ACTA UNIVERSITATIS OULUENSIS
A Scientiae Rerum Naturalium 811

YU-LIANG WU

STATISTICAL AND GEOMETRIC PROPERTIES OF TREE-SHIFTS AND FRACTALS

A dynamical perspective

Academic dissertation to be presented with the assent of the Doctoral Programme Committee of Technology and Natural Sciences of the University of Oulu for public defence in the OP auditorium (L10), Linnanmaa, on 25 October 2025, at 12 noon

UNIVERSITY OF OULU, OULU 2025

Acta Univ. Oul. A 811, 2025

Supervised by
Professor Meng Wu
Professor Ville Suomala
Associate Professor Sascha Troscheit

Reviewed by
Professor Lingmin Liao
Doctor Thomas Jordan

Opponent
Associate Professor Tuomas Sahlsten

ISBN 978-952-62-4629-1 (Paperback)
ISBN 978-952-62-4630-7 (PDF)

ISSN 0355-3191 (Printed)
ISSN 1796-220X (Online)

Cover Design
Raimo Ahonen

PUNAMUSTA
TAMPERE 2025

Wu, Yu-Liang, Statistical and geometric properties of tree-shifts and fractals. A dynamical perspective
University of Oulu Graduate School; University of Oulu, Faculty of Science
Acta Univ. Oul. A 811, 2025
University of Oulu, P.O. Box 8000, FI-90014 University of Oulu, Finland

Abstract

This thesis investigates the statistical and geometric properties of various symbolic and geometric fractals. Utilizing techniques from dynamical systems, ergodic theory, and large deviation theory, it provides a comprehensive analysis of key questions concerning Markov subshifts on rooted d-trees, self-conformal and self-affine fractals, and their associated measures. The research is organized into four central themes: the thermodynamic formalism of Markov subshifts over d-trees, the recurrence properties in Bedford--McMullen carpets, the decay in the Fourier spectrum of self-similar measures, and the local statistics of conformal measures.

The exploration of thermodynamic formalism for Markov subshifts over d-trees aligns with current research on statistical mechanical systems on trees, providing a large deviation description for d-tree-indexed Markov chains, characterizing pressure, entropy, and dimension in this non-Euclidean setting. Regarding recurrence in fractal systems, this work extends knowledge on recurrence in self-affine fractals by examining Bedford--McMullen carpets with coinciding Hausdorff and box-counting dimensions, providing new insights into quantitative recurrence under non-conformal dynamics. The investigation of Fourier decay in self-similar measures establishes a van der Corput lemma for one-dimensional non-atomic self-similar measures, with applications to metric theorems and absolutely pointwise normality. Finally, the study of local statistics of conformal measures expands the known class of fractal scaling measures by demonstrating that projections of ergodic measures through self-conformal iterated function systems are scaling measures, unveiling deep connections between scaling measures and ergodic theory.

These results collectively contribute to a deeper understanding of the interplay between fractal geometry, probability theory, and dynamical systems, providing novel approaches to addressing problems and offering potential directions for future research.

Keywords: fractal distribution, recurrence, tree-indexed Markov chains, van der Corput lemma

Wu, Yu-Liang, Puusiirto ja fraktaalien tilastolliset ja geometriset ominaisuudet. Dynaaminen näkökulma
Oulun yliopiston tutkijakoulu; Oulun yliopisto, luonnontieteellinen tiedekunta
Acta Univ. Oul. A 811, 2025
Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Tämä väitöskirja tutkii erilaisten symbolisten ja geometristen fraktaalien tilastollisia ja geometrisia ominaisuuksia. Hyödyntäen tekniikoita, jotka liittyvät dynaamisiin systeemeihin, ergodiateoriaan ja suurpoikkeamateoriaan, se tarjoaa kattavan analyysin keskeisistä kysymyksistä, jotka koskevat Markovin siirtoja juurellisilla d-puilla, itse-konformeja ja itse-affiinisia fraktaaleja sekä niihin liittyviä mittoja. Tutkimus on jaettu neljään keskeiseen teemaan: Markovin siirtojen termodynaaminen formalismi d-puilla, toistuvuuden ominaisuudet Bedford--McMullen -matoissa, Fourier-spektrin heikkeneminen itse-similaareille mitoille ja konformien mittojen paikallinen tilastollinen käyttäytyminen.

Markovin siirtojen termodynaamisen formalismin tutkiminen d-puilla on linjassa nykyisen tutkimuksen kanssa tilastollisista mekaanisista systeemeistä puilla, tarjoten suurpoikkeamateoreettisen kuvauksen d-puuindeksoiduille Markovin ketjuille, ja paineen, entropian ja ulottuvuuden karakterisoinnin tässä ei-Euklidisessa ympäristössä. Mitä tulee toistuvuuteen fraktaalisissa järjestelmissä, tämä työ laajentaa tietämystä itse-affiinisten fraktaalien toistuvuudesta tarkastelemalla Bedford--McMullen -mattoja, joiden Hausdorffin dimensio ja laatikkodimensio ovat samat, tuoden uusia näkökulmia kvantitatiiviseen toistuvuuteen ei-konformaalisissa dynamiikoissa. Fourierin heikkenemisen tutkiminen itse-similaareille mitoille luo van der Corput'n lemman yksiulotteisille ei-atomisille itse-similaareille mitoille, jolla on sovelluksia metrisiin teoreemiin ja absoluuttisesti pisteittäiseen normaalisuuteen. Viimeisenä, konformaalisten mittojen paikallisten tilastojen tutkiminen laajentaa tunnettua fraktaalisten skaalautuvien mittojen luokkaa osoittaen, että ergodisten mittojen projisoinnit itse-konformaalisten iteratiivisten funktiosysteemien kautta ovat skaalautuvia mittoja, paljastaen syviä yhteyksiä skaalautuvien mittojen ja ergodiateorian välillä.

Nämä tulokset yhdessä tarjoavat syvemmän ymmärryksen fraktaaligeometrian, todennäköisyysteorian ja dynaamisten systeemien välisestä vuorovaikutuksesta, tarjoten uusia lähestymistapoja avoimiin ongelmiin ja mahdollisia suuntaviivoja tulevalle tutkimukselle.

Asiasanat: fraktaali-jakauma, puuindeksoidut Markovin ketjut, toistuvuus, van der Corput'n lemma

Acknowledgements

I express my deepest gratitude to my supervisors, Professor Meng Wu, Professor Ville Suomala, and Associate Professor Sascha Troscheit. With great patience and constant encouragement you have guided me through this doctoral study. Our conversations were always insightful, and I have greatly benefited from your guidance and support.

My sincere thanks also go to Doctor Thomas Jordan and Professor Lingmin Liao for kindly agreeing to be my pre-examiners, and to Associate Professor Tuomas Sahlsten for accepting the role of opponent in my defense.

I am also grateful to Aleksi Pyörälä, Markus Myllyoja, and Roope Anttila for the countless memorable moments we shared during my time in Oulu. The laughter we shared truly brightened my days.

Special thanks go to the Finnish Academy of Science and Letters, whose support through the Vilho, Yrjö and Kalle Väisälä Fund helped me finalize my dissertation and complete my doctoral training. I sincerely appreciate their generous contribution.

To my dearest friends Lemon, Race, and Kai. Thank you for your unwavering emotional support and encouragement throughout the thesis writing process. I deeply cherish our friendship. Long live our friendships.

Last but not least, I owe my deepest gratitude to my parents. Thank you for your endless love, unconditional support, and belief in me. I could not have made it here without you.

List of original publications

The following is a list of original publications, referred to in the thesis by their Roman numerals (I–V).

I Wu, Y.-L., & Yuan, N. (2024). Quantitative recurrence problem on some Bedford–McMullen carpets. Journal of Mathematical Analysis and Applications, 543(2), 128938. https://doi.org/10.1016/j.jmaa.2024.128938

II Ban, J.-C., & Wu, Y.-L. (2025). On the topological pressure of axial product on trees. Stochastics and Dynamics. https://doi.org/10.1142/s0219493725500078

III Ban, J.-C., Lai, G.-Y., & Wu, Y.-L. (2025). Hausdorff dimensions of irreducible Markov hom tree-shifts. Journal of the London Mathematical Society, 111(6). https://doi.org/10.1112/jlms.70198

IV Algom, A., Chang, Y., Wu, M., & Wu, Y.-L. (2025). Van der Corput and metric theorems for geometric progressions for self-similar measures. Mathematische Annalen. https://doi.org/10.1007/s00208-025-03233-3

V Wu, M., & Wu, Y.-L. (2025). Scaling properties of invariant measures for conformal iterated function systems.

Article I reflects a collaborative effort in research development and writing. The doctoral researcher contributed to the design of the methodology in Articles II and III, as well as the writing. In Article IV, their role focused on proving a key theorem, while the co-authors formulated the research questions. In Article V, the researcher completed the proofs and writing. All authors contributed equally to each work.

Contents

1 Introduction

"What we observe is not nature itself, but nature exposed to our method of questioning."

– Werner Heisenberg

Excavations of the underlying principles of this world have never ceased since the birth of humankind. From the evolution of the universe to changes in weather, the desire for explanations of natural phenomena arises not just out of pure academic curiosity but also because of their practical importance. In retrospect, the study of dynamical systems has undeniably played a crucial role in such pursuits, with advancements often marking new milestones in the field of natural science. A notable example is statistical mechanics, which employs statistical methods and probability theory to study systems comprising a large number of microscopic entities. Its mathematical foundations later contributed to the development of ergodic theory and thermodynamic formalism, whose influences on other mathematical fields remain substantial.

In particular, the aforementioned influences extend to the field of fractal geometry. Inherently, fractals, as sets with fine structures at all scales, have the dynamics of magnification in their blood, which likewise flows through their associated measures. This nature suggests a deep connection with dynamical systems, which is captured and exploited throughout the development of the study as well as within the scope of this thesis. By employing techniques from dynamical systems, ergodic theory, and large deviation theory, the work explores a variety of geometric and statistical properties of symbolic and geometric fractals, with a focus on Markov subshifts over rooted d-trees for the former and self-conformal and self-affine sets for the latter. Categorized into four themes, the geometric and statistical properties in question are studied accordingly in the following projects: the thermodynamic formalism of Markov subshifts over d-trees, the recurrence in Bedford–McMullen carpets, the decay in the Fourier spectrum of self-similar measures, and the scaling properties of ergodic measures for conformal iterated function systems. Sections 1.2–1.5 provide a comprehensive overview of each project, with key findings summarized below.

Thermodynamic formalism of Markov subshifts over d-trees The investigation is motivated by the study of nearest-neighbor statistical mechanical systems on lattices with tree structures, focusing on the pressure and entropy of such systems. We aim to contribute to the ongoing research by supplementing the existing studies with a large deviation description for Markov subshifts and Markov measures. Consequently, the results obtained in the work not only refine the Shannon–McMillan-Breiman theorem by providing an explicit local entropy formula under the irreducible assumption, but also reveal, from a geometric perspective, how the degree d of the underlying lattice influences the entropy and the pressure.

Recurrence in Bedford–McMullen carpets The growing interest in the quantitative study of recurrent points in fractal sets, particularly those that inherit dynamics from iterated function systems, has encouraged this project. Concentrating on Bedford–McMullen carpets with uniform fibers, this study characterizes, in terms of dimension, the recurrence in the sets with respect to their associated $\times m_1$-$\times m_2$ dynamics and a prescribed inhomogeneous rate. The systems discussed in this framework are a representative example of fractal systems with non-conformal dynamics, contributing to the understanding of quantitative recurrence in such settings.

Decay in the Fourier spectrum of self-similar measures At the project's core is the pursuit of a van der Corput estimate for oscillatory integrals with respect to fractal measures—a problem rooted in harmonic analysis. Utilizing Tsujii's [1] large deviation estimate of the Fourier spectrum for self-similar measures, the desired estimate is established and applied to prove the decay in Fourier spectra of totally nonlinear self-conformal measures, as well as to an almost sure pseudo-randomness of the sequence $(\xi x^n)_{n\in\mathbb{N}}$, $\xi \neq 0$, with respect to a non-atomic self-similar measure.

Scaling properties of ergodic measures for conformal iterated function systems This project collects the local statistics of a fractal measure through a continuous magnification process, formulated as a scenery flow. Using ergodic theorems, this project demonstrates that ergodic measures projected via conformal systems are scaling measures with ergodic tangent distributions.

Through exploring the proposed projects, the thesis seeks to demonstrate the inextricably intertwined nature of fractal geometry, probability theory, and dynamical systems, offering inspiring perspectives and ultimately opening new avenues for future research in these fields.

1.1 Notation

Let (X, d_X) be a metric space. Denote the open r-ball centered at $x \in X$ as $B(x, r)$, and let $\mathscr{P}(X)$ represent the collection of Borel probability measures on X.

Shift spaces Throughout the thesis, let $\mathscr{A}$ be a finite alphabet and G be a free monoid with d generators ($d \geq 1$), with $G = \mathbb{N}$ when $d = 1$, a special case. Explicitly, the free monoid with d generators $\Sigma = \{1, 2, \cdots, d\}$ is represented as $G = \cup_{i=0}^{\infty} \Sigma^i$, where $\Sigma^0 := \{\varepsilon\}$ is the set containing the identity element ε. For the remainder of the thesis, the free monoid is identified by its Cayley graph as a rooted d-tree, in which ε is the *root* of the tree. The following is a list of tree-related notations.

Definition 1.1.1. *Let G be a rooted d-tree.*

1. $\Delta_n^{(d)} = \cup_{i=0}^{n} \Sigma^i$ *denotes the initial n-subtree in G.*
2. *For every vertex $g \in G \setminus \{\varepsilon\}$, denote by $\tilde{g}$ the parent of g.*
3. *For $g, h \in G$, write $g \leq h$ if g is on the (unique) path from ε to h.*
4. *For $g, h \in G$, denote by $g \wedge h$ the farthest vertex from ε such that $g \wedge h \leq g$ and $g \wedge h \leq h$.*

For conciseness, dependence on d is suppressed in the above notations when it is clear from the context.

Definition 1.1.2 (full shift)**.** *A G-*full shift *is a topological dynamical system $(\mathscr{A}^G, \sigma)$, where*

$$\mathscr{A}^G = \{(\mathtt{x}_g)_{g \in G} : \mathtt{x}_g \in \mathscr{A} \text{ for all } g \in G\}$$

is endowed with the prodiscrete topology with a compatible metric

$$d_{\mathscr{A}^G}(\mathtt{x}, \mathtt{y}) = \sup\{e^{-|\Delta_n|} : (\mathtt{x}_g)_{g \in \Delta_n} = (\mathtt{y}_g)_{g \in \Delta_n}\},$$

and $\sigma : G \times \mathscr{A}^G \to \mathscr{A}^G$ is a continuous map defined by

$$\sigma(g, \mathtt{x})_h = (\sigma_g \mathtt{x})_h = \mathtt{x}_{gh}.$$

Denote by $\mathscr{P}_\sigma(\mathscr{A}^G)$ the collection of σ-invariant measures and by $\mathscr{P}_{erg}(\mathscr{A}^G)$ the collection of ergodic measures.

As a special case, when $G = \mathbb{N}$, the full-shift is identified as the set of infinite strings

$$\mathscr{A}^{\mathbb{N}} = \{(\mathtt{a}_1 \mathtt{a}_2 \mathtt{a}_3 \cdots) : \mathtt{a}_i \in \mathscr{A} \text{ for all } i \in \mathbb{N}\}$$

with the initial tree $\Delta_n^{(1)} = [1, n+1] \subset \mathbb{N}$ and the shift map defined by

$$\sigma_n(\mathtt{x}_1 \mathtt{x}_2 \cdots) = \mathtt{x}_n \mathtt{x}_{n+1} \cdots.$$

Iterated function systems Let $\mathscr{A}$ be a finite alphabet, $G = \mathbb{N}$, and (X, d_X) be a complete metric space. For open $U, V \subset \mathbb{R}^n$ and $0 < \gamma < 1$, let $C^{1+\gamma}(U,V)$ (respectively, $C^{1+\gamma}(U)$) denote the class of continuously differentiable functions $f : U \to V$ (respectively, $f : U \to \mathbb{R}$) whose derivatives $D_x f$ are γ-Hölder continuous. Similarly, let $C^{\omega}(U,V)$ (respectively, $C^{\omega}(U)$) be the collection of analytic functions $f : U \to V$ (respectively, $f : U \to \mathbb{R}$).

