \mathcal{A}_0$. Thus, by Corollary 3.10,

$$\lim_{\epsilon\to 0^+}\lim_{n\to\infty} \frac{1}{|\Delta_{pn+j}|} \log \mathbb{P}\left(\frac{1}{|\Delta_{pn+j}|} \sum_{g\in\Delta_{pn+j}\setminus\{\epsilon\}} \log A_{X_g, X_{\varsigma(g)}} \in (\alpha-\epsilon, \alpha+\epsilon)\right)$$

almost surely. □

4.3. **Lower bound.** Suppose $\mathcal{T}_{\mathbf{A}}$ is a tree-shift with an irreducible adjacency matrix $\mathbf{A}$, and that $\mathbb{P}$ is a probability measure (supported on $\mathcal{T}_A$) as Corollary 4.7. Hence, the lower bound is obtained by the mass distribution principle. To be more precise, we note that, due to (28), for every $\delta < \min_{r\in\mathcal{R}_{p,d}} \min_{\mathcal{A}_j\in\mathcal{P}(\mathbf{A})} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r})$, we have

$$\limsup_{n\to\infty} \frac{\mathbb{P}(\mathsf{C}_n(X))}{e^{|\Delta_n|^{\delta}}} = 0 \quad \text{almost surely.}$$

Hence, by Egorov's theorem, there exists a subset $S \subset \mathcal{T}_{\mathbf{A}}$ with $\mathbb{P}(S) > 0$ and a constant C such that

$$\mathbb{P}(\mathsf{C}_n(t)) \le Ce^{-\delta|\Delta_n|} \text{ for all } n \in \mathbb{N}, t \in S.$$

As a consequence, supposing $\mathcal{C}$ is a cover of S consisting of disjoint cylinder sets, we have

$$\sum_{U\in\mathcal{C}} (\operatorname{diam}(U))^{\delta} \ge C^{-1} \sum_{U\in\mathcal{C}} \mathbb{P}(U) \ge C^{-1}\mathbb{P}(S) > 0.$$

Hence, we have $\mathcal{H}^{\delta}(S) \ge C^{-1}\mathbb{P}(S) > 0$ and $\dim_H \mathcal{T}_{\mathbf{A}} \ge \dim_H S \ge \delta$. Consequently, $\dim_H \mathcal{T}_{\mathbf{A}} \ge \min_{s\in\mathcal{R}_{p,d}} \min_{\mathcal{A}_j\in\mathcal{P}(\mathbf{A})} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r})$.

4.4. **Upper bound.** The upper bound of the Hausdorff dimension is proved by construction of suitable covers. Within the construction process, we will naturally encounter the nonlinear transfer operator $\mathcal{L}_{\mathbf{A},r}$, which gives rise to the discussion of Section 4.1

The study of the operator $\mathcal{L}_{\mathbf{A},r}$ is motivated by a naïve covering strategy for the space $\mathcal{T}_{\mathbf{A}}$. We recall that the feasible domain $W_{n:n+k}$ is asymptotically dense in the feasible domain (Lemma 3.4).

Now for $n + N > n \gg 0$ ($N \in \mathbb{N}$), we consider a covering $\mathcal{C}_{n,N}$ agreeing with the following philosophy:

(a) The covering $\mathcal{C}_{n,N}$ consists of cylinder sets of level n to level $n+N-1$, i.e.,

$$\mathcal{C}_{n,N} \subset \{\mathsf{C}_m(t) : t \in \mathcal{T}_{\mathbf{A}}, n \leq m < n+N\}.$$

(b) A cylinder set $\mathsf{C}_k(t)$ is contained in $\mathcal{C}_{n,N}$ if there exists $t \in \mathcal{T}_{\mathbf{A}}$ such that

$$\frac{|B_k(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_k|} \leq \min_{n\leq k<n+N} \frac{|B_m(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_m|}. \tag{30}$$

In other words, we include condition (a) in the belief that the Hausdorff dimension is approached by the intermediate dimension (see [3]) and impose condition (b) in the hope that the upper bound of the Hausdorff dimension is minimized. More precisely, under the circumstances, if

$$s > \liminf_{N\to\infty} \liminf_{n\to\infty} \max_{t\in\mathcal{T}_{\mathbf{A}}} \min_{n\leq k<n+N} \frac{|B_m(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_m|},$$

then,

$$\begin{aligned}
\mathcal{H}^s(\mathcal{T}_{\mathbf{A}}) &\leq \liminf_{N\to\infty} \liminf_{n\to\infty} \sum_{m=n}^{n+N-1} \sum_{[u]\in\mathcal{C}_{n,N}: u\in B_n(\mathcal{T}_{\mathbf{A}})} e^{-s|\Delta_m|} \\
&\leq \liminf_{N\to\infty} \liminf_{n\to\infty} \sum_{m=n}^{n+N-1} |W_{n:m}| e^{-(s-\min_{n\leq k<n+N} \frac{|B_k(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_k|})|\Delta_m|},
\end{aligned}$$

Now that Lemma 3.3 asserts

$$|W_{n:m}| \leq |D_{n:m}||S_{n:m}| \leq \left(\frac{|\Delta_{n:m}|}{m-n+1}+1\right)^{3(m-n+1)\cdot|\mathcal{A}|^2},$$

and thus $\mathcal{T}_{\mathbf{A}}$ is s-Hausdorff null:

$$\begin{aligned}
\mathcal{H}^s(\mathcal{T}_{\mathbf{A}}) \leq \liminf_{N\to\infty} \liminf_{n\to\infty} \sum_{m=n}^{n+N-1} e^{-(s-\min\limits_{n\leq k<n+N} \frac{|B_k(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_k|} - 3m|\mathcal{A}|^2 \frac{\log|\Delta_m|}{|\Delta_m|})|\Delta_m|} \\
\to 0.
\end{aligned}$$

This bounds the Hausdorff dimension from above by

$$\dim_H \mathcal{T}_{\mathbf{A}} \leq \liminf_{N\to\infty} \liminf_{n\to\infty} \max_{t\in\mathcal{T}_{\mathbf{A}}} \min_{n\leq k<n+N} \frac{|B_m(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_m|}.$$

In practice, to have an estimate of $\max_{t\in\mathcal{T}_{\mathbf{A}}} \min_{n\leq k<n+N} \frac{|B_m(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_m|}$, we apply Lemma 3.4 so that for any fixed N,

$$\begin{aligned}
&\max_{\substack{t\in\mathcal{T}_{\mathbf{A}}:\\ t_\epsilon\in\mathcal{A}_j}} \min_{n\leq m<n+N} \frac{|B_m(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_m|} \leq \min_{n\leq m<n+N} \max_{\substack{t\in\mathcal{T}_{\mathbf{A}}:\\ t_\epsilon\in\mathcal{A}_j}} \frac{|B_m(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_m|} \\
&\leq \min_{s\in\Gamma_{[N]}} \Big(\max_{(\mathsf{p},\mathsf{P})\in Z_{j+n+N}} \sum_{m=0}^{N-1} s_m f_{m,n+N}(\mathsf{p},\mathsf{P})\Big) + o_n(1),
\end{aligned}$$

where $o_n(1)$ according to Lemma 3.3 and Lemma 3.5 is a number independent of N that vanishes as $n\to\infty$. Now since $j = 0,1,\cdots,p-1$ has finitely many choices, we may fix $j = j_0$ such that

$$\liminf_{n\to\infty} \max_{\substack{t\in\mathcal{T}_{\mathbf{A}}:\\ t_\epsilon\in\mathcal{A}_{j_0}}} \min_{n\leq m<n+N} \frac{|B_m(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_m|} = \liminf_{n\to\infty} \max_{t\in\mathcal{T}_{\mathbf{A}}} \min_{n\leq m<n+N} \frac{|B_m(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_m|}.$$