Definition 1.1.3 (iterated function system)**.** *Let $\Phi = (f_{\mathtt{a}})_{\mathtt{a}\in\mathscr{A}}$ be a finite collection of functions $f_{\mathtt{a}} : X \to X$.*

1. *Φ is said to be an* iterated function system (IFS) *if each $f_{\mathtt{a}}$ is a contraction. Under the circumstances, there exists a non-empty compact set $K \subset X$ satisfying*
$$K = \bigcup_{\mathtt{a}\in\mathscr{A}} f_{\mathtt{a}}(K),$$
referred to as the attractor *of $(f_{\mathtt{a}})_{\mathtt{a}\in\mathscr{A}}$.*
2. *A* self-affine IFS *is an IFS consisting of linear maps of the form $f(x) = Ax + t$ for some invertible matrix $A \in GL_n(\mathbb{R})$ and vector $t \in \mathbb{R}^n$. The attractor of the IFS is referred to as a* self-affine set*.*
3. *A* self-similar IFS *is a self-affine IFS whose defining maps are of the form $f(x) = r \cdot Ax + t$ for some $0 < r < 1$, orthogonal matrix $A \in O_n(\mathbb{R})$ and vector $t \in \mathbb{R}^n$. Its associated attractor is referred to as a* self-similar set*.*
4. *A* self-conformal IFS *Φ is an IFS consisting of injective conformal maps $f \in C^{1+\gamma}(U, U)$ on a bounded open convex domain $U \subset \mathbb{R}^n$ satisfying $\|D_x f\|^{-1} D_x f \in O_n(\mathbb{R})$ and $\sup_{x\in U} \|D_x f\| < 1$. Its associated attractor is referred to as a* self-conformal set*.*

Equivalently, the attractor of an IFS Φ can be described by the canonical projection $\Phi : \mathscr{A}^{\mathbb{N}} \to X$
$$\Phi(\mathtt{x}) = \lim_{n\to\infty} f_{\mathtt{x}_1} \circ f_{\mathtt{x}_2} \circ \cdots \circ f_{\mathtt{x}_n}(*),$$
where $*$ is an arbitrary point in X. With the canonical projection, one can define the associated measures of the IFSes as follows.

Definition 1.1.4. *Let Φ be an IFS in $\mathbb{R}^n$ and $\mu \in \mathscr{P}_{erg}(\mathscr{A}^{\mathbb{N}})$ be a Bernoulli measure. Then, the pushforward measure $\Phi\mu$ is referred to as a* self-affine/self-similar/self-conformal measure *if Φ is self-affine/self-similar/self-conformal.*

Definition 1.1.5 (Bedford–McMullen carpet)**.** *Let $2 \le m_1 \le m_2$ be two integers and $\mathscr{A} \subset \{(\mathtt{a}^{(1)}, \mathtt{a}^{(2)}) \in \mathbb{Z}^2 : 0 \le \mathtt{a}^{(1)} < m_1, 0 \le \mathtt{a}^{(2)} < m_2\}$. A* Bedford–McMullen carpet *is a self-affine set $K \subset \mathbb{R}^2$ associated with some IFS $\Phi = (f_{\mathtt{a}})_{a\in\mathscr{A}}$ of the form*
$$f_{\mathtt{a}}(x) = \begin{bmatrix} \frac{1}{m_1} & 0 \\ 0 & \frac{1}{m_2} \end{bmatrix} x + \begin{bmatrix} \frac{\mathtt{a}^{(1)}}{m_1} \\ \frac{\mathtt{a}^{(2)}}{m_2} \end{bmatrix}.$$

Dimensions Let (X, d_X) be a metric space. The following dimensions of a set are considered in this thesis.

Definition 1.1.6 (box-counting dimension)**.** *The* upper/lower box-counting dimension *or* upper/lower Minkowski dimension *of a subset $S \subset X$ is respectively defined by*

$$\overline{\dim}_B S = \limsup_{r \to 0^+} \frac{\log \mathcal{N}_r(S)}{\log r} \quad \textit{and} \quad \underline{\dim}_B S = \liminf_{r \to 0^+} \frac{\log \mathcal{N}_r(S)}{\log r},$$

where $\mathcal{N}_r(S)$ is the minimal number of r-balls needed to cover the set S.

As an alternative measurement of subsets in X, the Hausdorff and packing dimensions intend to capture the "critical dimension" for the associated Hausdorff measure and packing measure in the following manner, both of which turn out to be Borel measures on the metric space; see [2, Chapter 2] for a discussion.

Definition 1.1.7 (Hausdorff dimension)**.** *The s-*dimensional Hausdorff measure *of a set S is an outer measure defined by*

$$H^s(S) = \lim_{\delta \to 0^+} H^s_\delta(S),$$

where

$$H^s_\delta(S) = \inf\left\{ \sum_{i=1}^\infty \operatorname{diam}(U_i) : S \subseteq \bigcup_{i=1}^\infty U_i \textit{ with } \operatorname{diam}(U_i) < \delta \right\}.$$

The Hausdorff dimension *of a set S is defined as the critical exponent for the Hausdorff measure to be finite-valued:*

$$\dim_H S = \inf\{s > 0 : H^s(S) = 0\} = \sup\{s > 0 : H^s(S) = \infty\}.$$

Definition 1.1.8 (packing dimension)**.** *The* s-dimensional packing pre-measure *of a set $S \subset X$ is defined as*

$$P^s_0(S) = \limsup_{\delta \to 0} P^s_\delta(S),$$

where

$$P^s_\delta(S) = \sup\left\{ \sum_{i=1}^\infty \operatorname{diam}(B_i) : B_i \textit{ are disjoint closed balls centered in } S, \operatorname{diam}(B_i) \le \delta \right\}.$$

*The s-*dimensional packing measure *is defined as*

$$P^s(S) = \inf\left\{ \sum_{i=1}^\infty P^s_0(S_i) : S \subset \bigcup_{i=1}^\infty S_i \right\},$$

and the packing dimension *of a set S is defined as the critical exponent for the packing measure to be finite-valued:*

$$\dim_P S = \inf\{s > 0 : P^s(S) = 0\} = \sup\{s > 0 : P^s(S) = \infty\}.$$

As artificial as Hausdorff measure and packing measure might seem, they prove natural in the pursuit of generalizing the Lebesgue measure, especially for characterizing Lebesgue-null Borel subsets—they indeed coincide, up to some multiplicative constant, with Lebesgue measure when s is integral.

Similar concepts are adopted to define the dimensions of measures. Below are some of these dimensions.

Definition 1.1.9 (local dimension)**.** *Let* $\mu \in \mathscr{P}(X)$. *The* upper/lower local dimension *of* μ *at* $x \in X$ *is defined, respectively, as*

$$\overline{\dim}_{loc}(\mu, x) = \limsup_{r \to 0} \frac{\log \mu(B(x,r))}{\log r} \quad \textit{and} \quad \underline{\dim}_{loc}(\mu, x) = \liminf_{r \to 0} \frac{\log \mu(B(x,r))}{\log r}.$$

Definition 1.1.10 (Hausdorff dimension)**.** *Let* $\mu \in \mathscr{P}(X)$. *The* upper/lower Hausdorff dimension *of* μ *is defined, respectively, as*

$$\overline{\dim}_H \mu = \operatorname*{ess\,sup}_{x \sim \mu} \overline{\dim}_{loc}(\mu, x) \quad \textit{and} \quad \underline{\dim}_H \mu = \operatorname*{ess\,inf}_{x \sim \mu} \underline{\dim}_{loc}(\mu, x).$$

Definition 1.1.11 (packing dimension)**.** *Let* $\mu \in \mathscr{P}(X)$. *The* upper/lower packing dimension *of* μ *is defined, respectively, as*

$$\overline{\dim}_P \mu = \operatorname*{ess\,sup}_{x \sim \mu} \overline{\dim}_{loc}(\mu, x) \quad \textit{and} \quad \underline{\dim}_P \mu = \operatorname*{ess\,inf}_{x \sim \mu} \overline{\dim}_{loc}(\mu, x).$$

Definition 1.1.12 (entropy dimension)**.** *Let* $\mu \in \mathscr{P}(X)$. *The* upper/lower entropy dimension *of* μ *is defined, respectively, as*

$$\overline{\dim}_e \mu = \limsup_{r \to 0} \int \frac{\log \mu(B(x,r))}{\log r}\, d\mu(x) \textit{ and } \underline{\dim}_e \mu = \liminf_{r \to 0} \int \frac{\log \mu(B(x,r))}{\log r}\, d\mu(x).$$

When $X = \mathbb{R}^d$, the relationship between dimensions of a set and those of a measure is discussed in [2, Chapter 10.1] and is summarized as follows.

Proposition 1.1.13. *If* $\mu \in \mathscr{P}(\mathbb{R}^d)$, *then the following hold.*

$$\overline{\dim}_H \mu = \inf\{\overline{\dim}_H S : \mu(S) = 1\} \quad \textit{and} \quad \underline{\dim}_H \mu = \inf\{\underline{\dim}_H S : \mu(S) > 0\},$$

$$\overline{\dim}_P \mu = \inf\{\overline{\dim}_P S : \mu(S) = 1\} \quad \textit{and} \quad \underline{\dim}_P \mu = \inf\{\underline{\dim}_P S : \mu(S) > 0\}.$$

Markov chains, entropy, and pressure The Markov chains considered in this thesis are those indexed by d-trees G, as defined below.

Definition 1.1.14 (Markov chain)**.** *Let G be a rooted d-tree. A G-indexed Markov chain $(\mathtt{X}_g)_{g\in G}$ is a stochastic process satisfying the* Markov property*: There exists a stochastic matrix $M \in \mathbb{R}_{\geq 0}^{\mathscr{A}\times\mathscr{A}}$ such that for all $g \in G$ and all $\mathtt{a},\mathtt{b} \in \mathscr{A}$,*

$$\mathbb{P}(\mathtt{X}_g = \mathtt{b} | \mathtt{X}_{\tilde{g}} = \mathtt{a}, \mathtt{X}_h \text{ for } g \wedge h \leq \tilde{g}) = \mathbb{P}(\mathtt{X}_g = \mathtt{b} | \mathtt{X}_{\tilde{g}} = \mathtt{a}) = M_{\mathtt{a},\mathtt{b}}.$$

In this context, the stochastic matrix $(M_{\mathtt{a},\mathtt{b}})_{\mathtt{a},\mathtt{b}\in\mathscr{A}\times\mathscr{A}}$ is referred to as the transition matrix*, and the probability vector $(\pi_{\mathtt{a}})_{\mathtt{a}\in\mathscr{A}} := (\mathbb{P}(\mathtt{X}_\varepsilon = \mathtt{a}))_{\mathtt{a}\in\mathscr{A}}$ is referred to as the* initial distribution*.*

Notably, a concrete construction of a Markov chain can be obtained through the Borel probability space $(\mathscr{A}^G, \mathscr{B}, \mathbb{P})$, where each $\mathtt{X}_g(\mathtt{x})$ is defined as $\mathtt{x}_g$ and the measure $\mathbb{P} \in \mathscr{P}(\mathscr{A}^G)$ is called the *Markov measure*. Moreover, these processes generalize one-sided Markov chains, as they reduce to the latter when $G = \mathbb{N}$ for $d = 1$. The following two important classes of Markov chains are considered in this thesis.

Definition 1.1.15. *Let $(\mathtt{X}_g)_{g\in G}$ be a d-tree-indexed Markov chain with transition matrix M.*

1. *$(\mathtt{X}_g)_{g\in G}$ is said to be* irreducible *if its transition matrix M is irreducible, namely, for every $\mathtt{a},\mathtt{b} \in \mathscr{A}$, there exists some $n \in \mathbb{N}$ such that $(M^n)_{\mathtt{a},\mathtt{b}} > 0$.*
2. *$(\mathtt{X}_g)_{g\in G}$ is labeled* primitive *if its transition matrix M is primitive, meaning for every $\mathtt{a},\mathtt{b}$, $(M^n)_{\mathtt{a},\mathtt{b}} > 0$ for all sufficiently large $n \in \mathbb{N}$.*

Naturally associated with the support of a Markov measure $\mathbb{P}$ is a Markov subshift on d-tree, which is defined as follows:

Definition 1.1.16 (Markov hom tree-shift[1])**.** *Let $A \in \mathbb{R}_{\geq 0}^{\mathscr{A}\times\mathscr{A}}$. A* Markov hom tree-shift *$\mathscr{T}_A^{(d)}$ on the d-tree associated with A is defined as*

$$\mathscr{T}_A^{(d)} = \{\mathtt{x} \in \mathscr{A}^T : A_{\mathtt{x}_{\tilde{g}},\mathtt{x}_g} > 0 \text{ for all } g \in G \setminus \{\varepsilon\}\}.$$

With these definitions in place, one can define pressure, energy, and entropy associated with Markov measures and Markov hom tree-shifts.

Definition 1.1.17. *Given an interaction matrix $A \in \mathbb{R}_{\geq 0}^{\mathscr{A}\times\mathscr{A}}$ and a measure $\mu \in \mathscr{P}(\mathscr{T}_A^{(d)})$, denote the set of admissible blocks in $\mathscr{T}_A^{(d)}$ of height n as*

$$B_n(\mathscr{T}_A^{(d)}) = \{(\mathtt{x}_g)_{g\in\Delta_n^{(d)}} : \mathtt{x} \in \mathscr{T}_A^{(d)}\},$$

[1] The term "hom-shifts" is introduced in [3], which refers to the $\mathbb{Z}^d$-shift spaces that arise as the space of graph homomorphisms from $\mathbb{Z}^d$ to a prescribed graph.

and define for every $n \in \mathbb{N}$ a function $\phi_{A,n}^{(d)} : \mathscr{T}_A^{(d)} \to \mathbb{R}$

$$\phi_{A,n}^{(d)}(\mathtt{x}) = \frac{1}{|\Delta_n^{(d)}|} \sum_{g \in \Delta_n^{(d)} \setminus \{\varepsilon\}} \log A_{\mathtt{x}_{\tilde{g}}, \mathtt{x}_g}.$$

1. *The* pressure *is defined as*

$$\mathbf{P}_A^{(d)} = \limsup_{n \to \infty} \frac{1}{|\Delta_n^{(d)}|} \log \left(\sum_{\mathtt{u} \in B_n(\mathscr{T}_A^{(d)})} \prod_{g \in \Delta_n^{(d)} \setminus \{\varepsilon\}} A_{\mathtt{u}_{\tilde{g}}, \mathtt{u}_g} \right).$$

   When $A \in \{0,1\}^{\mathscr{A}}$ is a 0-1 *matrix, then $h_{top}(\mathscr{T}_A^{(d)}) = \mathbf{P}_A^{(d)}$ is called the* topological entropy *of $\mathscr{T}_A^{(d)}$.*

2. *The* energy *of μ is defined as*

$$\phi_A^{(d)}(\mu) = \limsup_{n \to \infty} \int \phi_{A,n}(\mathtt{x})\, d\mu(\mathtt{x}).$$

3. *The* entropy *of μ is defined as*

$$h^{(d)}(\mu) = \limsup_{n \to \infty} \frac{1}{|\Delta_n^{(d)}|} \sum_{u \in B_n(\mathscr{T}_A^{(d)})} -\mu[\mathtt{u}] \log \mu[\mathtt{u}],$$

   where $[\mathtt{u}] = \{\mathtt{x} \in \mathscr{A}^G : (\mathtt{x}_g)_{g \in \Delta_n^{(d)}} = \mathtt{u}\}$ is the cylinder set associated with the block $\mathtt{u} \in B_n(\mathscr{T}_A^{(d)})$.

4. *The* upper/lower local entropy *of μ is defined as*

$$\overline{h}^{(d)}(\mu, \mathtt{x}) = \limsup_{n \to \infty} \frac{-\log \mu[\mathtt{x}|_{\Delta_n^{(d)}}]}{|\Delta_n^{(d)}|} \quad \textit{and} \quad \underline{h}^{(d)}(\mu, \mathtt{x}) = \liminf_{n \to \infty} \frac{-\log \mu[\mathtt{x}|_{\Delta_n^{(d)}}]]}{|\Delta_n^{(d)}|}.$$

The dependence on the degree d is again suppressed when it is clear from the context. Notably, [4] has proven that the pressure always converges in the limit. As [5] notes, the limits in the definitions of entropy and energy exist if μ is a homogeneous measure, or specifically, measures invariant under all automorphisms of G. This includes the class of homogeneous Markov measures, or, more explicitly, those satisfying $\pi M = \pi$.

Since the attention of this thesis is paid exclusively toward the Markov measures, it is noteworthy that for these measures the entropy can be alternatively expressed as

$$h^{(d)}(\mu) = \limsup_{n \to \infty} \int \phi_{M,n}^{(d)}(\mathtt{x})\, d\mu(\mathtt{x}).$$

The formulation unifies the investigation of entropy and energy by studying the ergodic average $\phi_{A,n}$ for a nonnegative interaction matrix A, recasting an asymptotic characterization of local entropy as an ergodic theorem.

Notably, when $G = \mathbb{N}$, a folklore theorem on the interrelation of the introduced notions is the *variational principle* due to Ruelle (see, e.g., [6, Theorem 12.1]), whereas the local behavior of the entropy and energy are characterized by the *Shannon–McMillan–Breiman theorem* ([6, Theorem 1.2.6]) and the *Birkhoff Ergodic Theorem* ([6, Theorem 9.3.1]), respectively. These results are summarized below.

Theorem 1.1.18 (Ruelle)**.** *Let $A \in \mathbb{R}_{\geq 0}^{\mathscr{A}\times\mathscr{A}}$ be a primitive matrix, $\mathscr{T}_A^{(1)} \subseteq \mathscr{A}^{\mathbb{N}}$ be the Markov subshift associated with A, and $\phi : \mathscr{T}_A^{(1)} \to \mathbb{R}$ be a Hölder continuous function. Then, there exists a unique equilibrium state μ for ϕ; namely, μ is the unique σ-invariant realizing*

$$\mathbf{P}_\phi^{(d)} = \sup\left\{ h^{(1)}(\mu) + \int \phi\, d\mu : \mu \in \mathscr{P}_\sigma(\mathscr{T}_A^{(d)}) \right\},$$

where

$$\begin{aligned}\mathbf{P}_\phi^{(d)} &= \limsup_{n\to\infty} \frac{1}{|\Delta_n^{(1)}|} \log\left(\sum_{\mathrm{u}\in B_n(\mathscr{T}_A^{(1)})} \sup_{\mathrm{x}\in[\mathrm{u}]} \exp\left(\sum_{g\in\Delta_n^{(1)}} \phi(\sigma_g(\mathrm{x}))\right)\right) \\ &= \limsup_{n\to\infty} \frac{1}{n} \log\left(\sum_{\mathrm{u}\in B_{n-1}(\mathscr{T}_A^{(1)})} \sup_{\mathrm{x}\in[\mathrm{u}]} \exp\left(\sum_{i=1}^{n} \phi(\sigma_i(\mathrm{x}))\right)\right)\end{aligned}$$

Moreover, μ is supported on $\mathscr{T}_A^{(1)}$ and is a Gibbs measure.