One can now apply Lemma 4.2 to derive

$$\begin{aligned}
&\liminf_{n\to\infty} \max_{t\in\mathcal{T}_{\mathbf{A}}} \min_{n\le m<n+N} \frac{|B_m(\mathcal{T}_{\mathbf{A}};\tau(t),\eta(t))|}{|\Delta_m|} \\
\le &\liminf_{n\to\infty} \min_{s\in\Gamma_{[N]}} (\max_{(\mathsf{p},\mathsf{P})\in Z_0} \sum_{m=0}^{N-1} s_m f_{m,pn-j_0+N}(\mathsf{p},\mathsf{P})) \\
\le &\min_{s\in\Gamma_{[N]}} \max_{(\mathsf{p},\mathsf{P})\in Z_0} \sum_{m=0}^{N-1} s_m f_m(\mathsf{p},\mathsf{P}) = \max_{(\mathsf{p},\mathsf{P})\in Z_0} \min_{s\in\Gamma_{[N]}} \sum_{m=0}^{N-1} s_m f_m(\mathsf{p},\mathsf{P}).
\end{aligned}$$

Next, if we let N tend to ∞, we may use Lemma 4.5 to illustrate the asymptotic behavior of the above formula:

$$\begin{aligned}
&\lim_{N\to\infty} \max_{(\mathsf{p},\mathsf{P})\in Z_0} \min_{s\in\Gamma_{[N]}} \sum_{m=0}^{N-1} s_m f_m(\mathsf{p},\mathsf{P}) \\
\le &\lim_{N\to\infty} \min_{s\in\Gamma_{[p]}} \max_{(\mathsf{p},\mathsf{P})\in Z_0} \sum_{m=0}^{p-1} s_m \sum_{n=0}^{pN-1} t^*_{n,N} f_{n+m}(\mathsf{p},\mathsf{P}) \\
= &\min_{r\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1} \prod_{i=0}^{\ell} r_i^{-1} \right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},s}).
\end{aligned}$$

The upper bound thus coincides with the lower bound.

4.5. **Upper bound for general $\mathcal{T}_{\mathbf{A}}$.** In this subsection, we derive an upper bound for $\mathcal{T}_{\mathbf{A}}$ without assumption (A1). To begin with, for every $a \in \mathcal{A}$ we define

$$\mathcal{A}^{(a)} = \{b \in \mathcal{A} : (\mathcal{A}^n)_{b,a} \text{ for some } n \in \mathbb{N}, b \in \mathcal{A}\} \cup \{a\},$$

associated with which is a submatrix $\mathbf{A}^{(a)}$ of $\mathbf{A}$ indexed by $\mathcal{A}^{(a)}$. In addition, we write

$$\hat{\mathcal{A}} = \{a \in \mathcal{A} : (\mathbf{A}^n)_{a,a} \text{ for infinitely may } n \in \mathbb{N}\},$$

so that the subsystem $\mathcal{T}_{\mathbf{A}^{(a)}}$ satisfies assumption (A1) if and only if $a \in \hat{\mathcal{A}}$. In such situation, $\mathcal{A}^{(a)}$ admits a partition $\mathcal{P}(\mathbf{A}^{(a)})$ consisting of $p^{(a)}$ sets as in (6). For convenience, we define

$$Z^{(a)} = \{(\mathsf{p},\mathsf{P}) \in \Gamma_{\mathcal{A}^{(a)}}^{\mathbb{Z}_+} \times \Upsilon_{\mathcal{A}^{(a)}}^{\mathbb{Z}_+} : \mathsf{p}^{(i)} = \mathsf{P}^{(i)}\mathsf{p}^{(i+1)}, \mathsf{P}^{(i)}_{a,b} = 0 \text{ if } \mathbf{A}^{(a)}_{a,b} = 0\}$$

and

$$Z^{(a),n} = \{(\mathsf{p},\mathsf{P}) \in \Gamma_{\mathcal{A}^{(a)}}^{n} \times \Upsilon_{\mathcal{A}^{(a)}}^{n-1} : \mathsf{p}^{(i)} = \mathsf{P}^{(i)}\mathsf{p}^{(i+1)}, \mathsf{P}^{(i)}_{a,b} = 0 \text{ if } \mathbf{A}^{(a)}_{a,b} = 0\}.$$

The goal of the subsection is to establish the following upper bound.

Proposition 4.8. *Suppose $\mathcal{T}_{\mathbf{A}}$ is a Markov hom tree-shift with $\mathbf{A}$. Then,*

$$\dim_H \mathcal{T}_{\mathbf{A}} \le \max_{a\in\hat{\mathcal{A}}} \min_{r\in\mathcal{R}_{p^{(a)},d}} \left(\sum_{\ell=0}^{p-1} \prod_{i=p^{(a)}}^{\ell} r_i^{-1} \right)^{-1} \cdot \log \rho_{\mathcal{A}_0^{(a)}}(\mathcal{L}_{\mathbf{A}^{(a)},r}). \tag{31}$$

More specifically, since

$$\dim_H \mathcal{T}_{\mathbf{A}} = \dim_H \cup_{a\in\mathcal{A}} \mathcal{T}_{\mathbf{A}^{(a)}} = \max_{a\in\mathcal{A}} \dim_H \mathcal{T}_{\mathbf{A}^{(a)}},$$

we may further focus on each $\mathcal{T}_{\mathbf{A}^{(a)}}$ without loss of generality. Now that the case $a \in \hat{\mathcal{A}}$ has been resolved in the previous subsection, in what follows we will assume $a \notin \hat{\mathcal{A}}$. Under the circumstances, there is a nonempty set

$$\mathcal{A}^{(a)\prime} := \{b \in \mathcal{A}^{(a)} : \sup\{n \in \mathbb{N} : ((\mathbf{A}^{(a)})^n)_{b,a}\} < \infty\},$$

which is associated with a number

$$k = \sup\{n+1 \in \mathbb{N} : ((\mathbf{A}^{(a)})^n)_{b,a} > 0, a \in \mathcal{A}^{(a)\prime}\}.$$

We then have a similar estimate as before:

$$\max_{t \in \mathcal{T}_{\mathbf{A}^{(a)}}} \min_{n \leq m < n+N} \frac{|B_m(\mathcal{T}_{\mathbf{A}^{(a)}}; \tau(t), \eta(t))|}{|\Delta_m|} \leq \max_{(\overleftarrow{\mathsf{p}}, \overleftarrow{\mathsf{P}}) \in Z^{(a),n+N}} (\min_{n \leq m < n+N} f_{m,n+N}(\mathsf{p}, \mathsf{P})) + o_n(1).$$

We can further use a rearranged form (as in (34)) and a similar argument as in Lemma 4.2 to derive that

$$\begin{aligned}
&\max_{(\mathsf{p},\mathsf{P}) \in Z^{(a),n+N}} \min_{s \in \Gamma_{[N]}} \sum_{m=0}^{N-1} s_m f_{m,n+N}(\mathsf{p}, \mathsf{P}) \\
\leq &\max_{\substack{(\mathsf{p},\mathsf{P}) \in Z^{(a),n+N-k} \\ \mathsf{p}_b^{(n+N-k)} = 0, \forall b \in \mathcal{A}^{(a)\prime}}} \min_{s \in \Gamma_{[N]}} \sum_{m=0}^{N-1} s_m f_{m,n+N-k}(\mathsf{p}, \mathsf{P}) + \frac{1}{d^{n-k}} \log |\mathcal{A}| \\
= &\min_{s \in \Gamma_{[N]}} \max_{\substack{(\mathsf{p},\mathsf{P}) \in Z^{(a),n+N-k} \\ \mathsf{p}_b^{(n+N-k)} = 0, \forall b \in \mathcal{A}^{(a)\prime}}} \sum_{m=0}^{N-1} s_m f_{m,n+N-k}(\mathsf{p}, \mathsf{P}) + \frac{1}{d^{n-k}} \log |\mathcal{A}| \\
= &\min_{s \in \Gamma_{[N]}} \max_{\substack{(\mathsf{p},\mathsf{P}) \in Z^{(a),n+N-k} \\ \mathsf{p}^{(n+N-k)} = \mathsf{e}_b, b \notin \mathcal{A}^{(a)\prime}}} \sum_{m=0}^{N-1} s_m f_{m,n+N-k}(\mathsf{p}, \mathsf{P}) + \frac{1}{d^{n-k}} \log |\mathcal{A}| \\
=: &\min_{s \in \Gamma_{[N]}} \max_{b \in \mathcal{A}^{(a)} \setminus \mathcal{A}^{(a)\prime}} \phi(s, b, n, N).
\end{aligned}$$