Theorem 1.1.19 (Shannon–McMillan–Breiman)**.** *Let $\mu \in \mathscr{P}_\sigma(\mathscr{T}_A^{(1)})$. Then, the upper and lower local entropy agree μ-almost everywhere as a μ-integrable function. Furthermore, if μ is ergodic, then the limit is $h^{(1)}(\mu)$ μ-almost everywhere.*

Theorem 1.1.20 (Birkhoff)**.** *Let $\mu \in \mathscr{P}_\sigma(\mathscr{T}_A^{(1)})$ and ϕ be μ-integrable. Then, the ergodic averages*

$$\lim_{n\to\infty} \frac{1}{|\Delta_n^{(1)}|} \sum_{g\in\Delta_n^{(1)}} \phi(\sigma_g(\mathrm{x})) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \phi(\sigma_i(\mathrm{x}))$$

converge μ-almost everywhere to a μ-integrable function. Furthermore, if μ is ergodic, then the limit equals $\int \phi\, d\mu$ μ-almost everywhere.

Scenery flows Let (X, d_{LP}) be a metric space with $X = \{\mu \in \mathscr{P}(B(0,1)) : 0 \in \operatorname{supp}\mu\}$ a subset of Borel measures in $\mathbb{R}^n$ and d_{LP} the *Lévy-Prohorov metric* defined by

$$d_{\mathrm{LP}}(\mu,\nu) := \inf\{\varepsilon : \mu(A) \leq \nu(A^{(\varepsilon)}) \text{ and } \nu(A) \leq \mu(A^{(\varepsilon)}), A \in \mathscr{B}(X)\},$$

where $A^{(\varepsilon)} = \{x \in \mathbb{R}^n : \inf_{y\in A} d_{\mathrm{LP}}(x,y) < \varepsilon\}$ is the ε-neighborhood of A. For every $\mu \in X$, define the *magnification* of μ at $x \in B(0,1)$ by e^t by

$$\mu_{x,t}(A) = \frac{\mu(e^{-t}(A \cap B(0,1)) + x)}{\mu(B(x,e^{-t}))},$$

using which the scenery flow is defined as follows.

Definition 1.1.21 (scenery flow)**.** *Let* $\mu \in \mathscr{P}(\mathbb{R}^d)$*. The* scenery flow *of* μ *at* x *is defined as*

$$\frac{1}{T}\int_0^T \delta_{\mu_{x,t}} dt, T \geq 0,$$

where $\delta_\cdot$ *is the Dirac delta function. Any accumulation point of the scenery flow is referred to as a* tangent distribution.

Definition 1.1.22 (scaling measure)**.** *Let* $\mu \in \mathscr{P}(\mathbb{R}^d)$*. Then,* μ *is referred to as a* scaling measure *if the scenery flow converges at* μ*-a.e.* x*, and is referred to as a* uniformly scaling measure *if it is a scaling measure with a unique tangent distribution.*

This thesis focuses on a specific class of tangent distributions, defined as follows.

Definition 1.1.23 (fractal distribution)**.** *Let* X *be equipped with the* scaling flow $S : \mathbb{R}_{\geq 0} \times X \to X$ *defined as* $S_t\mu := \mu_{0,t}$*, so that* (X,S) *forms a topological dynamical system.*

1. $P \in \mathscr{P}(X)$ *is said to be quasi-Palm if the following is satisfied: If* $\mathscr{S}$ *is a* P*-full Borel set, then for* P*-a.e* $\mu \in \mathscr{S}$ *and* μ*-a.e.* x *satisfying* $B(x,e^{-t}) \subset B(0,1)$*, we have* $\mu_{x,t} \in \mathscr{S}$*.*
2. *A distribution* $P \in \mathscr{P}(X)$ *is referred to as a* fractal distribution *if* P *is* S*-invariant and quasi-Palm.*
3. *An* ergodic fractal distribution P *is a fractal distribution that renders* (X,S,P) *ergodic.*

Indeed, [7, Theorem 1.7] indicates that given every $\mu \in \mathscr{P}(\mathbb{R}^n)$, for μ-a.e. x, all tangent distributions are fractal distributions.

1.2 Thermodynamic formalism of Markov subshifts over d-trees

The study of equilibria in statistical physics has a long history, with its foundations laid in Gibbs' seminal works [8, 9] on equilibrium thermodynamics and statistical mechanics. As the theory evolved, a mathematical framework inspired by equilibrium statistical mechanics gradually developed into an independent field known as thermodynamic formalism, which Ruelle [10] and others notably advanced in the 1970s. At its core, the study focuses on the characterization of equilibrium states, drawing implications for the macroscopic behavior of systems from these states.

Though successfully established for a wide range of systems, principles in classical thermodynamic formalism do not apply universally. In particular, Burton, Pfister, and Steif [5] hinted at a failure of the variational principle for Ising models on rooted d-trees—a fundamental theorem pointing to equilibrium states. The breakdown, among others, indicates that

$$\mathbf{P}_A^{(d)} \geq \sup\left\{h(\mu) + \phi_A(\mu) : \mu \text{ is homogeneous}\right\}. \tag{1}$$

With a possibly strict inequality in general, equation (2) contrasts with the classical variational principle for Ising models on the degenerate rooted tree $\mathbb{N}$. This phenomenon invites a closer inspection of both sides of the inequality and motivates the project.

Immediately, the discrepancy prompts the following question:

Question 1. *How does the tree structure affect the pressure?*

Philosophically, one would expect the pressure to grow with the degree of the tree, as an increase in the degree implies an increase in the freedom of the state at each site. As it turns out, the right-hand side of (1) is shown in the same article to be the spectral radius of E, independent of the underlying lattice structure; nevertheless, the left-hand side, systematically investigated by Petersen and Salama as *topological entropy* [11, 12] when $\phi = 0$ or more generally as *asymptotic pressure* [4], is observed in special cases to grow with the degree of the tree:

Theorem 1.2.1 ([11, Theorem 3.1])**.** *Let $\mathscr{T}_A^{(d)}$ be a golden mean subshift, namely,*

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

Then, $h_{top}(\mathscr{T}_A^{(d)})$ is strictly increasing in d.

Specifically, this sets the goal for this project to verify such monotonicity in general.

Particularly, when boiling down to its simplest form with $\phi = 0$, equation (1) can be paraphrased in terms of dimensions as follows:

$$\underline{\dim}_B(\mathscr{T}_A) \geq \sup\{\underline{\dim}_e \mu : \mu \text{ is homogeneous}\} \tag{2}$$

Leaving aside its dynamical interpretation, the expression seemingly compares two distinct notions from dimension theory: the box-counting dimension on one side and the entropy dimension on the other. To clarify the relationship, we propose examining various dimensions of the space $\underline{\dim}_B \mathscr{T}_A$ individually, leading to the following fundamental question:

Question 2. *What are the Hausdorff and packing dimensions of $\mathcal{T}_A$? What measures, if any, attain* $\sup\{\underline{\dim}_H\mu : \mu \in \mathcal{P}(\mathcal{T}_A)\}$ *and* $\sup\{\underline{\dim}_P\mu : \mu \in \mathcal{P}(\mathcal{T}_A)\}$*? Do these dimensions coincide with* $\underline{\dim}_B\mathcal{T}_A$*?*

Unlike the entropy investigated in (1), Question 2 provides further insights into the local structure of measures. The information needed is provided in statistical mechanics and dynamical systems through the *Shannon–McMillan–Breiman theorem*, also known as the *asymptotic equipartition property* in information theory. The theorem elaborates on the contribution of each typical configuration regarding local entropy or, in dimensional terminology, the local dimension of the measure. Particularly, recent developments of the theorem have paid extensive attention to Markov chains on tree-indexed systems [13, 14, 15, 16, 17], as Burton *et al.* highlighted, for the right-hand side of equation (1) can be attained by Markov measures. In summary, the contribution of the previously mentioned works was an almost sure convergence of the sample mean

$$\phi_{A,n}(\mathtt{X}) = \frac{1}{|\Delta_n|} \sum_{g \in \Delta_n \setminus \{\varepsilon\}} \log A_{\mathtt{X}_{\tilde{g}},\mathtt{X}_g} \tag{3}$$

with explicit characterization available only under the primitive assumption. Beyond the framework, its limiting behavior remains largely unexplored, encouraging further investigation into the topic, as explored in this thesis.

In this journey, the thesis follows the route of combinatorial reasoning, which is recognized in large deviation theory as the "method of types" (see, e.g., [18, Chapter 2.1.1] for its classical settings). The argument well catalogs the distribution of Markov chain realizations through the lens of combinatorics, specializing in providing a clear picture of rare events in terms of the following limit:

$$\frac{1}{|\Delta_n|} \log \mathbb{P}(\phi_{A,n}(\mathtt{X}) \in (\alpha, \beta))$$

In implementation, the foundation stone of the method of types is laid in Article II for Markov subshifts on rooted d-trees to address Question 1. Later, a large deviation principle ensues in Article III. As an immediate consequence of the results, for irreducible Markov chains of period p (see Definition 2.1.3), an analog of the law of large numbers is recovered for the conditional sample means $(\phi_{A,pn+j}(\mathtt{X})|\mathtt{X}_\varepsilon = a_0)$, further aiding the derivation of the Hausdorff dimension formula for $\mathcal{T}_A$.

1.3 Recurrence in Bedford–McMullen carpets

The characterization of orbits has long constituted a fundamental pillar in the study of dynamical systems. For instance, in number theory, Diophantine approximation examines the best approximation of a real number through its orbit under the Gauss map, taking the form of the number's continued fraction expansion. In this project, the topic is approached from the perspective of the recurrence set problem, attempting to explore the dimensions of these sets in a system with a fractal structure.

Recurrence as a concept is ubiquitous in dynamical systems, and a comprehensive conception of recurrent points provides key insights into a system's intrinsic structure. A notable example arises in the family of metric measure-preserving systems, for which Poincaré's recurrence theorem asserts almost every point revisits its vicinity infinitely often, applying which Poincaré proved that every bounded Hamiltonian system is stable in the sense of Poisson (see [19, Chapter 1] for a discussion). As a successor of such a qualitative portrayal, a quantitative description of the recurrence was pioneered by Boshernitzan, who offered a uniform bound for the typical return rate:

Theorem 1.3.1 ([20, Theorem 1.2])**.** *Let $(X,\mathscr{B},\mathrm{d},\mu,T)$ be a metric measure-preserving system and $\alpha>0$ be some constant such that α-Hausdorff measure H^α on X is σ-finite. Then, for almost every $x\in X$,*

$$\liminf_{n\to\infty} n^{1/\alpha}\mathrm{d}(x,T^n(x))<\infty.$$

If, in addition, $\mathscr{H}^\alpha(X)=0$, then for almost every $x\in X$,

$$\liminf_{n\to\infty} n^{1/\alpha}\mathrm{d}(x,T^n(x))=0.$$

Barreira and Saussol [21] later refined the theorem, considering the local dimensional structure of X.

Boshernitzan's result naturally spurs the research on the *recurrent set with speed* $\psi:\mathbb{N}\to\mathbb{R}_{>0}$:

$$R(X,T,\psi)=\{x\in X:\mathrm{d}(x,T^n(x))\le\psi(n)\text{ for infinitely many }n\},\tag{4}$$

as one can infer from the theorem that

$$\mu(R(X,T,n^{-1/\alpha}))=\begin{cases}1 & \text{if }\alpha>\dim_H X,\\ 0 & \text{if }\alpha<\dim_H X.\end{cases}$$

The formulation (4) yields a vast number of works characterizing recurrent sets. In the context of metric measure-preserving systems, some even reveal a complete dichotomy; specifically, $\mu(R(X,T,\psi))=1$ if ψ meets certain conditions and $\mu(R(X,T,\psi))=0$ otherwise. Notable contributions to this line of research include [22, 23, 24, 25, 26, 27]. As an alternative approach, researchers also examine the dimension of recurrent sets; results of this kind can be found in [28, 29].

Motivated by the developments, this thesis seeks to investigate the recurrence in fractal systems with dynamics arising from their associated iterated function systems:

Question 3. *Suppose Φ is an IFS with attractor X and $\mu \in \mathscr{P}(X)$*[2]*. Assume that μ-almost every $x = \Phi(\mathtt{x}) \in X$ admits a unique coding in $\mathtt{x} \in \mathscr{A}^{\mathbb{N}}$ so that $T(x) = \Phi(\sigma\mathtt{x})$. What is the size of $R(X,T,\psi)$?*

Within this framework, recent contributions toward the study include [30, 31] on characterizations of $\mu(R(X,T,\psi))$ and [29] regarding $\dim_H R(X,T,\psi)$—all of which deal with self-similar or self-conformal IFSes. Beyond the setting of conformal dynamics, however, characterizing recurrent sets becomes considerably more challenging, even for systems associated with self-affine IFSes. To the author's knowledge, the state-of-the-art results are provided by [26, 32], which focus on matrix transformations on tori, and by [33], which examines the generic self-affine systems regarding translation, but for a slightly different set $\Phi(R(\mathscr{A}^{\mathbb{N}},\sigma,\psi))$.

In addressing Question 3, our eyes are set on recurrence in Bedford–McMullen carpets[3] $X \subset \mathbb{T}^2$, in which case the dynamics $T : X \to X$ can be explicitly expressed as

$$T(x^{(1)},x^{(2)}) = \left(m_1 x^{(1)} \pmod 1, m_2 x^{(2)} \pmod 1\right). \tag{5}$$

The focus is on the dimensions of recurrent sets with different speeds in each coordinate:

$$R_\gamma(X,T,\psi) = \left\{ x \in X : \begin{cases} |(T^n(x))^{(1)} - x^{(1)}| \le \psi(n) \\ |(T^n(x))^{(2)} - x^{(2)}| \le \psi(n)^\gamma \end{cases} \text{for infinitely many } n \in \mathbb{N} \right\}, \tag{6}$$

where $\gamma \in [0, \log_{m_1} m_2]$, with (4) recovered as a special case $\gamma = 1$. Compared to extant results within the self-affine framework, ours is one of the exceptionally rare instances where the dimension of recurrent sets can be determined for Euclidean balls, rather than those projected from the symbolic space through the iterated function system.

To conclude the section, acknowledging the limitations of our approach is important. At its core, the naïve covering strategy employed in the derivation of the upper bound appears suboptimal beyond the aforementioned framework. For Bedford–McMullen carpets with non-uniform fibers, the resulting upper bound fails to align with the lower bound established by the piecewise Bernoulli measures. This discrepancy persists for rectangular neighborhoods with parameter γ exceeding $\log_{m_1} m_2$.

[2]Normally, the measure μ is a projected measure $\pi\overline{\mu}$ via the IFS for some $\mu \in \mathscr{P}_\sigma(\mathscr{A}^{\mathbb{N}})$ or a Hausdorff measure H^α if $0 < H^\alpha(X) < \infty$.

[3]Strictly speaking, Bedford–McMullen carpets are subsets of $[0,1]^2$. Nevertheless, the formulation of associated dynamics is more straightforward if one considers its natural projection on $\mathbb{T}^2$ through the map $(x^{(1)},x^{(2)}) \mapsto (x^{(1)} \pmod 1, x^{(2)} \pmod 1)$.

1.4 Decay in the Fourier spectrum of self-similar measures

Over the decades, Fourier analysis has proven to have numerous applications in scientific studies and in practice alike, ranging from physics to engineering; its impact is equally profound in mathematics itself, particularly in number theory.

In number theory, Weyl [34] initiated the study of equidistributed sequences on $[0,1]$. Precisely, a sequence $(a_n)_{n\in\mathbb{N}}$ in $[0,1]$ is said to be equidistributed on an interval I if it satisfies

$$\lim_{n\to\infty}\frac{|\{1\le k\le n : a_k\in J\}|}{n}=\frac{\operatorname{diam}(J)}{\operatorname{diam}(I)}\quad\text{for every interval } J\subset I.$$

In the spirit of Fourier analysis, Weyl's work provided the following equivalent characterization: For all $\lambda\in\mathbb{N}$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}e^{2\pi i a_n\lambda}=0, \tag{7}$$

paving the way for Fourier-analytic methods in number theory. Philosophically, an equidistributed sequence possesses the nature of identically and independently generated random numbers. This quality alone has numerous desirable consequences, with its application to numerical integration a notable example. Particularly, identity (7), also known as Weyl's criterion, was later applied by Davenport, Erdös, and LeVeque in a paper on the equidistribution of sequences of functions of real numbers, in which the authors found a Lebesgue-almost sure equidistribution given a sufficient decay in the Fourier spectrum of averages of the functions:

Theorem 1.4.1 ([35]). *Suppose $a_n:[a,b]\to\mathbb{R}$, with $a\le x\le b$ and $n\in\mathbb{N}$, is a sequence of real functions. If the series*

$$\sum_{N=1}^{\infty}\frac{1}{N}\int_a^b\left|\frac{1}{N}\sum_{n=1}^{N}e^{2\pi i a_n(x)\lambda}\right|^2dx$$

converges for each $\lambda\in\mathbb{N}$, then the sequence $a_n(x)$ is equidistributed for almost all x in $a<x<b$. Conversely, given any increasing function $(\psi(M))_{M\in\mathbb{N}}$ approaching infinity, there exists a sequence $a_n(x)$ which is not equidistributed for any x yet satisfies the inequality

$$\sum_{N=1}^{M}\frac{1}{N}\int_{\mathrm{a}}^b\left|\frac{1}{N}\sum_{n=1}^{N}e^{2\pi i a_n(x)\lambda}\right|^2dx<\psi(M).$$

Results regarding the almost sure equidistribution of a parametrized family as in Theorem 1.4.1 are referred to as metric theorems (see, e.g., [36]).

On a general level, the condition Erdös *et al.* proposed raises the question of estimating oscillatory integrals, which is of central interest in harmonic analysis. A renowned answer is attributed to van der Corput, stating the following:

Theorem 1.4.2 ([37, Proposition 2 in Chapter 8]). *Let f be a real-valued smooth function on an interval $J \subset \mathbb{R}$. Suppose $|f^{(k)}(x)| \geq 1$ for all $x \in J$, where $k \geq 1$ is an integer. If $k = 1$ and f' is monotonic, or if $k > 1$, then there is a constant c_k, which does not depend on f, such that*

$$\int e^{2\pi i f(x)\lambda} dx \leq c_k \lambda^{-1/k}.$$

As it turns out, a polynomial bound of this type is paramount in many studies, especially in those of fine-scale statistics of sequences initiated by [38]. Compared to the equidistributed property, such statistics describe the behavior of a sequence on the scale of mean gap $1/N$, which can distinguish between different equidistributed sequences, quantify the pseudo-randomness of sequences, and imply equidistribution. See [39, 40, 41] for some examples.

Previously described achievements on the Lebesgue measure light the way for researchers in their counterparts for fractal measures, producing the following question:

Question 4. *Is there a van der Corput lemma for fractal measures?*

Harnessing a large deviation estimate of the Fourier spectrum proved by Tsujii [1], the question is resolved in the affirmative for one-dimensional non-atomic self-similar measures, and its implications are fruitful. First, the estimate guarantees that given any non-atomic self-similar measure μ and any nonlinear analytic function $f : J \to \mathbb{R}$, the pushforward measure $f\mu$ has a polynomial decay in its Fourier spectrum, implying a polynomial Fourier decay for self-conformal measures associated with an analytic IFS containing a nonlinear map [42]. Consequently, this ensures μ-almost every number is a normal number, namely, $(\xi^n x \pmod 1)_{n\in\mathbb{N}}$ is equidistributed for all $\xi \in \mathbb{N}$. Second, a similar argument proves that given any non-atomic self-similar measure μ and any $\xi > 0$, all orders of the correlation of the sequence $(\xi x^n)_{n\in\mathbb{N}}$ are Poissonian (see Definition 2.3.5 for a formal definition) for μ-almost every x, and, in particular, implies the equidistribution of the sequence.

1.5 Scaling properties of ergodic measures for conformal iterated function systems

The ideology of approaching global characterization through local structures is pervasive across mathematical fields, and a geometric realization of the concept is through magnification.

Initiated by Furstenberg through CP-chains [43], geometrical statistics were gathered across scales via a constant multiplier magnifier and analyzed with the aid of ergodic theory. As derivatives, *scenery flows*, introduced by Gavish [44], replace the underpinning machinery with a continuous zoom-in process, and their connection with CP-chains is discussed in [7]. Both tools have proven effective in recent advances in fractal geometry, demonstrating their utility in establishing powerful theorems. Notable examples include [45], which concerns projection theorems, and [46], which illustrates the prevalence of normal numbers with respect to fractal measures. See [7] for a systematic exposition of both tools.

Philosophically, measures with uniform scaling statistics across sites are expected to exhibit identical geometric properties everywhere. For example, uniformly scaling measures are exactly dimensional [7, Proposition 1.19]; namely, $\dim_{\mathrm{loc}} \mu$ is constant almost everywhere. In light of this, posing the following question is logical:

Question 5. *What measures are uniformly scaling or scaling?*

Regarding Question 5, various fractal measures are shown to fall into this class; see [47, 48, 49] for self-similar measures and quasi-Bernoulli measures projected through conformal iterated function systems. Although many fractal measures are known to be uniformly scaling, researchers also provide examples of non-scaling measures (see, e.g., [50, 51]). This project further enriches the collection of fractal scaling measures known to researchers by considering conformal measures $\Phi\mu$ with $\mu \in \mathscr{P}_{erg}(\mathscr{A}^{\mathbb{N}})$, without imposing further separation conditions. Exploiting the ergodicity, this thesis demonstrates that $\Phi\mu$ is a scaling measure with its tangent distribution, up to an orthogonal transformation, unique and ergodic.

To conclude, the study of scaling measures offers deep insights into the geometric and statistical properties of fractal measures. By extending the class of known scaling measures to include conformal measures of ergodic origin, this work enriches the existing theory and provides a broader framework for understanding uniform scaling phenomena. The results presented here not only contribute to ongoing research in fractal geometry but also reinforce the connection between ergodic theory and scaling dynamics. Moreover, the application to Furstenberg-type problems highlights the relevance of these techniques in addressing broader conjectures.

2 Results and discussion

The chapter is dedicated to addressing Questions 1-5. The discussion follows a thematic approach, summarizing the relevant results in Articles I-V.

2.1 Thermodynamic formalism of shift spaces over trees

Questions 1 and 2 are addressed in Articles II and III, respectively.

2.1.1 Results

Regarding Question 1, the exposition begins with a scrutiny of the combinatorics of labeled root d-trees. Inspired by the method of types argument in large deviation theory, Article II reinstates its fundamental combinatorial estimates for the tree-indexed systems, laying the groundwork for the project. Manifested in the paper, the combinatorics are predominantly dictated by a nonlinear operator $\Psi_{A,d} : \mathbb{R}^{\mathscr{A}} \to \mathbb{R}^{\mathscr{A}}$, $A \in \mathbb{R}_{\geq 0}^{\mathscr{A}\times\mathscr{A}}$, defined by

$$(\Psi_{A,d}(x))_{\mathtt{a}} = ((Ax)_{\mathtt{a}})^d,$$

whose interplay with the asymptotic pressure is illustrated as follows.

Theorem 2.1.1 (Theorem 3.1(c) of Article II)**.** *Suppose A has neither zero columns nor zero rows. Let $\mathscr{A}_\infty = \{\mathtt{a} \in \mathscr{A} : \exists n \in \mathbb{N}, (A^n)_{\mathtt{a},\mathtt{a}} > 0\}$ and $L = \min\{n \in \mathbb{N} : (A^n)_{\mathtt{a},\mathtt{a}} > 0,$ $\forall \mathtt{a} \in \mathscr{A}_\infty\}$. Then,*

$$\mathbf{P}_A^{(d)} = \max_{\mathtt{a}\in\mathscr{A}_\infty} \lim_{n\to\infty} \frac{d-1}{d^{Ln+1}} \log(\Psi_{A,d}^{Ln}(x))_{\mathtt{a}}.$$

Notably, no generality is lost by assuming E has neither zero columns nor zero rows, for, as Article II highlighted, the asymptotic pressure remains unchanged by excluding corresponding states. By differentiating $\frac{d-1}{d^{Ln+1}} \log(\Psi_{A,d}^{Ln}(x))_{\mathtt{a}}$ with respect to the degree d unveils the monotonicity of asymptotic pressures, forming the following main theorem of the article.

Theorem 2.1.2 (Theorem 1.1 of Article II)**.** *Suppose A has neither zero columns nor zero rows. Then, $\mathbf{P}_A^{(d)}$ is increasing in d. Furthermore,*

$$\lim_{d\to\infty} \mathbf{P}_A^{(d)} = \max_{\mathtt{a}\in\mathscr{A}} \log \sum_{\mathtt{b}\in\mathscr{A}} A_{\mathtt{a},\mathtt{b}}.$$

Particularly, the theorem resolves Question 1 by confirming that the asymptotic pressure is increasing with respect to degree d of the tree.

As a sequel to Article II, Article III further leverages the established estimates to their fullest extent, resolving Question 2. Serving their original purposes, these estimates enable the development of a Cramér's theorem for finite-state Markov chains indexed by rooted d-trees as stated below. Recall first the following definition of the period of a state $a \in \mathscr{A}$.

Definition 2.1.3. *Suppose $(\mathtt{X}_g)_{g\in G}$ is a Markov chain indexed by the d-tree with transition matrix M.*

1. *The* period *of a state $\mathsf{a} \in \mathscr{A}$ is defined as*
$$p = \gcd\{n \in \mathbb{N} : (M^n)_{\mathsf{a},\mathsf{a}} > 0\},$$
where p is set to ∞ if the index set in question is empty.
2. *If $(\mathtt{X}_g)_{g\in G}$ is irreducible, then the* period *of the Markov chain is the common period of all states. In particular, $(\mathtt{X}_g)_{g\in G}$ is of period 1 if it is primitive.*

In what follows, $(\mathtt{X}_g)_{g\in G}$ is consistently a Markov chain indexed by the d-tree with transition matrix M.

Theorem 2.1.4 (Theorem 1.2 of Article III)**.** *If $\mathsf{a}_0 \in \mathscr{A}$ is a state of finite period p, then $(\phi_{A,pn+j}(\mathtt{X})|\mathtt{X}_\varepsilon = \mathsf{a}_0)_{n\in\mathbb{N}}$ satisfies the large deviation principle with speed $(|\Delta_{pn+j}|^{-1})_{n\in\mathbb{N}}$ and rate*
$$\Lambda_j^*(\alpha) = \sup_{\mu\in\mathbb{R}}[\mu\alpha - \Lambda_j(\mu)],$$
where, with $\mathbb{1} \in \mathbb{R}^{\mathscr{A}}$ denoting the all-1 vector,
$$\Lambda_j(\mu) = \lim_{n\to\infty} \frac{\log((\Psi_{M\odot A^{\odot\mu},d})^{pn+j}(\mathbb{1}))_{\mathsf{a}_0}}{|\Delta_{pn+j}|}.$$
Explicitly, for any Borel subset $S \subset \mathbb{R}$,
$$\sup_{\alpha\in\mathring{S}} -\Lambda_j^*(\alpha) \le \liminf_{n\to\infty} \frac{\log \mathbb{P}(\phi_{A,pn+j}(\mathtt{X}) \in S|\mathtt{X}_\varepsilon = \mathsf{a}_0)}{|\Delta_n|}$$
$$\le \limsup_{n\to\infty} \frac{\log \mathbb{P}(\phi_{A,pn+j}(\mathtt{X}) \in S|\mathtt{X}_\varepsilon = \mathsf{a}_0)}{|\Delta_n|} \le \sup_{\alpha\in\overline{S}} -\Lambda_j^*(\alpha).$$

Almost as a corollary of the previous theorem, a law of large numbers comes along under the irreducible assumption of the Markov chain.

Theorem 2.1.5 (Theorem 1.3 of Article III)**.** *Under the assumptions of Theorem 2.1.4, if the Markov chain is irreducible with period p, then for any state $\mathsf{a}_0 \in \mathscr{A}$,*
$$\lim_{n\to\infty}(\phi_{A,pn+j}(\mathtt{X})|\mathtt{X}_\varepsilon = \mathsf{a}_0) = \sum_{i=0}^{p-1} \frac{d^{-i}}{\sum_{\ell=0}^{p-1} d^{-\ell}} \sum_{\mathsf{a},\mathsf{b}\in\mathscr{A}} \pi_{\mathsf{a}}^{(i+1-j)} M_{\mathsf{a},\mathsf{b}} \log A_{\mathsf{a},\mathsf{b}} \qquad \mathbb{P}\text{-a.s.},$$

where $(\pi^{(i)})_{i\in\mathbb{Z}}$ are the (unique) left probability eigenvectors of M^p satisfying $\pi^{(i)} = \pi^{(i+1)}M$ and $\pi^{(i)}_{a_0} > 0$ if and only if $i \equiv 0 \pmod{p}$.

The explicit expression of the almost sure limits aids the subsequent derivation of the Hausdorff dimension of tree-indexed systems. Precisely, the dimension in question is described in terms of the eigenvalues of the nonlinear transfer operator $\mathscr{L}_{A,r} : \mathbb{R}^{\mathscr{A}} \to \mathbb{R}^{\mathscr{A}}$ ($r \in \mathbb{R}^{\mathscr{A}}_{>0}$), $\mathscr{L}_{A,r} = \Psi_{A,r_{p-1}} \circ \Psi_{A,r_{p-2}} \circ \cdots \circ \Psi_{A,r_0}$ as in the following theorem.

Theorem 2.1.6 (Theorem 1.4 of Article III)**.** *Let $\mathscr{T}_A^{(d)}$ be a Markov hom tree-shift with matrix $A \in \{0,1\}^{\mathscr{A}\times\mathscr{A}}$ an irreducible matrix of period p. Then,*

$$\dim_H \mathscr{T}_A^{(d)} = \min_{r\in\mathscr{R}_{p,d}} \left(\sum_{\ell=0}^{p-1} \prod_{i=0}^{\ell} r_i^{-1} \right)^{-1} \cdot \log \rho_*(\mathscr{L}_{A,r}), \tag{8}$$

where

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

$$\rho_*(\mathscr{L}_{A,r}) = \inf\{\alpha : \mathscr{L}_{A,r}(u) = \alpha \cdot u \in \mathbb{R} \setminus \{0\}\}.$$

In particular, if A is a primitive matrix, then $p = 1$, $\mathscr{L}_{A,r} = A$ for $r \in \mathscr{R}_{1,d} = \{1\}$, and $\dim_H \mathscr{T}_A^{(d)} = \log \rho_*(A) = \log \rho(A)$, where $\rho(A)$ is the spectral radius of A.

In summary, the findings combined answer Question 2. Assuming $A \in \{0,1\}^{\mathscr{A}\times\mathscr{A}}$ is an irreducible incidence matrix, then Article II with Appendix A in Article III asserts

$$\begin{aligned} &\max_{\mu\in\mathscr{P}(\mathscr{T}_A^{(d)})} \overline{\dim}_P \mu = d \cdot \max_{\mu\in\mathscr{P}(\mathscr{T}_A^{(d)})} \underline{\dim}_P \mu \\ &= \dim_P \mathscr{T}_A^{(d)} = \overline{\dim}_B \mathscr{T}_A^{(d)} = d \cdot \underline{\dim}_B \mathscr{T}_A^{(d)} \\ &= \lim_{n\to\infty} \max_{a\in\mathscr{A}} \frac{d-1}{d^{n+1}-1} \log \Psi_{A,d}(\mathbb{1})_a, \end{aligned}$$

which increases as d grows. Furthermore, Article III demonstrates

$$\begin{aligned} &\max_{\mu\in\mathscr{P}(\mathscr{T}_A^{(d)})} \underline{\dim}_H \mu = d^{-1} \max_{\mu\in\mathscr{P}(\mathscr{T}_A^{(d)})} \overline{\dim}_H \mu \\ &= \dim_H \mathscr{T}_A^{(d)} \\ &= \min_{r\in\mathscr{R}_{p,d}} \left(\sum_{\ell=0}^{p-1} \prod_{i=0}^{\ell} r_i^{-1} \right)^{-1} \cdot \log \rho_*(\mathscr{L}_{A,r}), \end{aligned}$$

which, according to the last expression, decreases as d approaches infinity.

2.1.2 Further questions

Since Questions 1 and 2 have been addressed only for nearest-neighbor models on d-trees under the assumption of an irreducible transition matrix, considering the following questions is natural.

- Given the implicit formula of limiting behavior provided in the literature (e.g., [16]), what is the limit of $\phi_{A,n}^{(d)}(\mathtt{X})$ for general Markov chains?
- Is $\Lambda_j^*(\alpha)$ in Theorem 2.1.4 differentiable?
- What is the multifractal spectrum of the Markov measure?
- What is $\lim_{d\to\infty}\dim_H \mathscr{T}_A^{(d)}$?
- Is there a large deviation estimate for non-nearest-neighbor systems on general trees?

2.2 Recurrence in Bedford–McMullen carpets

Regarding Question 3, Article I provides a partial answer.

2.2.1 Results

Throughout the section, let $X\subset\mathbb{T}^2$ denote a Bedford–McMullen carpet with the dynamics $T:X\to X$ (together with $m_1,m_2\in\mathbb{N}$) defined as in (5). The function $\psi:\mathbb{N}\to\mathbb{R}_{>0}$ represents the speed of recurrence and $R_\gamma(X,T,\psi)$ is the recurrent set given in (6). Additionally, define

$$\tau_1=\liminf_{n\to\infty}\frac{\log_{m_1}\psi(n)}{n}\quad\text{and}\quad\tau_2=\liminf_{n\to\infty}\frac{\log_{m_2}\psi(n)}{n}.$$

Recall that Bedford [52] and McMullen [53] independently obtained the box-counting and Hausdorff dimensions of Bedford–McMullen carpets.