Now we can choose subsequences N_j of N and n_ℓ of n, such that

$$\lim_{j \to \infty} \lim_{\ell \to \infty} \min_{s \in \Gamma_{[N_j]}} \max_{b \in \mathcal{A}^{(a)} \setminus \mathcal{A}^{(a)\prime}} \phi(s, b, n_\ell, N_j)$$

is well-defined, associated with which are optimizers $s(n_\ell, N_j)$ and $b(n_\ell, N_j)$ of

$$\min_{s \in \Gamma_{[N]}} \max_{b \notin \hat{\mathcal{A}}^{(a)\prime}} \phi(s, b, n, N).$$

Without loss of generality, we may further assume that $b(n_\ell, N_j) = b^*$ is constant, for if otherwise, one can still take subsequences of N_j and n_ℓ to have the properties fulfilled. It is noteworthy that since $b^* \notin \mathcal{A}^{(a)\prime}$, we may find an $a^* \in \mathcal{A}^{(a)} \cap \hat{\mathcal{A}}$ such that $b \in \mathcal{A}^{(a^*)}$. Hence, as a consequence of duality and Lemma 4.5, we have

$$\begin{aligned}
&\lim_{j \to \infty} \lim_{\ell \to \infty} \min_{s \in \Gamma_{[N_j]}} \max_{b \notin \hat{\mathcal{A}}^{(a)\prime}} \phi(s, b, n_\ell, N_j) = \lim_{j \to \infty} \lim_{\ell \to \infty} \min_{s \in \Gamma_{[N_j]}} \phi(s, b^*, n_\ell, N_j) \\
\leq &\limsup_{N \to \infty} \limsup_{n \to \infty} \min_{s \in \Gamma_{[N]}} \max_{(\mathsf{p},\mathsf{P}) \in Z^{(a^*),n+N-k}} \sum_{m=0}^{N-1} s_m f_{m,n+N-k}(\mathsf{p}, \mathsf{P}) \\
= &\limsup_{N \to \infty} \limsup_{n \to \infty} \max_{(\mathsf{p},\mathsf{P}) \in Z^{(a^*),n+N-k}} \min_{s \in \Gamma_{[N]}} \sum_{m=0}^{N-1} s_m f_{m,n+N-k}(\mathsf{p}, \mathsf{P}) \\
\leq &\max_{a \in \hat{\mathcal{A}}} \min_{r \in \mathcal{R}_{p^{(a)},d}} \left(\sum_{\ell=0}^{p-1} \prod_{i=p^{(a)}}^{\ell} r_i^{-1} \right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A}^{(a)},r}).
\end{aligned}$$

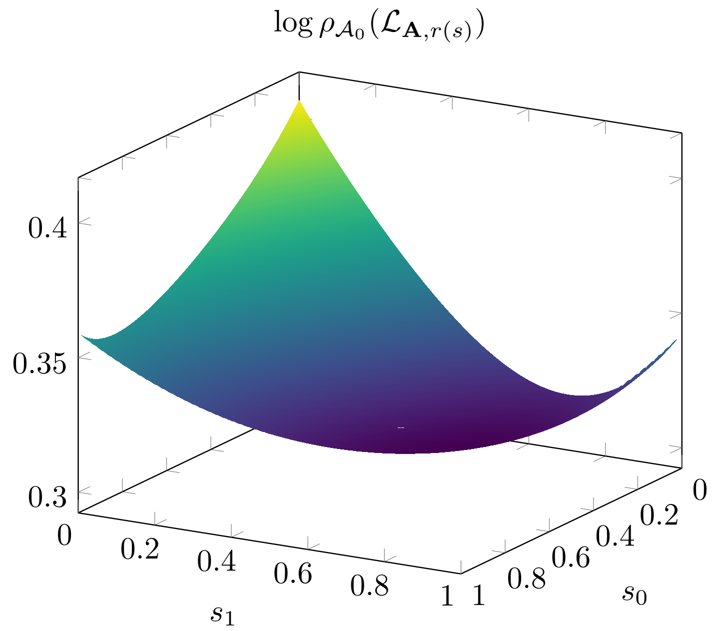

FIGURE 3. spectral radius $\log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r(s)})$

Proof of Corollary 1.3. It suffices observe that for each $a \in \hat{\mathcal{A}}$, we may take $r = \mathbb{1} \in \mathcal{R}_{p^{(a)},d}$ so that

$$\left(\sum_{\ell=0}^{p-1} \prod_{i=p^{(a)}}^{\ell} r_i^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A}^{(a)},r}) = \frac{1}{p^{(a)}} \log \rho_{\mathcal{A}_0^{(a)}}(({\mathbf{A}^{(a)}}^T)^{p^{(a)}}) \leq \log \rho(\mathbf{A}).$$

The theorem is thus proved. □

Here we provide an example demonstrating our formula.

Example 4.9. *Consider a $\mathcal{T}_{\mathbf{A}}$ over the 3-tree defined by an irreducible adjacency matrix.*

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Then, one can numerically compute the Hausdorff dimension by finding the eigenvalues and eigenvectors, according to Proposition 4.1, via iteration, for which the logarithm of the eigenvalue is plotted in Figure 3. Numerically, the optimal s^* is around $(0.312, 0.588, 0.010)$ *with the associated eigenvalue* $\log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r(s^*)}) \approx 0.3027$*, which is strictly less than the logarithm of spectral radius of* $\log \rho(\mathbf{A}^T) \approx 0.3208$ *and is consistent with Corollary 1.3.*

5. Discussion

Inspired by the classical large deviation principle for Markov chains, this article analogously establish the large deviation principle for Markov chains indexed by rooted d-trees with irreducible transition matrices. In the belief of its potential for providing a better conception of the tree-shifts as well as further applications, the expositions in the previous sections showcase the capability of the method of types argument: it not only provides a theorem regarding almost sure convergence, but acts as the key machinery in the exposition of Hausdorff dimension. Despite the aforementioned accomplishments, our work is not meant to be a complete investigation, but would be just a starting point. Indeed, several questions still remain unanswered in this works, which we list them as follows.

- Is the rate function $\lambda_j^*(s)$ in Theorem 1.1 continuously differentiable if $\mathbf{A}$ is irreducible?
- If $\mathbf{A}$ is irreducible, can the the optimal $r \in \mathcal{R}_{p,d}$ in Theorem 3.11 be determined? Can the optimal Markov measure in Lemma 4.5 be determined?
- What is the Hausdorff dimension for general $\mathcal{T}_{\mathbf{A}}$ without the assumption $\mathbf{A}$ being irreducible?

We hope this paper will inspire further studies in this direction to push the frontier of our knowledge about the tree-indexed Markov chains as well as the tree-shifts.