Theorem 2.2.1. *Let X be a Bedford-McMullen carpet associated with the alphabet $\mathscr{A}$ as in Definition 1.1.5, and*

$$M=|\{\mathtt{a}^{(1)}:\mathtt{a}\in\mathscr{A}\}|\quad\textit{and}\quad N_{\mathtt{a}^{(1)}}=|\{\mathtt{b}\in\mathscr{A}:\mathtt{b}^{(1)}=\mathtt{a}^{(1)}\}|.$$

Then,

$$\dim_P X=\dim_B X=\log_{m_1}M+\log_{m_2}(|\mathscr{A}|/M),$$
$$\dim_H X=\log_{m_1}\left(\sum_{\mathtt{a}\in\mathscr{A}}N_{\mathtt{a}^{(1)}}^{\log_{m_2}m_1}\right).$$

In light of this, $\dim_B X=\dim_H X$ if and only if there exist M nonempty columns, each containing N rectangles. Throughout, M and N will consistently retain these meanings.

Theorem 2.2.2 (Theorem 1.1 of Article I)**.** *Let X be a Bedford–McMullen carpet satisfying* $\dim_{\mathrm{B}} X = \dim_H X$ *and* $0 \le \gamma \le \log_{m_1} m_2$. *Then,* $\dim_H R_\gamma(X,T,\psi) = \min\{t_1, t_2\}$, *where*

$$t_1 = \begin{cases} \left(1 - \frac{\tau_1 \log_{m_2} m_1}{1+\tau_2}\right) \cdot \log_{m_1} M + \frac{\log_{m_2} N}{1+\tau_2} & \text{if } 1+(1-\gamma)\tau_1 \le \log_{m_1} m_2, \\ \frac{\log_{m_2} M + \log_{m_2} N}{1+\tau_2} & \text{if } 1+(1-\gamma)\tau_1 > \log_{m_1} m_2. \end{cases}$$

$$t_2 = \begin{cases} \frac{\log_{m_1} M}{1+\tau_1} + \log_{m_2} N & \text{if } 1+\tau_1 \le \log_{m_1} m_2, \\ \frac{\log_{m_1} M + \log_{m_1} N}{1+\tau_1} & \text{if } 1+(1-\gamma)\tau_1 \le \log_{m_1} m_2 < 1+\tau_1, \\ \frac{\log_{m_1} M + (1+(1-\gamma)\tau_1)\log_{m_2} N}{1+\tau_1} & \text{if } 1+(1-\gamma)\tau_1 > \log_{m_1} m_2. \end{cases}$$

In particular, if $\gamma = 1$, *this is reduced to the following form.*

(1) *If* $\log_{m_1} m_2 > 1+\tau_1$, *then*

$$\dim_H R_\gamma(X,T,\psi) = \min\left\{\frac{\log_{m_1} M + \log_{m_2} N}{1+\tau_2}, \frac{\log_{m_1} M}{1+\tau_1} + \log_{m_2} N\right\}.$$

(2) *If* $\log_{m_1} m_2 \le 1+\tau_1$, *then*

$$\dim_H R_\gamma(X,T,\psi) = \min\left\{\frac{\log_{m_1} M + \log_{m_2} N}{1+\tau_2}, \frac{\log_{m_1} M + \log_{m_1} N}{1+\tau_1}\right\},$$

Notably, when $\gamma = 1$, the set $R_\gamma(X,T,\psi)$ is a recurrent set associated with balls with respect to the L^∞-distance. Nevertheless, the dimension remains unchanged if one replaces the canonical balls with respect to the L^2-distance, since the two norms are equivalent up to a multiplicative constant that is subsumed into the rate function ψ. In addition, although our formulae formally match those found for shrinking targets by Bárány and Rams [54], the key difference, as Article I Remark 3.1 highlights, lies in the underlying structure of the sets, necessitating the construction of distinct measures in the derivation of lower bounds.

2.2.2 Further questions

As Question 3 remains largely unexplored in the self-affine setting, we would like to ask the following questions:

- What is $R_\gamma(X,T,\psi)$ for $\gamma > \log_{m_1} m_2$? What is the dimension formula if X is a general Bedford–McMullen carpet?
- What is $\dim_H R_\gamma(X,T,\psi)$ for other carpets, e.g., Baránski carpets, Lalley–Gatzouras carpets, Feng–Wang carpets?
- Suppose X is a self-affine set. What is $\mu(R_\gamma(X,T,\psi))$ if μ is a Hausdorff measure or an associated self-affine measure?

2.3 Decay in the Fourier spectrum of self-similar measures

Toward Question 4, Article IV proposed and proved a van der Corput estimate for self-similar measures.

2.3.1 Results

Heuristically, the phase $f : J \to \mathbb{R}$ in the van der Corput lemma shall not be too "flat" as though it were a constant function, for the polynomial estimate would have failed for an obvious reason. Considering this, the following condition on non-flatness of f is apropos.

Definition 2.3.1. *Let $0 < \delta \leq 1$. A function $h : \mathbb{R} \to \mathbb{R}$ is called δ-non-flat on an interval $J \subset \mathbb{R}$ if the following hold: For all $\varepsilon > 0$ there exists $n_0 = n_0(\varepsilon) \geq 1$ such that for all $n > n_0$ and every bounded subset $A \subset J$ we have*

$$\mathcal{N}_{2^{-\delta n}}(A) \geq 2^{\varepsilon n} \Rightarrow \operatorname{diam}(h(A)) \geq 2^{-n}.$$

With the definition, the van der Corput lemma for self-similar measures is stated as follows.

Theorem 2.3.2 (Theorem 1.6 of Article IV)**.** *Let μ be a non-atomic self-similar measure supported on a closed interval $J \subseteq \mathbb{R}$. Let $f \in C^{1+\alpha}(J)$ for some $0 < \alpha \leq 1$. Suppose that f' is δ-non-flat on J. Then there exist constants $\tau = \tau(\mu, \delta, \alpha)$ and $C = C(f) > 0$ such that*

$$\left| \int e^{2\pi i f(x)\lambda} d\mu(x) \right| \leq C\lambda^{-\tau} \quad \textit{for } \lambda > 0.$$

In fact, for smooth phase $f \in C^k(J)$, the non-flatness can be translated into conditions on derivatives, as observed in the following lemma.

Lemma 2.3.3 (Lemma 2.6 of Article IV)**.** *Let $J \subset \mathbb{R}$ be a closed interval and $h \in C^k(J)$ for some integer $k \geq 1$. If*

$$\max_{1 \leq j \leq k} |f^{(j)}(x)| \geq 1 \textit{ for any } x \in J,$$

then there exists $\rho > 0$ such that f is $\frac{1}{2k}$-non-flat on all sub-intervals $I \subset J$ with $\operatorname{diam}(J) \leq \rho$.

Combining the above gives rise to the following theorem, which more closely resembles the original van der Corput lemma.

Theorem 2.3.4 (Theorem 1.1 of Article IV). *Let μ be a non-atomic self-similar measure supported on an interval $J \subseteq \mathbb{R}$. Let $f \in C^k(J)$ for some integer $k \geq 2$ be such that for some $c_0 > 0$, $\max_{2\leq j\leq k} |f^{(j)}(x)| \geq c_0$ for any $x \in J$. Then there exist constants $\tau = \tau(\mu, k)$ and $C = C(f) > 0$ such that*

$$\left| \int_J e^{2\pi i f(x)\lambda} d\mu(x) \right| \leq C\lambda^{-\tau} \quad \text{for } \lambda > 0. \tag{9}$$

Importantly, the first order derivative of phase f is intentionally excluded from the hypothesis, as certain Bernoulli convolutions (see, e.g., [55, Proposition 15.3.2]) exhibit no decay whatsoever in their Fourier spectra, in which case the theorem fails for $g(x) = x$.

The backbone of the theorems above is a large deviation estimate on the Fourier spectra of self-similar measures proved by Tsujii [1, Theorem 1]. By the same argument, Article IV proves the following two applications.

The first application focuses on the following fine-scale statistics of sequences.

Definition 2.3.5. *Suppose $N \geq 1$ and $k \geq 2$.*

1. *Let $\mathscr{U}_k = \mathscr{U}_k(N)$ denote the set of distinct integral k-tuples taking values in $\{1, \ldots, N\}$; that is,*

$$\mathscr{U}_k = \left\{ \boldsymbol{u} = (u_1, \ldots, u_k) : u_i \in \{1, \ldots, N\}, u_i \neq u_j \text{ for all } i \neq j \right\}.$$

For each $\boldsymbol{u} \in \mathscr{U}_k$ and real-valued sequence $(x_n)_{n\geq 1}$ consider the difference vector

$$\Delta(\boldsymbol{u}, (x_n)_{n\geq 1}) = (x_{u_1} - x_{u_2}, \ldots, x_{u_{k-1}} - x_{u_k}) \in \mathbb{R}^{k-1}.$$

Given $f : \mathbb{R}^{k-1} \to \mathbb{R}$ with compact support, the k-level correlation function *is defined as*

$$R_k(f, (x_n)_{n\geq 1}, N) := \frac{1}{N} \sum_{\boldsymbol{u}\in\mathscr{U}_k} \sum_{\boldsymbol{l}\in\mathbb{Z}^{k-1}} f(N(\Delta(\boldsymbol{u}, (x_n)_{n\geq 1}) + \boldsymbol{l})). \tag{10}$$

If

$$\lim_{N\to\infty} R_k(f, (x_n)_{n\geq 1}, N) = \int_{\mathbb{R}^{k-1}} f(\boldsymbol{x}) d\boldsymbol{x}, \; \forall f \in C_c^\infty(\mathbb{R}^{k-1}) \tag{11}$$

then we say the k-level correlation of $(x_n)_{n\geq 1}$ is Poissonian.

2. *Let us reorder the sequence $(x_n)_{n=1}^N$ and label them as*

$$0 \leq \theta_{1,N} \leq \theta_{2,N} \leq \cdots \leq \theta_{N,N} \leq 1 \text{ with } \theta_{0,N} := \theta_{N,N} - 1 \pmod 1.$$

Suppose that for every $s \geq 0$ the limit, as $N \to \infty$, of the function

$$G(s, (x_n)_{n\geq 1}, N) := \frac{1}{N} |\{1 \leq n \leq N : N(\theta_{n,N} - \theta_{n-1,N}) \leq s\}|$$

exists and equals $G(s) = 1 - e^{-s}$. We say that the level spacings are Poissonian.

Each definition above alludes to the similarity between the deterministic sequence $(x_n)_{n\in\mathbb{N}}$ and the one associated with the Poisson process with intensity 1; see [56] for further discussion. In particular, either property implies that $(x_n)_{n\in\mathbb{N}}$ is equidistributed on $[0,1]$. In Article IV, we proved that $(\xi x^n)_{n\in\mathbb{N}}$ is Poissonian in the sense of fine-scale statistics above, augmenting Aistleitner *et al.*'s [56] results.

Theorem 2.3.6 (Theorem 1.2 of Article IV)**.** *Let μ be a non-atomic self-similar measure on $\mathbb{R}$ such that μ-a.e. x is larger than 1. Let $\xi \in \mathbb{R}\setminus\{0\}$. Then for μ-almost every x the k-level correlation of $(\xi x^n)_{n\geq 1}$ is Poissonian for all $k \geq 2$. In particular, the level spacings are also Poissonian, and the sequence $(x^n)_{n\geq 1}$ is equidistributed on $[0,1]$.*

The second application is the completion of Algom *et al.*'s [42] results, asserting a decay in the Fourier spectrum for totally nonlinear conformal measures.

Corollary 2.3.7 (Corollary 1.3 of Article IV)**.** *Let μ be a non-atomic self-similar measure supported on an interval $J \subseteq \mathbb{R}$, and let $f \in C^\omega(J)$ be a non-affine real analytic function. Then there exist $\tau > 0$ and $C > 0$ such that*

$$\left|\int_J e^{2\pi i f(x)\lambda} d\mu(x)\right| \leq C|\lambda|^{-\tau} \text{ as } |\lambda| \to \infty. \tag{12}$$

As [42, Section 6.2] illustrated, the above corollary completes the puzzle of the following theorem.

Theorem 2.3.8 (Theorem 1.4 of Article IV and Corollary 1.2 of [42])**.** *Let Φ be a $C^\omega(\mathbb{R})$ IFS. If Φ contains a non-affine map then every non-atomic self-conformal measure ν admits some $\tau > 0$ such that*

$$\widehat{\nu}(\lambda) = O\left(\frac{1}{|\lambda|^\tau}\right) \text{ as } |\lambda| \to \infty,$$

where $\widehat{\nu}$ is the Fourier transform of ν:

$$\widehat{\nu}(\lambda) = \int e^{2\pi i x\cdot\lambda} d\nu(x).$$

In particular, Baker and Banaji [57] obtain Theorem 2.3.8 simultaneously using a different argument. Furthermore, using an argument essentially identical to the Davenport–Erdös-LeVeque theorem (Theorem 1.4.1), one deduces the pointwise absolute normality of the measure, namely, almost every x (with respect to $f\mu$ or ν) is a normal number with respect to all bases greater than 1.

2.3.2 Further questions

We include the following list of problems that remain unanswered in the thesis.

- Is there a (sharp) bound for the exponent τ in Theorems 2.3.2 and 2.3.4?
- Does the van der Corput lemma hold for self-similar measures in higher dimensions? Does it hold for self-affine measures, given a generalization of Tsujii's large deviation estimate available in higher dimensions [58, Corollary 1.7]?

These questions highlight areas where future work may yield valuable insights or new results.

2.4 Scaling properties of ergodic measures for conformal iterated function systems

Article V provides a resolution to Question 5.

2.4.1 Results

The following theorem answers Question 5.

Theorem 2.4.1 (Theorem 1.1 of Article V)**.** *Suppose Φ is a conformal iterated function system. If $\mu \in \mathscr{P}_{erg}(\mathscr{A}^{\mathbb{N}})$, then $\pi\mu$ is a scaling measure. Moreover, there exists an ergodic fractal distribution $P \in \mathscr{P}(\mathscr{P}(\mathbb{R}^d))$ such that the limit of the scenery flow P_x is of the form $A_x P$ for some orthogonal matrix A_x.*

Essentially, the proof relies on the fact that $\Phi\mu$ is exactly dimensional if $\mu \in \mathscr{P}_{erg}(\mathscr{A}^{\mathbb{N}})$ along with the local entropy average techniques from [45] and an application of Pinsker's inequality.

2.4.2 Further questions

Here we list a few further research problems.

- Theorem 2.4.1 ensures that $\Phi\mu$ is a scaling measure if $\mu \in \mathscr{P}_{erg}(\mathscr{A}^{\mathbb{N}})$ is projected via self-conformal IFS. Is $\Phi\mu$ uniformly scaling if μ is projected via a self-similar IFS?
- Does Theorem 2.4.1 hold if $\mu \in \mathscr{P}_{\sigma}(\mathscr{A})$ is only invariant?
- Are all self-affine measures scaling measures?

References

[1] M. Tsujii, "On the Fourier transforms of self-similar measures," *Dynamical Systems*, vol. 30, no. 4, pp. 468–484, Aug. 2015. [Online]. Available: http://dx.doi.org/10.1080/14689367.2015.1078291

[2] K. Falconer, *Techniques in Fractal Geometry*. Chichester, England: John Wiley & Sons, Mar. 1997.

[3] N. Chandgotia and B. Marcus, "Mixing properties for hom-shifts and the distance between walks on associated graphs," *Pacific Journal of Mathematics*, vol. 294, no. 1, pp. 41–69, 2018.

[4] K. Petersen and I. Salama, "Asymptotic pressure on some self-similar trees," *Stochastics and Dynamics*, vol. 23, no. 02, Dec. 2022. [Online]. Available: http://dx.doi.org/10.1142/S0219493723500090

[5] R. M. Burton, C.-E. Pfister, and J. E. Steif, "The variational principle for Gibbs states fails on trees," *Markov Processes And Related Fields*, vol. 1, no. 3, pp. 387–406, 1995. [Online]. Available: https://infoscience.epfl.ch/handle/20.500.14299/31900

[6] M. Viana and K. Oliveira, *Foundations of Ergodic Theory*, ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2016.

[7] M. Hochman, "Dynamics on fractals and fractal distributions," 2010. [Online]. Available: https://arxiv.org/abs/1008.3731

[8] J. W. Gibbs, "On the equilibrium of heterogeneous substances," *Transactions of the Connecticut Academy of Arts and Sciences*, vol. 3, pp. 108–248, 1876, published October 1875 – May 1876.

[9] ——, "On the equilibrium of heterogeneous substances," *Transactions of the Connecticut Academy of Arts and Sciences*, vol. 3, pp. 343–524, 1878, published May 1877 – July 1878.

[10] D. Ruelle, *Thermodynamic Formalism: The Mathematical Structures of Equilibrium Statistical Mechanics*, 2nd ed. Cambridge University Press, 2004, first edition published in 1978.

[11] K. Petersen and I. Salama, "Tree shift topological entropy," *Theoretical Computer Science*, vol. 743, pp. 64–71, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304397518303724

[12] ——, "Entropy on regular trees," *Discrete and Continuous Dynamical Systems*, vol. 40, no. 7, pp. 4453–4477, Jul. 2020.

[13] W. Yang and W. Liu, "Strong law of large numbers for Markov chains field on a Bethe tree," *Statistics & Probability Letters*, vol. 49, no. 3, pp. 245–250, Sep. 2000. [Online]. Available: https://doi.org/10.1016/s0167-7152(00)00053-5

[14] W. Yang, "Some limit properties for Markov chains indexed by a homogeneous tree," *Statistics & Probability Letters*, vol. 65, no. 3, pp. 241–250, Nov. 2003. [Online]. Available: http://dx.doi.org/10.1016/j.spl.2003.04.001

[15] W. Yang and Z. Ye, "The asymptotic equipartition property for nonhomogeneous Markov chains indexed by a homogeneous tree," *IEEE Transactions on Information Theory*, vol. 53, no. 9, pp. 3275–3280, 2007.

[16] H. Huang, "The generalized entropy ergodic theorem for nonhomogeneous Markov chains indexed by a homogeneous tree," *Probability in the Engineering and Informational Sciences*, vol. 34, no. 2, pp. 221–234, Jan. 2019. [Online]. Available: https://doi.org/10.1017/s0269964818000554

[17] Z. Shi, Z. Wang, P. Zhong, and Y. Fan, "The generalized entropy ergodic theorem for nonhomogeneous bifurcating Markov chains indexed by a binary tree," *Journal of Theoretical Probability*, vol. 35, no. 3, pp. 1367–1390, Sep. 2022. [Online]. Available: https://doi.org/10.1007/s10959-021-01117-1

[18] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Springer Berlin Heidelberg, 2010. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-03311-7

[19] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*. Princeton University Press, Dec. 1981. [Online]. Available: http://dx.doi.org/10.1515/9781400855162

[20] M. D. Boshernitzan, "Quantitative recurrence results," *Inventiones mathematicae*, vol. 113, no. 1, pp. 617–631, Dec. 1993. [Online]. Available: https://doi.org/10.1007/BF01244320

[21] L. Barreira and B. Saussol, "Hausdorff dimension of measures via Poincaré recurrence," *Communications in Mathematical Physics*, vol. 219, no. 2, pp. 443–463, May 2001. [Online]. Available: https://doi.org/10.1007/s002200100427

[22] M. Hussain, B. Li, D. Simmons, and B. Wang, "Dynamical Borel-Cantelli lemma for recurrence theory," *Ergodic Theory and Dynamical Systems*, pp. 1–15, Apr. 2021.

[23] M. Kirsebom, P. Kunde, and T. Persson, "On shrinking targets and self-returning points," *Annali Scuola Normale Superiore - Classe Di Scienze*, pp. 1499–1535, Sep. 2023.

[24] T. Persson, "A strong Borel–Cantelli lemma for recurrence," *Studia Mathematica*, vol. 268, no. 1, pp. 75–89, 2023. [Online]. Available: http://dx.doi.org/10.4064/sm220216-2-7

[25] D. Kleinbock and J. Zheng, "Dynamical Borel–Cantelli lemma for recurrence under Lipschitz twists," *Nonlinearity*, vol. 36, no. 2, pp. 1434–1460, Jan. 2023. [Online]. Available: http://dx.doi.org/10.1088/1361-6544/acafcb

[26] Y. He and L. Liao, "Quantitative recurrence properties for piecewise expanding map on $[0,1]^d$," *arXiv*, Feb. 2023, preprint, arXiv2302.05149. [Online]. Available: http://arxiv.org/abs/2302.05149

[27] D. Allen, S. Baker, and B. Bárány, "Recurrence rates for shifts of finite type," *Advances in Mathematics*, vol. 460, p. 110039, Jan. 2025. [Online]. Available: http://dx.doi.org/10.1016/j.aim.2024.110039

[28] B. Tan and B.-W. Wang, "Quantitative recurrence properties for beta-dynamical system," *Advances in Mathematics*, vol. 228, no. 4, pp. 2071–2097, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001870811002258

[29] S. Seuret and B.-W. Wang, "Quantitative recurrence properties in conformal iterated function systems," *Advances in Mathematics*, vol. 280, pp. 472–505, 2015.

[30] Y. Chang, M. Wu, and W. Wu, "Quantitative recurrence properties and homogeneous self-similar sets," *Proceedings of the American Mathematical Society*, vol. 147, no. 4, pp. 1453–1465, Dec. 2018. [Online]. Available: https://www.ams.org/proc/2019-147-04/S0002-9939-2018-14287-2/

[31] S. Baker and M. Farmer, "Quantitative recurrence properties for self-conformal sets," *Proceedings of the American Mathematical Society*, vol. 149, no. 3, pp. 1127–1138, Mar. 2021. [Online]. Available: https://doi.org/10.1090/proc/15285

[32] Z.-N. Hu and T. Persson, "Hausdorff dimension of recurrence sets," *Nonlinearity*, vol. 37, no. 5, p. 055010, Apr. 2024. [Online]. Available: http://dx.doi.org/10.1088/1361-6544/ad3597

[33] B. Bárány and S. Troscheit, "Dynamically defined subsets of generic self-affine sets," *Nonlinearity*, vol. 35, no. 10, pp. 4986–5013, Aug. 2022. [Online]. Available: https://doi.org/10.1088/1361-6544/ac8219

[34] H. Weyl, "Über die gleichverteilung von zahlen mod. eins," *Mathematische Annalen*, vol. 77, no. 3, pp. 313–352, Sep. 1916. [Online]. Available: http://dx.doi.org/10.1007/BF01475864

[35] H. Davenport, P. Erdős, and W. J. LeVeque, "On Weyl's criterion for uniform distribution," *Michigan Mathematical Journal*, vol. 10, no. 3, Aug. 1963. [Online]. Available: http://dx.doi.org/10.1307/mmj/1028998917

[36] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, ser. Pure & Applied Mathematics Monograph. Nashville, TN: John Wiley & Sons, Jun. 1974.

[37] E. M. Stein and T. S. Murphy, *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals*. Princeton University Press, 1993. [Online]. Available: http://www.jstor.org/stable/j.ctt1bpmb3s

[38] Z. Rudnick and P. Sarnak, "The pair correlation function of fractional parts of polynomials," *Communications in Mathematical Physics*, vol. 194, no. 1, pp. 61–70, May 1998. [Online]. Available: http://dx.doi.org/10.1007/s002200050348

[39] C. Aistleitner, T. Lachmann, and F. Pausinger, "Pair correlations and equidistribution," *Journal of Number Theory*, vol. 182, pp. 206–220, Jan. 2018. [Online]. Available: http://dx.doi.org/10.1016/j.jnt.2017.06.009

[40] S. Grepstad and G. Larcher, "On pair correlation and discrepancy," *Archiv der Mathematik*, vol. 109, no. 2, pp. 143–149, Jun. 2017. [Online]. Available: http://dx.doi.org/10.1007/s00013-017-1060-1

[41] J. Marklof, "Pair correlation and equidistribution on manifolds," *Monatshefte für Mathematik*, vol. 191, no. 2, pp. 279–294, Jun. 2019. [Online]. Available: http://dx.doi.org/10.1007/s00605-019-01308-3

[42] A. Algom, F. R. Hertz, and Z. Wang, "Polynomial Fourier decay and a cocycle version of Dolgopyat's method for self conformal measures," *arXiv preprint arXiv:2306.01275*, 2023.

[43] H. Furstenberg, *Intersections of Cantor Sets and Transversality of Semigroups*. Princeton: Princeton University Press, 1971, pp. 41–60. [Online]. Available: https://doi.org/10.1515/9781400869312-004

[44] M. Gavish, "Measures with uniform scaling scenery," *Ergodic Theory and Dynamical Systems*, vol. 31, no. 1, pp. 33–48, 2011.

[45] M. Hochman and P. Shmerkin, "Local entropy averages and projections of fractal measures," *Annals of Mathematics*, vol. 175, no. 3, pp. 1001–1059, May 2012. [Online]. Available: http://dx.doi.org/10.4007/annals.2012.175.3.1

[46] ——, "Equidistribution from fractal measures," *Inventiones mathematicae*, vol. 202, no. 1, pp. 427–479, Feb. 2015. [Online]. Available: http://dx.doi.org/10.1007/s00222-014-0573-5

[47] A. Ferguson, J. M. Fraser, and T. Sahlsten, "Scaling scenery of $(\times m, \times n)$ invariant measures," *Advances in Mathematics*, vol. 268, pp. 564–602, Jan. 2015. [Online]. Available: http://dx.doi.org/10.1016/j.aim.2014.09.019

[48] A. Pyörälä, "The scenery flow of self-similar measures with weak separation condition," *Ergodic Theory and Dynamical Systems*, pp. 1–24, Aug. 2021. [Online]. Available: http://dx.doi.org/10.1017/etds.2021.86

[49] B. Bárány, A. Käenmäki, A. Pyörälä, and M. Wu, "Scaling limits of self-conformal measures," 2023. [Online]. Available: https://arxiv.org/abs/2308.11399

[50] T. Kempton, "The scenery flow for self-affine measures," 2015. [Online]. Available: https://arxiv.org/abs/1505.01663

[51] J. M. Fraser and M. Pollicott, "Micromeasure distributions and applications for conformally generated fractals," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 159, no. 3, pp. 547–566, Oct. 2015. [Online]. Available: http://dx.doi.org/10.1017/S0305004115000523

[52] T. Bedford, "Crinkly curves, Markov partitions and box dimensions in self-similar sets." Ph.D. dissertation, Thesis (Ph. D.)–The University of Warwick, 1984.

[53] C. McMullen, “The Hausdorff dimension of general Sierpiński carpets,” *Nagoya Mathematical Journal*, vol. 96, pp. 1–9, Dec. 1984.

[54] B. Bárány and M. Rams, “Shrinking targets on Bedford–McMullen carpets,” *Proceedings of the London Mathematical Society*, vol. 117, no. 5, p. 951–995, 2018.

[55] M. J. Bertin, A. Decomps-Guilloux, M. Grandet-Hugot, M. Pathiaux-Delefosse, and J. P. Schreiber, *Pisot and Salem Numbers*. Birkhäuser Basel, 1992. [Online]. Available: http://dx.doi.org/10.1007/978-3-0348-8632-1

[56] C. Aistleitner, S. Baker, N. Technau, and N. Yesha, “Gap statistics and higher correlations for geometric progressions modulo one,” *Mathematische Annalen*, vol. 385, no. 1–2, p. 845–861, Jan. 2022. [Online]. Available: http://dx.doi.org/10.1007/s00208-022-02362-3

[57] S. Baker and A. Banaji, “Polynomial Fourier decay for fractal measures and their pushforwards,” *Mathematische Annalen*, Jan. 2025. [Online]. Available: http://dx.doi.org/10.1007/s00208-025-03091-z

[58] O. Khalil, “Exponential mixing via additive combinatorics,” 2023. [Online]. Available: https://arxiv.org/abs/2305.00527

Original publications

I Wu, Y.-L., & Yuan, N. (2024). Quantitative recurrence problem on some Bedford–McMullen carpets. Journal of Mathematical Analysis and Applications, 543(2), 128938. https://doi.org/10.1016/j.jmaa.2024.128938

II Ban, J.-C., & Wu, Y.-L. (2025). On the topological pressure of axial product on trees. Stochastics and Dynamics. https://doi.org/10.1142/s0219493725500078

III Ban, J.-C., Lai, G.-Y., & Wu, Y.-L. (2025). Hausdorff dimensions of irreducible Markov hom tree-shifts. Journal of the London Mathematical Society, 111(6). https://doi.org/10.1112/jlms.70198

IV Algom, A., Chang, Y., Wu, M., & Wu, Y.-L. (2025). Van der Corput and metric theorems for geometric progressions for self-similar measures. Mathematische Annalen. https://doi.org/10.1007/s00208-025-03233-3

V Wu, M., & Wu, Y.-L. (2025). Scaling properties of invariant measures for conformal iterated function systems.

Original publications are not included in the electronic version of the dissertation.

I

J. Math. Anal. Appl. 543 (2025) 128938

Contents lists available at ScienceDirect

Journal of Mathematical Analysis and Applications

journal homepage: www.elsevier.com/locate/jmaa

Regular Articles

Quantitative recurrence problem on some Bedford-McMullen carpets

Yu-Liang Wu [a], Na Yuan [b,*]

[a] *Department of Mathematical Sciences, University of Oulu, Pentti Kaiteran katu 1, Oulu, 90014, Finland*
[b] *School of Mathematics, Guangdong University of Education, Xingang Middle Road 351, Guangzhou, 510303, China*

ARTICLE INFO

Article history:
Received 17 March 2024
Available online 9 October 2024
Submitted by L. Olsen

Keywords:
Bedford–McMullen carpet
Quantitative recurrence
Hausdorff dimension

ABSTRACT

In this paper, we study the Hausdorff dimension of the quantitative recurrent set of the canonical endomorphism on the Bedford–McMullen carpets whose Hausdorff dimension and box dimension are equal.

1. Introduction

1.1. Background

The concept of recurrence plays an important role in dynamical systems and ergodic theory. Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system equipped with a compatible metric d, i.e., (X, d) is a metric space and $\mathcal{B}$ is a Borel σ-algebra of X. If (X, d) is a separable metric space, the Poincaré Recurrence Theorem implies that μ-almost every $x \in X$ is recurrent in the sense that

$$\liminf_{n\to\infty} d(T^n x, x) = 0.$$

In nature, the result provides no information about the rate at which an orbit will return to the initial point or in what manner a neighborhood of the initial point will shrink under the iteration. The interest in this quantitative characterization has provoked a rich subsequent literature on the so-called quantitative recurrent sets: given a *rate function* $\psi : X \times \mathbb{N} \to (0, \infty)$, the *quantitative recurrent set with respect to* ψ is defined as

* Corresponding author.
E-mail addresses: yu-liang.wu@oulu.fi (Y.-L. Wu), yuanna199@gmail.com (N. Yuan).

https://doi.org/10.1016/j.jmaa.2024.128938
0022-247X/

$$R(T,\psi)=\{x\in X: d(T^n x,x)<\psi(n,x) \text{ for infinitely many } n\in\mathbb{N}\}. \tag{1.1}$$

Boshernitzan [8] gave an outstanding result for general systems concerning the size in measure of $R(T,\psi)$. Later, Barreira and Saussol [6] stated a finer result.

In recent years, many authors have turned their eyes to recurrent sets on fractals. On the one hand, some researchers showed that the μ-measure of the set $R(T,\psi)$ is null or full according to convergence or divergence of a certain series in some dynamical systems (see Chang-Wu-Wu [10], Baker-Farmer [2], Hussain-Li-Simmons-Wang [15], Kirsebom-Kunde-Persson [16], Persson [24], Kleinbock-Zheng [17] and Baker-Koivusalo [3]). On the other hand, many researchers studied the Hausdorff dimension of the set $R(T,\psi)$ in some dynamical systems (see Tan-Wang [28] and Seuret-Wang [26]). Note that when we require $\{T^n x\}_{n\geq 1}$ to return to the neighborhoods of a chosen point $x_0\in X$ rather than the initial point x, the problem becomes the so-called shrinking target problem, which was first investigated by Hill and Velani [13]. Since then, many authors have contributed to the study of the shrinking target problem. To name but a few, see [1,4,9,12,14,18,20–22,25,27,29] and references within.

As for the Hausdorff dimension, it is to be noted that the aforementioned works mainly involve systems of $\mathbb{R}^1$ or dynamical systems in conformal dynamics, and hardly anything is known as far as high-dimensional non-conformal dynamics are concerned. The only known result was presented by Bárány and Troscheit [5], who investigated dimensions for both the quantitative recurrence and the shrinking target problems for dynamically defined subsets of generic self-affine sets (in the sense of Lebesgue almost all translations). In this work, we intend to consider the Bedford-McMullen carpets, a typical family of planar self-affine sets introduced in [7] and [23], and discuss the recurrence based on the Euclidean metric. This is the first time the problem of recurrent sets for deterministic self-affine fractal sets is discussed. We would like to point out that due to the similarity of the quantitative recurrence problem and the shrinking target problem in nature, techniques for one problem sometimes apply to the other. For our study, arguments closely follow those from Barany and Rams [4] with a technical nuance mentioned in Remark 3.1.