Appendix A. Proofs of Lemmas

A.1. **Proof of Lemma 4.2.** We fix $N' \gg N$ be some large integer. Since the inequality "$\leq$" is automatic by definition, and it remains to show the other inequality. Additionally, due to the continuity of the objective function in $(t, \mathsf{p}, \mathsf{P})$, it is equivalent to prove the lemma by showing the inequality with $\inf_{t\in\mathring{\Gamma}_{[N]}}$ in place of $\min_{t\in\Gamma_{[N]}}$. By putting

$$q_i(t) = \begin{cases} \sum_{n=0}^{i} t_n \frac{d-1}{d^{i-n+1}} & \text{if } i < N; \\ \sum_{n=0}^{N-1} t_n \frac{d-1}{d^{i-n+1}} & \text{if } i \geq N, \end{cases}$$

we can apply the minimax theorem to deduce

$$\begin{aligned}
&\max_{(\mathsf{p},\mathsf{P})\in\mathbb{Z}_0} \inf_{t\in\mathring{\Gamma}_{[N]}} \sum_{n=0}^{N-1} t_n \cdot f_n(\mathsf{p},\mathsf{P}) \\
= &\max_{(\mathsf{p},\mathsf{P})\in\mathbb{Z}_0} \inf_{t\in\mathring{\Gamma}_{[N]}} \sum_{i=0}^{\infty} q_i(t) \cdot \mathbb{1}^T \left(\mathsf{P}^{(i)} \odot \log \frac{1}{\mathsf{P}^{(i)}}\right) \mathsf{p}^{(i+1)} \\
\geq &\max_{(\mathsf{p},\mathsf{P})\in\mathbb{Z}_0} \inf_{t\in\mathring{\Gamma}_{[N]}} \sum_{i=0}^{N'-1} q_i(t) \cdot \mathbb{1}^T \left(\mathsf{P}^{(i)} \odot \log \frac{1}{\mathsf{P}^{(i)}}\right) \mathsf{p}^{(i+1)} \\
= &\max_{\mathsf{p}^{(N)}\in\mathcal{C}_0} \max_{\mathsf{P}^{(1:N-1)}} \inf_{t\in\mathring{\Gamma}_{[N]}} {\lambda^{(1)}}^T \mathsf{p}^{(1)} + \sum_{i=1}^{N'-1} q_i(t) \cdot \mathbb{1}^T \left(\mathsf{P}^{(i)} \odot \log \frac{1}{\mathsf{P}^{(i)}}\right) \mathsf{p}^{(i+1)},
\end{aligned}$$

where $\lambda^{(0)} = 0$ and $\lambda^{(1)} = q_0(t) \log A^T e^{q_0(t)^{-1}\lambda^{(0)}}$ are convex in t. Next, we define iteratively that $\lambda^{(k+1)} = q_k(t) \log A^T e^{q_k(t)^{-1}\lambda^{(k)}}$, which are convex in t for all k due to Lemma 3.6. Then, we may recursively apply the minimax theorem to derive

$$\begin{aligned}
&\max_{(\mathsf{p},\mathsf{P})\in Z_0} \inf_{t\in\mathring{\Gamma}_{[N]}} \sum_{n=0}^{N-1} t_n \cdot f_n(\mathsf{p},\mathsf{P}) \\
\geq &\inf_{t\in\mathring{\Gamma}_{[N]}} \max_{(\mathsf{p},\mathsf{P})\in\mathbb{Z}_0} \sum_{i=0}^{N'-1} q_i(t) \cdot \mathbb{1}^T \left(\mathsf{P}^{(i)} \odot \log \frac{1}{\mathsf{P}^{(i)}}\right) \mathsf{p}^{(i+1)} \\
\geq &\inf_{t\in\mathring{\Gamma}_{[N]}} \max_{(\mathsf{p},\mathsf{P})\in\mathbb{Z}_0} \sum_{i=0}^{\infty} q_i(t) \cdot \mathbb{1}^T \left(\mathsf{P}^{(i)} \odot \log \frac{1}{\mathsf{P}^{(i)}}\right) \mathsf{p}^{(i+1)} - \frac{\log|\mathcal{A}|}{d^{N'}}
\end{aligned}$$

The lemma is proved by letting N' tend to infinity.

A.2. **Proof of Lemma 4.3.** Due to the continuity of the objective function, it suffices to show the equality restricted to the interior $\mathring{\Gamma}_{[p]}$ of $\Gamma_{[p]}$. Namely, it is equivalent to show the equality with $\inf_{s\in\mathring{\Gamma}_{[p]}}$ in place of all appearances of $\min_{s\in\Gamma_{[p]}}$ in (23). Under the circumstances, we first rephrase, by virtue of periodicity, the left-hand side of (23) as

$$\max_{\substack{\mathsf{P}\in\Upsilon_{\mathcal{A}}^{p}\\ \mathsf{p}^{(p)}\in\mathcal{C}_0}} \inf_{\substack{s\in\mathring{\Gamma}_{[p]}\\ \mu\in\mathbb{R}^{\mathcal{A}}}} \sum_{i=0}^{p-1} \frac{\sum_{m=1}^{p} s_{m+i}d^m}{\sum_{m=1}^{p} d^m} \mathbb{1}^T \left(\mathsf{P}^{(i)} \odot \log\frac{\mathbf{A}}{\mathsf{P}^{(i)}}\right) \prod_{\ell=i+1}^{p-1} \mathsf{P}^{(\ell)}\mathsf{p}^{(p)} + \mu^T \left(\prod_{\ell=0}^{p-1} \mathsf{P}^{(\ell)}\mathsf{p}^{(p)} - \mathsf{p}^{(p)}\right), \tag{32}$$

where s_{p+i} are simply aliases of s_i for $i = 0, \cdots, p-1$, respectively. For short, we denote by $F(\mathsf{p}^{(p)}, \mathsf{P}, s, \mu)$ the objective function.

As always, the expression on the left of (23) is by definition no larger than the right, so it suffices to show the other inequality. The proof essentially exploits the following sequence of functions $(\lambda^{(i)}(s,\mu))_{i=0}^{p-1}$, whose convexity is guaranteed in Lemma 3.6. We let $q_i(s) = (\sum_{m=1}^{p} d^m)^{-1}(\sum_{m=1}^{p} s_{m+i}d^m)$, and define

$$\lambda^{(i+1)}(s,\mu) = \begin{cases} \mu & \text{if } i = 0; \\ q_i(s)\log(A^T e^{q_i(s)^{-1}\lambda^{(i)}(s,\mu)}) & \text{if } i = 1, 2, \cdots, p-1. \end{cases}$$

One can now apply the minimax theorem to $\mathsf{P}^{(0)}$ and (s,μ), since the objective function is concave in the former variable and convex in the latter. Swapping $\max_{\mathsf{P}^{(0)}}$ and $\inf_{(s,\mu)}$ turns (32) into

$$\max_{\substack{\mathsf{P}^{(1:p-1)}\in\Upsilon_{\mathcal{A}}^{p}\\ \mathsf{p}^{(p)}\in\mathcal{C}_0}} \inf_{\substack{s\in\mathring{\Gamma}_{[p]}\\ \mu\in\mathbb{R}^{\mathcal{A}}}} \sum_{i=1}^{p-1} \frac{\sum_{m=1}^{p} s_{m+i}d^m}{\sum_{m=1}^{p} d^m} \mathbb{1}^T \left(\mathsf{P}^{(i)} \odot \log\frac{\mathbf{A}}{\mathsf{P}^{(i)}}\right) \prod_{\ell=i+1}^{p-1} \mathsf{P}^{(\ell)}\mathsf{p}^{(p)} + {\lambda^{(1)}}^T \prod_{\ell=1}^{p-1} \mathsf{P}^{(\ell)}\mathsf{p}^{(p)} - \mu^T\mathsf{p}^{(p)},$$

Recursively using convexity of $\lambda^{(i)}$ and applying the minimax theorem, one will move all max in (32) to the right of inf while retaining equality. In particular,

$$\begin{aligned}
&\max_{\substack{\mathsf{P}\in\Upsilon_{\mathcal{A}}^{p}\\ \mathsf{p}^{(p)}\in\mathcal{C}_0}} \inf_{\substack{s\in\mathring{\Gamma}_{[p]}\\ \mu\in\mathbb{R}^{\mathcal{A}}}} F(\mathsf{p}^{(p)}, \mathsf{P}, s, \mu) = \inf_{\substack{s\in\mathring{\Gamma}_{[p]}\\ \mu\in\mathbb{R}^{\mathcal{A}}}} \max_{\substack{\mathsf{P}\in\Upsilon_{\mathcal{A}}^{p}\\ \mathsf{p}^{(p)}\in\mathcal{C}_0}} F(\mathsf{p}^{(p)}, \mathsf{P}, s, \mu) \\
&\geq \inf_{\substack{s\in\mathring{\Gamma}_{[p]}\\ \mu\in\mathbb{R}^{\mathcal{A}}}} \max_{\substack{\mathsf{P}\in\Upsilon_{\mathcal{A}}^{p}\\ \mathsf{p}^{(p)}\in\mathcal{C}_0\\ \mathsf{p}^{(p)}=\prod_{\ell=0}^{p-1}\mathsf{P}^{(\ell)}\mathsf{p}^{(p)}}} F(\mathsf{p}^{(p)}, \mathsf{P}, s, \mu) \\
&= \inf_{s\in\mathring{\Gamma}_{[p]}} \max_{\substack{\mathsf{P}\in\Upsilon_{\mathcal{A}}^{p}\\ \mathsf{p}^{(p)}\in\mathcal{C}_0\\ \mathsf{p}^{(p)}=\prod_{\ell=0}^{p-1}\mathsf{P}^{(\ell)}\mathsf{p}^{(p)}}} F(\mathsf{p}^{(p)}, \mathsf{P}, s, 0) = \min_{s\in\Gamma_{[p]}} \max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ p\text{-periodic}}} \sum_{m=0}^{p-1} s_m f_m(\mathsf{p},\mathsf{P}).,
\end{aligned}$$

which agrees with the right-hand side of (23) (with $\inf_{s\in\mathring{\Gamma}_{[p]}}$ in place of $\min_{s\in\Gamma_{[p]}}$.) The lemma is hence proved.

A.3. **Proof of Lemma 4.5.** We first prove the following auxiliary lemma:

Lemma A.1. *Suppose* $s \in \Gamma_{[p]}$,

$$q_i(s) = \sum_{j=0}^{p-1} \frac{s_{i-j}}{d^j} \frac{d^p - d^{p-1}}{d^p - 1} \text{ and } r(s) = \left(\frac{q_0(s)}{q_1(s)}, \frac{q_1(s)}{q_2(s)}, \cdots, \frac{q_{p-1}(s)}{q_0(s)} \right).$$

Then, $s \mapsto q(s)$ *is a bijection between* $\Gamma_{[p]}$ *and* $\{q \in \Gamma_{[p]} : q_0 \le dq_1 \le \cdots \le d^{p-1}q_{p-1} \le d^p q_0\}$*, and* $q(s) \mapsto r(s)$ *is a bijection between* $\{q \in \Gamma_{[p]} : q_0 \le dq_1 \le \cdots \le d^{p-1}q_{p-1} \le d^p q_0\}$ *and* $\mathcal{R}_{p,d}$.

Proof. We note that q_i is a bijection between $\Gamma_{[p]}$ and $\{q \in \Gamma_{[p]} : q_0 \le dq_1 \le \cdots \le d^{p-1}q_{p-1} \le d^p q_0\}$. Indeed, $s \mapsto q_i(s)$ is an injective linear transformation, and every element in the latter set admits a preimage in the former. It remains to show that $q(s) \mapsto r(s)$ maps $\{q \in \Gamma_{[p]} : q_0 \le dq_1 \le \cdots \le d^{p-1}q_{p-1} \le d^p q_0\}$ bijectively into $\mathcal{R}_{p,d}$. It is not hard to see that such map is well-defined and $\mathcal{R}_{p,d}$ is its image. The injectivity follows from the constraint that $\sum_{i=0}^{p-1} q_i(s) = 1$, while bijectivity follows by noting that its inverse can be explicitly found as $q_0(s) = (\sum_{i=0}^{p-1} \prod_{\ell=0}^{i} r_\ell^{-1}(s))^{-1}$ and $q_i(s) = \prod_{\ell=0}^{i} r_\ell^{-1}(s) q_0(s)$. □

For the sake of simplicity, we denote by $F_N(\mathsf{p}, \mathsf{P}, s)$ the objective function of (25). Before we proceed with the proof, we first present a useful identity: for every $s, s' \in \Gamma_{[p]}$, it yields, due to the uniform boundedness $0 \le f_n(\mathsf{p}, \mathsf{P}) \le \log |\mathcal{A}|$, that

$$\left| \max_{(\mathsf{p},\mathsf{P}) \in Z_0} F_N(\mathsf{p}, \mathsf{P}, s) - \max_{(\mathsf{p},\mathsf{P}) \in Z_0} F_N(\mathsf{p}, \mathsf{P}, s') \right| \le 2 \cdot \log |\mathcal{A}| \cdot \mathsf{d}_v(s, s') \tag{33}$$

which is independent of N. Also, will first prove the equality between (24) and (25) by substituting $\inf_{r \in \mathcal{R}_{p,d}}$ for $\min_{r \in \mathcal{R}_{p,d}}$ and then demonstrate the infimum can indeed be attained.

Our general setup for our proof is as follows. By denoting $s_{p+i} = s_i$ for $i \in \mathbb{Z}_+$ and

$$q_{i,N}(s) = \begin{cases} \sum\limits_{j=0}^{i} \frac{s_{i-j}}{d^j} \left(N + \frac{1}{d^p - 1} \right)^{-1} \frac{d^p - d^{p-1}}{d^p - 1} & \text{if } 0 \le i < p; \\ \sum\limits_{j=0}^{p-1} \frac{s_{i-j}}{d^j} \left(N + \frac{1}{d^p - 1} \right)^{-1} \frac{d^p - d^{p-1}}{d^p - 1} & \text{if } p \le i < pN - 1; \\ d^{-(i - pN + 1)} q_{i,pN-1}(s) & \text{if } pN \le i, \end{cases}$$

we derive an alternative expression of the objective function F_N

$$F_N(\mathsf{p}, \mathsf{P}, s) = \sum_{i=0}^{\infty} q_{i,N}(s) \mathbb{1}^T \left(\mathsf{P}^{(i)} \log \frac{\mathbf{A}}{\mathsf{P}^{(i)}} \right) \mathsf{p}^{(i+1)}. \tag{34}$$

Based on this expression, we may consider the following approximation

$$F_{N,m}(\mathsf{p}, \mathsf{P}, s) := \sum_{i=0}^{m-1} q_{i,N}(s) \mathbb{1}^T \left(\mathsf{P}^{(i)} \log \frac{\mathbf{A}}{\mathsf{P}^{(i)}} \right) \mathsf{p}^{(i+1)}$$

so that $|F_N(\mathsf{p}, \mathsf{P}, s) - F_{N,m}(\mathsf{p}, \mathsf{P}, s)| \le \sum_{i=m-1}^{\infty} q_{i,N}(s) \log |\mathcal{A}|$ and thus

$$\left| \max_{(\mathsf{p},\mathsf{P}) \in Z_0} F_N(\mathsf{p}, \mathsf{P}, s) - \max_{(\mathsf{p},\mathsf{P}) \in Z_0} F_{N,pN}(\mathsf{p}, \mathsf{P}, s) \right| \le \frac{(d-1) q_{pN-1,N}(s)}{d} \log |\mathcal{A}|. \tag{35}$$