1.2. The main theorem

It is the purpose of this paper to study the quantitative recurrent set problem on the system (K,T) of Bedford-McMullen carpet K. Let $2\leq m_1\leq m_2$ be two integers and $\Sigma_{m_i}=\{0,1,\cdots,m_i-1\}$ for $i=1,2$. Define for every $\mathtt{a}=(\mathtt{a}^{(1)},\mathtt{a}^{(2)})\in\Sigma_{m_1}\times\Sigma_{m_2}$ a map $\phi_{\mathtt{a}}:[0,1]^2\to[0,1]^2$ as

$$\phi_{\mathtt{a}}(x^{(1)},x^{(2)}):=\left(\frac{x^{(1)}+\mathtt{a}^{(1)}}{m_1},\frac{x^{(2)}+\mathtt{a}^{(2)}}{m_2}\right).$$

Given any nonempty subset $A\subseteq\Sigma_{m_1}\times\Sigma_{m_2}$, the *Bedford-McMullen carpet K associated with A* is the unique attractor of the iterated function system $\{\phi_{\mathtt{a}}:\mathtt{a}\in A\}$. If we consider the coding map $\pi:(\Sigma_{m_1}\times\Sigma_{m_2})^{\mathbb{N}}\to[0,1]^2$ defined as

$$\pi(\mathtt{x})=\left(\sum_{n=1}^{\infty}\frac{\mathtt{x}_n^{(1)}}{m_1^n},\sum_{n=1}^{\infty}\frac{\mathtt{x}_n^{(2)}}{m_2^n}\right),$$

where $\mathtt{x}=(\mathtt{x}_1^{(1)},\mathtt{x}_1^{(2)})(\mathtt{x}_2^{(1)},\mathtt{x}_2^{(2)})(\mathtt{x}_3^{(1)},\mathtt{x}_3^{(2)})\cdots$, then the Bedford-McMullen carpet K can be expressed as $K=\pi(A^{\mathbb{N}})$. It is noteworthy that the latter definition naturally endows the carpet K with a map $T:K\to K$ defined as

$$T(x)=(T_{m_1}(x^{(1)}),T_{m_2}(x^{(2)})):=\left(m_1x^{(1)} \pmod 1, m_2x^{(2)} \pmod 1\right).$$

Let $\psi : \mathbb{N} \to (0, \infty)$ be a *rate function* and $\gamma > 0$, and define the *recurrent set* $W_\gamma(K, T, \psi)$ *with respect to* ψ as:

$$W_\gamma(K,T,\psi) := \left\{ x \in K : \begin{cases} |x^{(1)} - T_{m_1}^n(x^{(1)})| < \psi(n) \\ |x^{(2)} - T_{m_2}^n(x^{(2)})| < \psi(n)^\gamma \end{cases} \text{for infinitely many } n \right\}. \tag{1.2}$$

Note that for the rest of the discussion, we assume no monotonicity of ψ. Denote by $\dim_{\mathrm{H}}$ and $\dim_{\mathrm{B}}$ the Hausdorff and box dimensions, respectively. The following notations are used throughout the discussions.

$$\ell_1(n) = -\log_{m_1} \psi(n) \quad \text{and} \quad \ell_2(n) = -\log_{m_2} \psi(n)^\gamma \tag{1.3}$$

and

$$\frac{\ell_i(n)}{n} = \tau_i(n) \quad \text{and} \quad \liminf_{n\to\infty} \frac{\ell_i(n)}{n} = \tau_i, \quad \text{for } i = 1, 2. \tag{1.4}$$

Denote

$$\hat{\ell}_i(n) = \lceil \ell_i(n) \rceil := \min\{k \in \mathbb{N} : k \geq \ell_i(n)\}. \tag{1.5}$$

for $i = 1, 2$ and $n \in \mathbb{N}$. It is noteworthy that if $0 \leq \gamma \leq \log_{m_1} m_2$, then $\ell_1(n) \geq \ell_2(n)$ and $\hat{\ell}_1(n) \geq \hat{\ell}_2(n)$ for all $n \in \mathbb{N}$.

Our main result related to the Hausdorff dimension of the set $W_\gamma(K, T, \psi)$ is as follows.

Theorem 1.1. *Let* $W_\gamma(K, T, \psi)$ *be defined in* (1.2), $\tau_1 \geq 0$, *and* $0 \leq \gamma \leq \log_{m_1} m_2$. *Let* M *be the number of columns containing at least one chosen rectangle and* N_i *the number of rectangles chosen from the i-th non-empty column. Suppose that* $\dim_{\mathrm{B}} K = \dim_{\mathrm{H}} K$, *or equivalently, there exists an integer* $N \geq 1$ *such that* $N_i \in \{0, N\}$ *for all i (see [7] or [23]), we have* $\dim_{\mathrm{H}} W_\gamma(K, T, \psi) = \min\{t_1, t_2\}$, *where*

$$t_1 = \begin{cases} \left(1 - \frac{\tau_1 \log_{m_2} m_1}{1+\tau_2}\right) \cdot \log_{m_1} M + \frac{\log_{m_2} N}{1+\tau_2} & \text{if } 1 + (1-\gamma)\tau_1 \leq \log_{m_1} m_2, \\ \frac{\log_{m_2} M + \log_{m_2} N}{1+\tau_2} & \text{if } 1 + (1-\gamma)\tau_1 > \log_{m_1} m_2. \end{cases}$$

$$t_2 = \begin{cases} \frac{\log_{m_1} M}{1+\tau_1} + \log_{m_2} N & \text{if } 1 + \tau_1 \leq \log_{m_1} m_2, \\ \frac{\log_{m_1} M + \log_{m_1} N}{1+\tau_1} & \text{if } 1 + (1-\gamma)\tau_1 \leq \log_{m_1} m_2 < 1 + \tau_1, \\ \frac{\log_{m_1} M + (1+(1-\gamma)\tau_1) \log_{m_2} N}{1+\tau_1} & \text{if } 1 + (1-\gamma)\tau_1 > \log_{m_1} m_2. \end{cases}$$

In particular, if $\gamma = 1$, *this is reduced to the following form.*

(1) *If* $\log_{m_1} m_2 > 1 + \tau_1$, *then*

$$\dim_{\mathrm{H}} W_\gamma(K,T,\psi) = \min\left\{ \frac{\log_{m_1} M + \log_{m_2} N}{1+\tau_2}, \frac{\log_{m_1} M}{1+\tau_1} + \log_{m_2} N \right\}.$$

(2) *If* $\log_{m_1} m_2 \leq 1 + \tau_1$, *then*

$$\dim_{\mathrm{H}} W_\gamma(K,T,\psi) = \min\left\{ \frac{\log_{m_1} M + \log_{m_2} N}{1+\tau_2}, \frac{\log_{m_1} M + \log_{m_1} N}{1+\tau_1} \right\},$$

Remark 1.1. We note that Theorem 1.1 covers the cases of all nontrivial τ_1 for if $\tau_1 < 0$, then $W_\gamma(K, T, \psi) = K$. In fact, for the rest of the discussion, we further assume $0 < \tau_1 < \infty$ since the cases $\tau_1 = 0$ and $\tau_1 = \infty$ could be inferred from the remaining cases by noting that $W_\gamma(K, T, \psi) \subset W_\gamma(K, T, \phi)$ if $\psi \leq \phi$.

Remark 1.2. The set $W_1(K,T,\psi)$ can be interpreted as a quantitative recurrent set defined in (1.1) with the maximum norm of $\mathbb{R}^2$. Nevertheless, Theorem 1.1 implies that the dimension of the quantitative recurrent set is invariant under all equivalent metrics, including the one induced by the Euclidean norm.

Remark 1.3. The homogeneity property (i.e., $N_i \in \{0, N\}$ for all i) in Theorem 1.1 is imposed only for the lower bound (derived in Section 3) and upper bound (derived in Section 4) to coincide. Nevertheless, our arguments for the lower and upper bounds remain valid (with necessary modifications) without this assumption. As a related issue, in Theorem 1.1, our study of $W_\gamma(K,T,\psi)$ is limited to a restrictive class of γ so that $\tau_1 \geq \tau_2$ and that the homogeneity property could be exploited to deduce matching lower bound and upper bound. We note that this is also the main difficulty we face if we were to further generalize the theorem by having independent rate functions in the two coordinates.

1.3. Organization of the paper

In this paper, we will present a Hausdorff dimension formula of the recurrent set valid for the subfamily of Bedford–McMullen carpets whose Hausdorff dimensions are equal to the box dimensions. The paper is organized as follows. In Section 2, we present necessary notation and preliminaries. Section 3 and Section 4 are devoted to the proof of Theorem 1.1. Finally, in Section 5, we provide some examples to which the main theorem is applicable.

2. Preliminaries

In this section, we introduce necessary notations and preliminaries. For convenience's sake, the following notations of the symbolic space are also introduced. For $i = 1, 2$, let $\Sigma_{m_i}^n = \{\mathtt{u} : \mathtt{u} = (\mathtt{u}_1, \ldots, \mathtt{u}_n), \mathtt{u}_j \in \Sigma_{m_i},\ j = 1, \ldots, n\}$. Denote by $|u|$ the length of $u \in \Sigma_{m_i}^n$ for $i = 1, 2$. Firstly, for any $l \in \mathbb{N} \cup \{\infty\}$ and $u \in \Sigma_{m_i}^n$, write $(u)^l$ for the word $(u, \cdots, u)$ (l times repeated concatenation of the word). More generally, for any positive number $l > 0$, denote by $(u)^l$ the word $(u)^{\lfloor l \rfloor} u'$, where u' is the prefix of u with length $\lfloor (l - \lfloor l \rfloor)|u| \rfloor$ and $\lfloor l \rfloor = \max\{k \in \mathbb{N} : k \leq l\}$. Secondly, we identify the spaces $(\Sigma_{m_1} \times \Sigma_{m_2})^{\mathbb{N}}$ and $(\Sigma_{m_1}^{\mathbb{N}} \times \Sigma_{m_2}^{\mathbb{N}})$ by setting

$$(\mathtt{x}_1^{(1)}, \mathtt{x}_1^{(2)})(\mathtt{x}_2^{(1)}, \mathtt{x}_2^{(2)})(\mathtt{x}_3^{(1)}, \mathtt{x}_3^{(2)}) \cdots \sim (\mathtt{x}_1^{(1)}\mathtt{x}_2^{(1)}\mathtt{x}_3^{(1)} \cdots, \mathtt{x}_1^{(2)}\mathtt{x}_2^{(2)}\mathtt{x}_3^{(2)} \cdots),$$

which is a one-to-one correspondence. With this identification, we introduce the following notation of cylinder set with different lengths in its coordinates:

$$[(\mathtt{w}^{(1)}, \mathtt{w}^{(2)})] = \{(\mathtt{x}^{(1)}, \mathtt{x}^{(2)}) : \mathtt{x}_{1:n_i}^{(i)} = \mathtt{w}^{(i)}, i = 1, 2\}, \quad (\mathtt{w}^{(1)}, \mathtt{w}^{(2)}) \in \Sigma_{m_1}^{n_1} \times \Sigma_{m_2}^{n_2}\},$$

each of which is associated with a half-open half-closed rectangle

$$I([(\mathtt{w}^{(1)}, \mathtt{w}^{(2)})]) = \left[\sum_{i=1}^{n_1} \frac{\mathtt{w}^{(1)}}{m_1^i}, \sum_{i=1}^{n_1} \frac{\mathtt{w}^{(1)}}{m_1^i} + m_1^{-n_1}\right) \times \left[\sum_{i=1}^{n_2} \frac{\mathtt{w}^{(2)}}{m_2^i}, \sum_{i=1}^{n_2} \frac{\mathtt{w}^{(2)}}{m_2^i} + m_2^{-n_2}\right).$$

We call $I([(\mathtt{w}^{(1)}, \mathtt{w}^{(2)})])$ an n_1*-th level rectangle* when $n_1 = n_2$. In particular, we call $I([(\mathtt{w}^{(1)}, \mathtt{w}^{(2)})])$ an n_1*-th level approximate square* when $n_2 = \lceil \log_{m_2} m_1 \cdot n_1 \rceil$. For any $\mathtt{x} \in (\Sigma_{m_1} \times \Sigma_{m_2})^{\mathbb{N}}$, let $\mathtt{x}_{1:n}^{(i)} := \mathtt{x}_1^{(i)}\mathtt{x}_2^{(i)} \ldots \mathtt{x}_n^{(i)}$. If $A \subset \Sigma_{m_1} \times \Sigma_{m_2}$, let

$$A^n = \{(\mathtt{x}_1^{(1)}\mathtt{x}_2^{(1)}\mathtt{x}_3^{(1)} \cdots \mathtt{x}_n^{(1)}, \mathtt{x}_1^{(2)}\mathtt{x}_2^{(2)}\mathtt{x}_3^{(2)} \cdots \mathtt{x}_n^{(2)}) : (\mathtt{x}_i^{(1)}, \mathtt{x}_i^{(2)}) \in A\}.$$

For any $\mathtt{u} = (\mathtt{u}_1^{(1)}\mathtt{u}_2^{(1)} \ldots \mathtt{u}_n^{(1)}, \mathtt{u}_1^{(2)}\mathtt{u}_2^{(2)} \ldots \mathtt{u}_n^{(2)}) \in A^n$, let $\mathtt{u}^{(i)} = (\mathtt{u}_1^{(i)}\mathtt{u}_2^{(i)} \ldots \mathtt{u}_n^{(i)})$, $i = 1, 2$.

3. Proof of the lower bound

In this section, we will prove the lower bound of the Hausdorff dimension of $W_\gamma(K,T,\psi)$, and to this end, we are going to prove Lemma 3.1-3.3.

In the following, we outline our strategy of proof. To begin with, we note that it suffices to prove the lower bound of Theorem 1.1 with an additional assumption that $0 \le \hat{\ell}_2(n) \le \hat{\ell}_1(n)$ for all n. Indeed, if Theorem 1.1 holds under the assumption, then for any $\psi(n)$ one may choose $\phi(n) = \min\{1, \psi(n)\}$ so that $\dim_{\mathrm{H}} W_\gamma(K,T,\psi) \ge \dim_{\mathrm{H}} W_\gamma(K,T,\phi)$ yields the desired lower bound. Let $p \in [0,1]^A$ be a probability vector indexed by A. For every $\mathtt{a} \in A$, let

$$p_{\mathtt{a}^{(1)}} = \sum_{\mathtt{b}\in A,\ \mathtt{b}^{(1)}=\mathtt{a}^{(1)}} p_{\mathtt{b}} \quad \text{and} \quad A^{(1)} = \{\mathtt{a}^{(1)} : \mathtt{a} = (\mathtt{a}^{(1)}, \mathtt{a}^{(2)}),\ \mathtt{a}\in A\}.$$

We take an increasing sequence of natural numbers n_i such that

$$2^i \sum_{j=1}^{i} n_j \ll n_{i+1} \qquad \text{and} \qquad \lim_{i\to\infty} \tau_1(n_j) = \tau_1, \tag{3.1}$$

and define the measure μ (with respect to probability measure p) as follows. To begin with, write $\mathcal{W}_1 = A^{n_1}$ and set

$$\mathcal{W}_{i+1} = \Big\{\mathtt{wuvw}' \in A^{n_{i+1}} : \mathtt{w}\in\mathcal{W}_i, \mathtt{u}\in A^{\hat{\ell}_2(n_i)}, \mathtt{v}\in A^{\hat{\ell}_1(n_i)-\hat{\ell}_2(n_i)}, \mathtt{w}'\in A^{n_{i+1}-(n_i+\hat{\ell}_1(n_i))}, \\ (\mathtt{uv})^{(1)} = (\mathtt{w}^{(1)})^{\frac{\hat{\ell}_1(n_i)}{n_i}}, \mathtt{u}^{(2)} = (\mathtt{w}^{(2)})^{\frac{\hat{\ell}_2(n_i)}{n_i}}\Big\}$$

Observe that by construction, a word $\mathtt{w}$ is in $\mathcal{W}_i$ if and only if it can be written in the form of

$$\mathtt{w} = \underline{\mathtt{w}_1}\,\underline{\mathtt{u}_1}\,\underline{\mathtt{v}_1}\,\underline{\mathtt{w}_2}\cdots\underline{\mathtt{w}_{i-1}}\,\underline{\mathtt{u}_{i-1}}\,\underline{\mathtt{v}_{i-1}}\,\underline{\mathtt{w}_i}, \tag{3.2}$$

where, by setting $n_0 = 0$, $\underline{\mathtt{w}_j} \in A^{n_j-(n_{j-1}+\hat{\ell}_1(n_{j-1}))}$ is arbitrary, and $\underline{\mathtt{u}_j} \in A^{\hat{\ell}_2(n_j)}$, $\underline{\mathtt{v}_j} \in A^{\hat{\ell}_1(n_{j-1})-\hat{\ell}_2(n_{j-1})}$ satisfy

$$\begin{cases} \underline{\mathtt{u}_j}^{(1)}\,\underline{\mathtt{v}_j}^{(1)} = \left(\underline{\mathtt{w}_1}^{(1)}\cdots\underline{\mathtt{u}_{j-1}}^{(1)}\,\underline{\mathtt{v}_{j-1}}^{(1)}\,\underline{\mathtt{w}_j}^{(1)}\right)^{\frac{\hat{\ell}_1(n_j)}{n_j}}, \\ \underline{\mathtt{u}_j}^{(2)} = \left(\underline{\mathtt{w}_1}^{(2)}\cdots\underline{\mathtt{u}_{j-1}}^{(2)}\,\underline{\mathtt{v}_{j-1}}^{(2)}\,\underline{\mathtt{w}_j}^{(2)}\right)^{\frac{\hat{\ell}_2(n_j)}{n_j}}, \end{cases} \tag{3.3}$$

for all $1 \le j \le i$. For convenience, denote

$$L_i = |\underline{\mathtt{w}_1}\cdots\underline{\mathtt{u}_i}| = n_i + \hat{\ell}_2(n_i),$$
$$R_i = |\underline{\mathtt{w}_1}\cdots\underline{\mathtt{v}_i}| = n_i + \hat{\ell}_1(n_i).$$

Next, we inductively define μ on cylinder sets of length $n_i + \hat{\ell}_1(n_i)$ for $i \in \mathbb{N}$ as

$$\mu[\mathtt{w}] = \begin{cases} \mu[\mathtt{w}'] \displaystyle\prod_{\ell=1}^{|\mathtt{v}|} \frac{p_{\mathtt{v}_\ell}}{p_{\mathtt{v}_\ell^{(1)}}} \prod_{\ell=1}^{|\mathtt{w}''|} p_{\mathtt{w}''_\ell}; & \text{if } \mathtt{w}'\in\mathcal{W}_{i-1} \text{ and } \mathtt{w} = \mathtt{w}'\mathtt{uvw}''\in\mathcal{W}_i \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