The maximum of $F_{N,pN}(\mathsf{p}, \mathsf{P}, s)$ is found in [2, Proposition 11]:

$$\max_{(\mathsf{p},\mathsf{P}) \in Z_0} F_{N,pN}(\mathsf{p}, \mathsf{P}, s) = \max_{\mathsf{p}^{(pN)} = \mathsf{e}_a : a \in \mathcal{A}_0} (\lambda^{(pN),N})^T \mathsf{p}^{(pN)} \tag{36}$$

where

$$\lambda^{(0),N} = 0 \text{ and } \lambda^{(i+1),N} = \begin{cases} q_{i,N}(s)\log(\mathbf{A}^T e^{\frac{\lambda^{(i),N}}{q_{i,N}(s)}}) & \text{if } q_{i,N} \neq 0; \\ 0 & \text{otherwise,} \end{cases}$$

or, equivalently,

$$e^{\frac{\lambda^{(0),N}}{q_{0,N}(s)}} = \mathbb{1} \text{ and } e^{q_{i+1,N}(s)^{-1}\lambda^{(i+1),N}} = \begin{cases} \Psi_{\mathbf{A},\frac{q_{i,N}(s)}{q_{i+1,N}(s)}}(e^{\frac{\lambda^{(i),N}}{q_{i,N}(s)}}) & \text{if } q_{i,N} \neq 0; \\ \mathbb{1} & \text{otherwise.} \end{cases} \tag{37}$$

We stress that since the number

$$\overline{q}_i(s) = q_{i,N}(s) \cdot \left(N + \frac{1}{d^p - 1}\right)$$

is independent of N, so is $\frac{q_{i,N}(s)}{q_{i+1,N}(s)} = \frac{\overline{q}_i(s)}{\overline{q}_{i+1}(s)}$. Thus, the expression $e^{q_{i,N}(s)^{-1}\lambda^{(i),N}}$ is identical for all $N \geq i/p$, and we will then consider the following vector as in Lemma A.1:

$$r(s) = \left(\frac{\overline{q}_p(s)}{\overline{q}_{p+1}(s)}, \frac{\overline{q}_{p+1}(s)}{\overline{q}_{p+2}(s)}, \cdots, \frac{\overline{q}_{2p-1}(s)}{\overline{q}_p(s)}\right), \tag{38}$$

We now prove (25)≤(24). According to its definition, for every $\epsilon > 0$ one may choose $r^* \in \mathcal{R}_{p,d}$ such that

$$\inf_{r\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} {r_i}^{-1}\right) \log \rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r}) < \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{*-1}\right) \log \rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r^*}) + \epsilon.$$

Then, by Lemma A.1, we can find some $s^* \in \Gamma_{[p]}$ such that $r^* = r(s^*)$. As a consequence, we have

$$\left\| e^{\frac{\lambda^{(pN),N}}{q_{pN,N}(s^*)}} \right\| = \left\| \mathcal{L}^{N-1}_{\mathbf{A},r^*}(e^{\frac{\lambda^{(p),N}}{q_{p,N}(s^*)}}) \right\| \leq (\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r^*}))^{N-1} \left\| e^{\frac{\lambda^{(p),N}}{q_{p,N}(s^*)}} \right\|.$$

On account of this, (35) and (37),

$$\begin{aligned} &\limsup_{N\to\infty} \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_N(\mathsf{p},\mathsf{P},s^*) \leq \limsup_{N\to\infty} \max_{\mathsf{p}^{(pN)}=\mathsf{e}_a : a\in\mathcal{A}_0} (\lambda^{(pN),N})^T \mathsf{p}^{(pN)} \\ &\leq \overline{q}_p(s)\log\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r^*}) = \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{*-1}\right)^{-1} \log\rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r^*}) \\ &< \min_{r\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{*-1}\right)^{-1} \log\rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r}) + \epsilon, \end{aligned}$$

where the equality in the second line follows from that

$$\sum_{i=p}^{2p-1} \overline{q}_i(s^*) = 1 \text{ and } \overline{q}_{p+i} = {r_{i-1}^*}^{-1} \cdots r_1^{*-1} r_0^{*-1} \overline{q}_p.$$

To prove (25)≥(24), we note that one can always find a sequence $s(N)$ such that

$$\max_{(\mathsf{p},\mathsf{P})\in Z_0} F_N(\mathsf{p},\mathsf{P},s(N)) = \min_{s\in\Gamma_{[p]}} \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_N(\mathsf{p},\mathsf{P},s),$$

which, as a consequence of (33), admits a convergent subsequence $s(N_i) \in \Gamma_{[p]}$ with its limit s^* satisfying

$$\liminf_{N\to\infty} \min_{s\in\Gamma_{[p]}} \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_N(\mathsf{p},\mathsf{P},s) = \lim_{i\to\infty} \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_{N_i}(\mathsf{p},\mathsf{P},s^*). \tag{39}$$

In addition, we set $r^* = r(s^*)$ as in (38) and $\lambda^{(i),j}$ as in (36) with $s = s^*$. We suppose, according to Proposition 4.1, $v^{(0)} \in \mathcal{C}_0$ is an eigenvector such that $\mathcal{L}_{\mathbf{A},r^*}(v^{(0)}) = \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r^*}) \cdot v^{(0)}$. Similar to (36), we claim that

$$\begin{aligned}\liminf_{N\to\infty} \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_{N,pN}(\mathsf{p},\mathsf{P},s^*) &= \liminf_{N\to\infty} \max_{\mathsf{p}^{(pN)}=\mathsf{e}_a : a\in\mathcal{A}_0} (\lambda^{(pN),N})^T \mathsf{p}^{(pN)} \\ \geq \liminf_{N\to\infty} \max_{\mathsf{p}^{(pN)}=\mathsf{e}_a : a\in\mathcal{A}_0} (\nu^{(pN),N})^T \mathsf{p}^{(pN)} &= \log \rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{A},r^*})\end{aligned} \tag{40}$$

where

$$e^{\frac{\nu^{(0),N}}{q_{0,N}(s^*)}} = \mathbb{1}_{\{a: v_a^{(0)}>0\}} \text{ and } e^{\frac{\nu^{(i+1),N}}{q_{i+1,N}(s^*)}} = \begin{cases} \Psi_{\mathbf{A},\frac{q_{i,N}(s^*)}{q_{i+1,N}(s^*)}}(e^{\frac{\nu^{(i),N}}{q_{i,N}(s^*)}}) & \text{if } q_{i,N} \neq 0; \\ \mathbb{1}_{\mathcal{A}_0} & \text{otherwise.}\end{cases} \tag{41}$$

Indeed, since $\Psi_{\mathbf{A},\frac{q_{i,N}(s^*)}{q_{i+1,N}(s^*)}}$ is order-preserving,

$$e^{\frac{\lambda^{(i+1),N}}{q_{i+1,N}(s^*)}} = \Psi_{\mathbf{A},\frac{q_{i,N}(s^*)}{q_{i+1,N}(s^*)}}(e^{\frac{\lambda^{(i),N}}{q_{i,N}(s^*)}}) \geq \Psi_{\mathbf{A},\frac{q_{i,N}(s^*)}{q_{i+1,N}(s^*)}}(e^{\frac{\lambda^{(i),N}}{q_{i,N}(s^*)}}) = e^{\frac{\nu^{(i+1),N}}{q_{i+1,N}(s^*)}} \text{ for all } i \geq 0,$$

and thus $\nu_a^{(pN),N} \leq \lambda_a^{(pN),N}$ for all $a \in \mathcal{A}_0$. This proves the inequality. Now by definition of $\Psi_{\mathbf{A},\frac{q_{i,N}(s^*)}{q_{i+1,N}(s^*)}}$, $e^{q_{pn_0,N}(s^*)^{-1}\nu^{(p),N}} \in \mathcal{C}_0^+$. Hence, there exists a constant $C > 0$ (independent of N since $e^{q_{pn_0,N}(s^*)^{-1}\nu^{(p),N}}$ is independent of N) such that $e^{q_{pn_0,N}(s^*)^{-1}\nu^{(pn_0),N}} \leq Cv^{(0)}$. Consequently, by the order-preserving property and homogeneity of $\mathcal{L}_{\mathbf{A},r^*}$

$$\left\| e^{q_{pN,N}(s^*)^{-1}\nu^{(pN),N}} \right\| = \left\| \mathcal{L}_{\mathbf{A},r^*}^{N-1}(e^{q_{p,N}(s^*)^{-1}\nu^{(p),N}}) \right\| \geq C \cdot \left\| \mathcal{L}_{\mathbf{A},r^*}^{N-1}(v^{(0)}) \right\|.$$

Therefore,

$$\liminf_{N\to\infty} \max_{(\mathsf{p},\mathsf{P})\in Z_0^+} F_{N,pN}(\mathsf{p},\mathsf{P},s^*) \geq \left(\sum_{\ell=0}^{p-1} \prod_{i=0}^{\ell} r_i^{*-1} \right)^{-1} \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r^*}),$$

and last equality in (40) is proved. In addition, the argument here also justifies that "min" in (24) is attained. Precisely, for the very same r^*, we see from the part (25)≤(24) that for any $s \in \Gamma_{[p]}$,

$$\begin{aligned}\left(\sum_{\ell=0}^{p-1} \prod_{i=0}^{\ell} r_i^{*-1} \right)^{-1} \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r^*}) &\leq \lim_{i\to\infty} \min_{s\in\Gamma_{[p]}} \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_{N_i}(\mathsf{p},\mathsf{P},s) \\ \leq \limsup_{N\to\infty} \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_N(\mathsf{p},\mathsf{P},s) &\leq \left(\sum_{\ell=0}^{p-1} \prod_{i=0}^{\ell} r_i(s)^{-1} \right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r}),\end{aligned}$$

which is as desired.

We now prove (25)≤(26) with the assumption of irreducibility of $\mathbf{A}$, under which $\mathcal{P}(\mathbf{A})$ is a partition of $\mathcal{A}$. To begin with, we take s^* to be an minimizer of (25) so that a maximizer of (40) regardless of the choice of $\mathsf{p}^{(pN)} \in \mathcal{A}_0$ is

$$\mathsf{P}_{a,b}^{(i)\,*} = \begin{cases} \dfrac{e^{q_{-i,N}(s)^{-1}\nu_a^{(-i),N}}}{\sum_{c:\mathbf{A}_{c,b}=1} e^{q_{-i,N}(s)^{-1}\nu_a^{(-i),N}}} & \text{if } \mathbf{A}_{a,b} = 1, b \in \mathcal{A}_{-i-1}; \\ 0 & \text{otherwise,}\end{cases}$$

which is also independent of N. In particular, according to Lemma 4.1, there exists $w \in \mathcal{C}_j$ and $0 \le \theta < 1$ such that

$$\limsup_{N\to\infty} \left\| \frac{e^{\frac{\nu^{(pN),N}}{q_{pN,N}(s)}}}{\left\| e^{\frac{\nu^{(pN),N}}{q_{pN,N}(s)}} \right\|} - w \right\|^{1/N} = \limsup_{N\to\infty} \left\| \frac{\mathcal{L}^N_{\mathbf{A},r}(e^{\frac{\nu^{(p),N}}{q_{p,N}(s)}})}{\left\| \mathcal{L}^N_{\mathbf{A},r}(e^{\frac{\nu^{(p),N}}{q_{p,N}(s)}}) \right\|} - w \right\|^{1/N} < \theta. \tag{42}$$

By writing

$$w^{(j+1)} = \Psi_{\mathbf{A},r_j}(w) / \left\|\Psi_{\mathbf{A},r_j}(w)\right\|,$$

we define the matrix

$$\mathsf{Q}^{(j)*}_{a,b} = \begin{cases} \frac{w^{(j)}_a}{\sum_{c:c\in\mathcal{A}^+_{j-1}} w^{(j)}_c} & \text{if } w^{(j-1)}_b > 0; \\ 0 & \text{otherwise.} \end{cases} \tag{43}$$

Since

$$\mathcal{L}^n_{\mathbf{A},r}(e^{q_{p,N}(s^*)^{-1}\nu^{(p),N}}) \in \mathcal{C}^+_0 = \{x \in \mathbb{R}^{\mathcal{A}}_{\ge 0} : x_a > 0 \text{ if } a \in \mathcal{A}_0\}$$

for all $n \in \mathbb{Z}_+$ and $x \mapsto x\log x$ is Lipschitz on every compact subset of $(0,1)$, and therefore there exists a constant $C > 0$, and $0 < \theta < 1$ (all independent of N) such that

$$\left| \sum_{a\in\mathcal{A}} \left(\mathsf{P}^{(i)*} \log \frac{1}{\mathsf{P}^{(i)*}}\right)_{a,b} - \sum_{a\in\mathcal{A}} \left(\mathsf{Q}^{(i)*} \log \frac{1}{\mathsf{Q}^{(i)*}}\right)_{a,b} \right| \le C\cdot\theta^n \text{ for } b \in \mathcal{A}_{-i-1}.$$

Now observe that

$$\limsup_{N\to\infty} \max_{\mathsf{p}^{(pN)}=\mathsf{e}_a : a\in\mathcal{A}_0} (\nu^{(pN),N})^T \mathsf{p}^{(pN)} = \limsup_{N\to\infty} (\nu^{(pN),N})^T \mathsf{e}_{a_0},$$

and that $(\mathsf{Q}^*)^{pn}\mathsf{e}_{a_0} \in \mathcal{C}^+_0$ for all sufficiently large n. As a consequence, there exists $(\mathsf{Q}^*)^{pn}\mathsf{e}_{a_0} \in \mathcal{C}^+_0$ converges to $\mathsf{q}^{(0)^*}$ due to [11, Theorem 4.7]. Moreover, the theorem also asserts that if we set $\mathsf{q}^{(pn)^*} = \mathsf{q}^{(0)^*}$ for all n and $\mathsf{q}^{(i)^*} = \mathsf{Q}^{(i)^*}\mathsf{q}^{(i+1)^*} \in \mathcal{C}_i \cap \Gamma_{\mathcal{A}}$, then there exist $C' > C$ and $\theta < \theta' < 1$ such that

$$\mathsf{d}_v\left(\prod_{\ell=i}^{pN-1} \mathsf{Q}^{(\ell)^*}\mathsf{e}_{a_0}, \mathsf{q}^{(0)^*}\right) \le C'\cdot\theta'^{pN-i} \tag{44}$$

so that

$$\begin{aligned} &\mathsf{d}_v\left(\prod_{\ell=i}^{pN-1} \mathsf{P}^{(i)^*}\mathsf{e}_{a_0}, \mathsf{q}^{(0)^*}\right) \\ &\le \mathsf{d}_v\left(\prod_{\ell=i}^{pN-1} \mathsf{P}^{(i)^*}\mathsf{e}_{a_0}, \prod_{i=0}^{pN-1} \mathsf{Q}^{(i)^*}\mathsf{e}_{a_0}\right) + \mathsf{d}_v\left(\prod_{\ell=i}^{pN-1} \mathsf{Q}^{(i)^*}\mathsf{e}_{a_0}, \mathsf{q}^{(0)^*}\right) \\ &\le C'\cdot\frac{\theta'^i}{1-\theta'} + C'\cdot\theta'^{pN-i} \end{aligned}$$

Hence, by writing $\mathsf{p}^{(i)} = \prod_i^{pN-1} \mathsf{e}_{a_0}$, we have

$$\begin{aligned} &|F_{N,pN}(\mathsf{Q}^*, \mathsf{q}^*, s^*) - F_{N,pN}(\mathsf{P}^*, \mathsf{p}^*, s^*)| \\ &\le \sum_{i=0}^{pN-1} q_{i,N}(s)\left(C'\cdot\frac{\theta'^i}{1-\theta'} + 2C'\cdot\theta'^{pN-i}\right) + \sum_{i=pN}^{\infty} q_{i,N}(s)\log|\mathcal{A}| \\ &\le \frac{1}{N}\frac{3C'}{(1-\theta')^2} + \frac{1}{d-1}\frac{1}{N}\log|\mathcal{A}| \to 0 \text{ as } N\to\infty, \end{aligned}$$

and

$$\min_{s\in\Gamma_{[p]}} \max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ \mathsf{P}\text{ is 1-periodic}}} \sum_{m=0}^{p-1} s_m f_m(\mathsf{p},\mathsf{P}) \geq \lim_{N\to\infty} \sum_{m=0}^{p-1} s_m^* \sum_{n=0}^{pN-1} t_{n,N}^* f_m(\mathsf{q}^*,\mathsf{Q}^*)$$
$$\geq \lim_{N\to\infty} \max_{(\mathsf{p},\mathsf{P})\in Z_0} \sum_{m=0}^{p-1} s_m^* \sum_{n=0}^{pN-1} t_{n,N}^* f_{n+m}(\mathsf{p},\mathsf{P})$$

The inequality is then proved.

The rest of the proof is rather straightforward: (25)≥(26) follows by definition, while (26)≥(27) is the weak duality.

As for irreducible $\mathbf{A}$, we may assume without loss of generality that (24) ¿ 0, for if otherwise, $\mathbf{A}$ contains exactly one 1 in each column and in each row, and one can choose $\mathsf{P}^* = \mathbf{A}$ and $\mathsf{p}^{(0)^*} = \mathsf{e}_{a_0}$, where $\mathcal{A}_0 = \{a_0\}$. Under the circumstances, we suppose that if $(s^*,\mathsf{p}^*,\mathsf{P}^*)$ is an optimizer of (27), $v^{(0)^*} \in \Gamma_{\mathcal{A}_0}$ is the associated eigenvector, and r^* is a minimizer of (24). Then, if there exists some $\mathsf{P}_{a',b'} = 0$ such that $\mathbf{A}_{a',b'} = 1$, we may define a matrix

$$\mathbf{A}'_{a,b} = \begin{cases} \mathbf{A}_{a,b} & \text{if } (a,b) \neq (a',b'); \\ 0 & \text{otherwise.} \end{cases}$$

Then, by denoting

$$Z_0' = \{(\mathsf{p},\mathsf{P}) \in Z_0 : \mathsf{P}^{(i)}_{a',b'} = 0 \text{ for all } i \in \mathbb{Z}_+\},$$

we can reproduce (24)≥(25) as before that

$$\min_{r(s)\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i(s)^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A}',r})$$
$$\geq \lim_{N\to\infty} \min_{s\in\Gamma_{[p]}} \max_{(\mathsf{p},\mathsf{P})\in Z_0'} \sum_{m=0}^{p-1} s_m \sum_{n=0}^{pN-1} t_{n,N}^* f_{n+m}(\mathsf{p},\mathsf{P}).$$

Next, we note that (25)≥(26)≥(27) follows naturally by definition. These altogether imply that

$$\min_{r(s)\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i(s)^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A}',r})$$
$$\geq \max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ \mathsf{P}\text{ is 1-periodic}\\ \mathsf{p}\text{ is } p\text{-periodic}}} \min_{s\in\Gamma_{[p]}} \sum_{m=0}^{p-1} s_m f_m(\mathsf{p},\mathsf{P}) = \sum_{m=0}^{p-1} s_m^* f_m(\mathsf{p}^*,\mathsf{P}^*) \tag{45}$$
$$= \min_{r(s)\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i(s)^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r}).$$

However, since $\mathbf{A}$ is irreducible, there exists n such that $\mathcal{L}^n_{\mathbf{A}',r^*}(v^{(0)^*})_a < \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A}',r}) v_a^{(0)^*}$ for all $a \in \mathcal{A}_0$. Now we may apply a Collatz–Wielandt formula [10, Theorem 5.6.1] to deduce that

$$\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A}',r^*}) = \rho_{\mathcal{A}_0}(\mathcal{L}^n_{\mathbf{A}',r^*})^{1/n} \leq \max_{a\in\mathcal{A}_0} \left(\frac{\mathcal{L}^n_{\mathbf{A}',r^*}(v^{(0)^*})_a}{v^{(0)^*}_a}\right)^{1/n} < \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r^*}).$$

This implies

$$\begin{aligned}
&\min_{r(s)\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i(s)^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A}',r}) \\
&\leq \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^*(s)^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A}',r^*}) \\
&< \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^*(s)^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r^*}) \\
&= \min_{r(s)\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i(s)^{-1}\right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{A},r}),
\end{aligned}$$

which contradicts (45). Hence, $\mathsf{P}^*_{a,b} > 0$ if and only if $\mathbf{A}_{a,b} > 0$, which implies the irreducibility. In this case, the only feasible $\mathsf{p}^{(0)} \in \mathcal{C}_0$ is the right eigenvector of $(\mathsf{P}^{(0)})^p$.

References

[1] N. Aubrun and M.-P. Béal, *Decidability of conjugacy of tree-shifts of finite type*, Automata, Languages and Programming, Springer Berlin Heidelberg, 2009, pp. 132–143.

[2] J.-C. Ban and Y.-L. Wu, *On the topological pressure of axial product on trees*, 2023.

[3] A. Banaji and I. Kolossváry, *Intermediate dimensions of Bedford-McMullen carpets with applications to Lipschitz equivalence*, 2021.

[4] I. Benjamini and Y. Peres, *Markov chains indexed by trees*, The Annals of Probability **22** (1994), no. 1.

[5] I. Benjamini and O. Schramm, *Percolation Beyond Z^d, Many Questions And a Few Answers*, Electronic Communications in Probability **1** (1996), no. none, 71 – 82.

[6] T. Berger and Z. Ye, *Entropic aspects of random fields on trees*, IEEE Transactions on Information Theory **36** (1990), no. 5, 1006–1018.

[7] A. Dembo, P. Morters, and S. Sheffield, *Large deviations of Markov chains indexed by random trees*, Annales de l'Institut Henri Poincare (B) Probability and Statistics **41** (2005), no. 6, 971–996.

[8] H. Huang, *The generalized entropy ergodic theorem for nonhomogeneous Markov chains indexed by a homogeneous tree*, Probability in the Engineering and Informational Sciences **34** (2019), no. 2, 221–234.

[9] H. Komiya, *Elementary proof for Sion's minimax theorem*, Kodai Mathematical Journal **11** (1988), no. 1.

[10] B. Lemmens and R. Nussbaum, *Nonlinear Perron-Frobenius theory*, Cambridge University Press, May 2012.

[11] B. Lins, *Convergence of iterates in nonlinear Perron-Frobenius theory*, Discrete and Continuous Dynamical Systems - B **28** (2023), no. 7, 3868–3886.

[12] W. Liu and W. Yang, *The markov approximation of the sequences of n-valued random variables and a class of small deviation theorems*, Stochastic Processes and their Applications **89** (2000), no. 1, 117–130.

[13] W. Peng, W. Yang, and B. Wang, *A class of small deviation theorems for functionals of random fields on a homogeneous tree*, Journal of Mathematical Analysis and Applications **361** (2010), no. 2, 293–301.

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

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

[16] W. Yang, *A class of strong deviation theorems for the random fields associated with nonhomogeneous Markov chains indexed by a Bethe tree*, Stochastic Analysis and Applications **30** (2012), no. 2, 220–237.

[17] W. Yang and W. Liu, *Strong law of large numbers for Markov chains field on a bethe tree*, Statistics & Probability Letters **49** (2000), no. 3, 245–250.