It is readily checked by definition that

$$1 = \sum_{\mathtt{w}' \in A^{n_1+\hat{\ell}_1(n_1)}} \mu[\mathtt{w}'] \quad \text{and} \quad \mu[\mathtt{w}] = \sum_{\mathtt{w}' \in A^{n_{i+1}+\hat{\ell}_1(n_{i+1})}} \mu[\mathtt{w}'] \text{ for every } \mathtt{w} \in A^{n_i+\hat{\ell}_1(n_i)},$$

which extends μ to an additive set function on the algebra generated by all cylinders sets and hence, by Carathéodory extension theorem, to a Borel probability measure on $A^{\mathbb{N}}$. It then follows from this definition that $\operatorname{supp}(\mu \circ \pi^{-1}) \subseteq W_\gamma(K, T, \psi)$. We note that our choice of measure is similar to the piecewise Bernoulli measure considered in [4]. For convenience, we write

$$F(\mathtt{x}, k) = \mu\big[(\mathtt{x}^{(1)}_{1:k}, \mathtt{x}^{(2)}_{1:\lceil k \log_{m_2} m_1 \rceil})\big], \tag{3.5}$$

and for the rest of this section abuse the notation $\mathtt{x}_i$ (similarly for $\mathtt{x}^{(1)}_i$ and $\mathtt{x}^{(2)}_i$) to denote the random variable that indicates the i-th letter of the elements in the sample space $(\Sigma_{m_1} \times \Sigma_{m_2})^{\mathbb{N}}$, which consequently renders $F(\mathtt{x}, k)$ a random variable in this context. Under the circumstances, we recover in Lemma 3.2 an analog of [4, Lemma 4.2]:

$$\liminf_{k\to\infty} \frac{-\log_{m_1} F(\mathtt{x}, k)}{k} = \liminf_{k\to\infty} \mathbb{E}_p \left[\frac{-\log_{m_1} F(\mathtt{x}, k)}{k}\right] \quad \mu\text{-a.e.},$$

where $\mathbb{E}_p$ denotes the mathematical expectation with respect to μ. It is noteworthy that the above limit coincides with the lower local dimension function almost everywhere (see for example [19]), i.e.,

$$\liminf_{k\to\infty} \frac{-\log_{m_1} F(\mathtt{x}, k)}{k} = \liminf_{r\to 0} \frac{\log \mu \circ \pi^{-1}(B(\pi(\mathtt{x}), r))}{\log r} \quad \mu\text{-a.e.}$$

We then apply [11, Proposition 2.3] to conclude that

$$\dim_{\mathrm{H}}(W_\gamma(K, T, \psi)) \geq \sup_p \left[\liminf_{k\to\infty} \mathbb{E}_p \left[\frac{-\log_{m_1} F(\mathtt{x}, k)}{k}\right]\right],$$

for which the lower limit on the right-hand side is calculated in Lemma 3.2 with the optimal value obtained thereafter. For convenience, we introduce the following notations.

$$H(p) = \sum_{\mathtt{a}\in A} -p_{\mathtt{a}} \log_{m_1} p_{\mathtt{a}}, \tag{3.6}$$

$$H_1(p) = \sum_{\mathtt{a}^{(1)}\in A^{(1)}} -p_{\mathtt{a}^{(1)}} \log_{m_1} p_{\mathtt{a}^{(1)}}, \tag{3.7}$$

and

$$H_2(p) = \sum_{\mathtt{a}\in A} -p_{\mathtt{a}} \log_{m_1} \frac{p_{\mathtt{a}}}{p_{\mathtt{a}^{(1)}}}. \tag{3.8}$$

Now, we begin to prove the lemmas.

Lemma 3.1. *Let $\{X_i : i \in \mathbb{N}\}$ be independent random variables with zero means and $M := \sup_i \mathbb{E}_p|X_i|^4 < \infty$. Then, for all $n, m \in \mathbb{N}$ and $\epsilon > 0$, $\mu(|m^{-1}\sum_{i=1}^n X_i| > \epsilon) \leq 3(\frac{m^4}{n^2})^{-1} M \epsilon^{-4}$.*

Proof. It is an immediate consequence of Markov's inequality. □

Lemma 3.2. *Let $F(\mathtt{x}, k)$ be as defined in (3.5) and suppose $0 \leq \tau_1 < \infty$. Then,*

$$\liminf_{k\to\infty} \frac{-\log_{m_1} F(\mathtt{x}, k)}{k} = \liminf_{k\to\infty} \mathbb{E}_p \left[\frac{-\log_{m_1} F(\mathtt{x}, k)}{k}\right] \quad \mu\text{-a.e.} \tag{3.9}$$

Proof. Recall that according to equation (3.2), one can express $\mathtt{x}$ in the following form:

$$\mathtt{x} = \underline{\mathtt{w}_1}\,\underline{\mathtt{u}_1}\,\underline{\mathtt{v}_1}\,\underline{\mathtt{w}_2}\,\underline{\mathtt{u}_2}\,\underline{\mathtt{v}_2}\,\underline{\mathtt{w}_3}\cdots.$$

We first prove the claim that for any $\epsilon > 0$,

$$\mu\left(\left|\frac{\log_{m_1} F(\mathtt{x},k)}{k} - \mathbb{E}_p\left(\frac{\log_{m_1} F(\mathtt{x},k)}{k}\right)\right| > \epsilon\right) = O(k^{-2}),$$

where the $O(k^{-2})$ may depend on p and ϵ but not on k, and thus ϵ throughout the discussion is treated as a constant. Due to the assumption $n_i \gg R_{i-1}$, every $k \in \mathbb{N}$ must fall in one of the following cases.

Case 1: $n_i < k$ **and** $\lfloor k \log_{m_2} m_1 \rfloor \le L_i$**.** For such k, we have

$$\begin{aligned}\frac{\log_{m_1} F(\mathtt{x},k)}{k} =& \frac{1}{k}\Bigg[\log_{m_1}\mu[\underline{\mathtt{w}_1}\cdots\underline{\mathtt{v}_{i-1}}] + \sum_{\ell=1}^{|\underline{\mathtt{w}_i}|}\log_{m_1} p_{(\underline{\mathtt{w}_i})_\ell^{(1)}} + \sum_{\ell=1}^{\min\{|\underline{\mathtt{w}_i}|,\lfloor k\log_{m_2} m_1\rfloor - R_{i-1}\}}\log_{m_1}\frac{p_{(\underline{\mathtt{w}_i})_\ell}}{p_{(\underline{\mathtt{w}_i})_\ell^{(1)}}} \\ &+ \sum_{\ell=1}^{\max\{k-R_i,0\}}\log_{m_1} p_{(\underline{\mathtt{w}_{i+1}})_\ell^{(1)}}\Bigg].\end{aligned} \tag{3.10}$$

Observe that, by virtue of $n_i \gg R_{i-1}$,

$$\left|\frac{1}{k}\log_{m_1}\mu[\underline{\mathtt{w}_1}\cdots\underline{\mathtt{v}_{i-1}}]\right| \le \frac{R_{i-1}}{k}\cdot\max_{\mathtt{a}\in A:p_\mathtt{a}\neq 0}\left|\log_{m_1} p_\mathtt{a}\right| = o(1). \tag{3.11}$$

In addition, according to (3.4) the remaining terms are, respectively, summations over i.i.d. random variables

$$\left\{\log_{m_1} p_{(\underline{\mathtt{w}_i})_\ell^{(1)}} : 1 \le \ell \le |\underline{\mathtt{w}_i}|\right\} \text{ and } \left\{\log_{m_1}\frac{p_{(\underline{\mathtt{w}_i})_\ell}}{p_{(\underline{\mathtt{w}_i})_\ell^{(1)}}} : 1 \le \ell \le |\underline{\mathtt{w}_i}|\right\}$$

with finite fourth moments

$$\mathbb{E}_p\left|\log_{m_1} p_{(\underline{\mathtt{w}_i})_\ell^{(1)}}\right|^4 \le \max_{\mathtt{a}^{(1)}\in A^{(1)}:p_{\mathtt{a}^{(1)}}\neq 0}\left|\log_{m_1} p_\mathtt{a}\right|^4 \quad\text{and}\quad \mathbb{E}_p\left|\log_{m_1}\frac{p_{(\underline{\mathtt{w}_i})_\ell}}{p_{(\underline{\mathtt{w}_i})_\ell^{(1)}}}\right|^4 \le \max_{\mathtt{a}\in A:p_\mathtt{a}\neq 0}\left|\log\frac{p_\mathtt{a}}{p_{\mathtt{a}^{(1)}}}\right|^4.$$

Therefore, Lemma 3.1, together with (3.11), yields

$$\mu\left(\left|\frac{\log_{m_1} F(\mathtt{x},k)}{k} - \mathbb{E}_p\left(\frac{\log_{m_1} F(\mathtt{x},k)}{k}\right)\right| > \epsilon\right) = O(k^{-2})$$

Case 2: $k \le n_{i+1}$ **and** $L_i < \lfloor k\log_{m_2} m_1\rfloor$**.** We begin with a similar expression as above:

$$\begin{aligned}\frac{\log_{m_1} F(\mathtt{x},k)}{k} =& \frac{1}{k}\Bigg[\log_{m_1}\mu[\underline{\mathtt{w}_1}\cdots\underline{\mathtt{v}_{i-1}}] + \sum_{\ell=1}^{|\underline{\mathtt{w}_i}|}\log_{m_1} p_{(\underline{\mathtt{w}_i})_\ell} + \sum_{\ell=1}^{\min\{|\underline{\mathtt{v}_i}|,\lfloor k\log_{m_2} m_1\rfloor - L_i\}}\log_{m_1}\frac{p_{(\underline{\mathtt{v}_i})_\ell}}{p_{(\underline{\mathtt{v}_i})_\ell^{(1)}}} \\ &+ \sum_{\ell=1}^{\max\{k-R_i,0\}}\log_{m_1} p_{(\underline{\mathtt{w}_{i+1}})_\ell^{(1)}} + \sum_{\ell=1}^{\max\{k\log_{m_2} m_1 - R_i,0\}}\frac{p_{(\underline{\mathtt{w}_{i+1}})_\ell}}{p_{(\underline{\mathtt{w}_{i+1}})_\ell^{(1)}}}\Bigg].\end{aligned} \tag{3.12}$$

The argument for i.i.d. random variables proceeds as above except for the summation involving $\underline{\mathtt{v}_i}$, which are not necessary i.i.d. However, this issue could be resolved by considering the conditional measure $\mu_{\mathtt{w}^*}$ of μ on the cylinder set $[\mathtt{w}^*]$, $\mathtt{w}^* = \underline{\mathtt{w}_1}\,\underline{\mathtt{u}_1}\cdots\underline{\mathtt{w}_i} \in \mathcal{W}_i$. Hence, with respect to any well-defined $\mu_{\mathtt{w}^*}$,

$$\sum_{\ell=1}^{n} \log_{m_1} \frac{p_{(\underline{\mathtt{v}_i})_\ell}}{p_{(\underline{\mathtt{v}_i})_\ell^{(1)}}}$$

is a summation over independent random variables with uniformly bounded fourth moments

$$\mathbb{E}_p \left| \log_{m_1} \frac{p_{(\underline{\mathtt{v}_i})_\ell}}{p_{(\underline{\mathtt{v}_i})_\ell^{(1)}}} \right|^4 \leq \max_{\mathtt{a} \in A: p_{\mathtt{a}} \neq 0} \left| \log_{m_1} \frac{p_{\mathtt{a}}}{p_{\mathtt{a}^{(1)}}} \right|^4.$$

Applying Lemma 3.1 again yields that for every $\mathtt{w}^* \in \mathcal{W}_i$,

$$\mu_{\mathtt{w}^*} \left(\left| \frac{1}{k} \sum_{\ell=1}^{n} \log_{m_1} \frac{p_{(\underline{\mathtt{v}_i})_\ell}}{p_{(\underline{\mathtt{v}_i})_\ell^{(1)}}} - \mathbb{E}_p \left(\frac{1}{k} \sum_{\ell=1}^{n} \log_{m_1} \frac{p_{(\underline{\mathtt{v}_i})_\ell}}{p_{(\underline{\mathtt{v}_i})_\ell^{(1)}}} \middle| \mathtt{w}^* \right) \right| > \epsilon \right) = O(k^{-2}).$$

We note that the bound $O(k^{-2})$ could be chosen uniform over $\mathtt{w}^*$, and thus it remains to show that

$$\mu \left(\left| \mathbb{E}_p \left(\frac{1}{k} \sum_{\ell=1}^{n} \log_{m_1} \frac{p_{(\underline{\mathtt{v}_i})_\ell}}{p_{(\underline{\mathtt{v}_i})_\ell^{(1)}}} \right) - \mathbb{E}_p \left(\frac{1}{k} \sum_{\ell=1}^{n} \log_{m_1} \frac{p_{(\underline{\mathtt{v}_i})_\ell}}{p_{(\underline{\mathtt{v}_i})_\ell^{(1)}}} \middle| \mathtt{w}^* \right) \right| > \epsilon \right) = O(k^{-2}).$$

Recall that equation (3.4) asserts μ-almost surely,

$$(\underline{\mathtt{u}_i}\,\underline{\mathtt{v}_i})^{(1)} = \left(\underline{\mathtt{w}_1}^{(1)} \cdots \underline{\mathtt{u}_{i-1}}^{(1)}\, \underline{\mathtt{v}_{i-1}}^{(1)}\, \underline{\mathtt{w}_i}^{(1)} \right)^{\frac{\ell_1(n_i)}{n_i}},$$

which implies that if we associate each index $\ell \in [1, |\underline{\mathtt{v}_i}|]$ of $\underline{\mathtt{v}_i}$ with $\ell' \in [1, n_i]$ satisfying $\ell' = L_i + \ell \pmod{n_i}$, then

$$(\underline{\mathtt{v}_i})_\ell = (\underline{\mathtt{w}_i})_{\ell' - R_{i-1}} \text{ whenever } R_{i-1} < \ell' \leq n_i.$$

Due to $n_i \gg R_{i-1}$, we again deduce that for every $1 \leq n \leq |\underline{\mathtt{v}_i}|$,

$$\frac{1}{k} \sum_{\ell=1}^{n} \log_{m_1} \frac{p_{(\underline{\mathtt{v}_i})_\ell}}{p_{(\underline{\mathtt{v}_i})_\ell^{(1)}}} = \frac{1}{k} \sum_{\substack{\ell \in [1, |\underline{\mathtt{v}_i}|]: \\ \ell' \in (R_{i-1}, n_i]}} \log_{m_1} \frac{p_{(\underline{\mathtt{v}_i})_\ell}}{p_{(\underline{\mathtt{w}_i})_{\ell'-R_{i-1}}^{(1)}}} + o(1),$$

where $o(1)$ is uniform due to a similar argument as (3.11). Thus, writing

$$t_{n,\ell} := \#\{j \in [1, n] : \ell = j \pmod{n_i}\} \in \left\{ \left\lfloor \frac{n}{n_i} \right\rfloor, \left\lfloor \frac{n}{n_i} \right\rfloor + 1 \right\} \qquad (1 \leq \ell \leq |\mathtt{w}^{[i]}|),$$

we have

$$\mathbb{E}_p \left(\frac{1}{k} \sum_{\ell=1}^{n} \log_{m_1} \frac{p_{(\underline{\mathtt{v}_i})_\ell}}{p_{(\underline{\mathtt{v}_i})_\ell^{(1)}}} \middle| \mathtt{w}^* \right) = \frac{1}{k} \sum_{\ell=1}^{|\underline{\mathtt{w}_i}|} t_{n,\ell} \sum_{\substack{\mathtt{a} \in A: \\ \mathtt{a}^{(1)} = (\underline{\mathtt{w}_i})_\ell^{(1)}}} \frac{p_{\mathtt{a}_\ell}}{p_{(\underline{\mathtt{w}_i})_\ell^{(1)}}} \log_{m_1} \frac{p_{\mathtt{a}_\ell}}{p_{(\underline{\mathtt{w}_i})_\ell^{(1)}}} + o(1), \tag{3.13}$$

which again is a summation of i.i.d. random variables and Lemma 3.1 applies. In summary, we have

$$\mu \left(\left| \frac{\log_{m_1} F(\mathtt{x}, k)}{k} - \mathbb{E}_p \left(\frac{\log_{m_1} F(\mathtt{x}, k)}{k} \right) \right| > \epsilon \right) = O(k^{-2}).$$

To wrap up, we note the lemma follows from the Borel-Cantelli lemma applied to our claim with $\epsilon \to 0$. □

Table 1
Values of S_i.

k_i	Case	$\liminf_{i\to\infty} \mathbb{E}_p\left[\frac{\log_{m_1} F(\mathtt{x},k_i)}{k_i}\right]$
n_i	1	N/A
	2	$H_1(p) + \log_{m_2} m_1 \cdot H_2(p)$
L_i	1	$(1+\tau_2)^{-1} \cdot H_1(p) + \max\{(1+\tau_2)^{-1}, \log_{m_2} m_1\} \cdot H_2(p)$
	2	N/A
$n_i \log_{m_1} m_2$	1	$\max\{1 - \tau_1 \log_{m_2} m_1, \log_{m_2} m_1\} \cdot H_1(p) + \log_{m_2} m_1 \cdot H_2(p)$
	2	N/A
$L_i \log_{m_1} m_2$	1	$\max\{1 - (1+\tau_2)^{-1}\tau_1, (1+\tau_2)^{-1}\} \cdot H_1(p) \log_{m_2} m_1 + (1+\tau_2)^{-1} \log_{m_2} m_1 \cdot H_2(p)$
	2	N/A
R_i	1	$(1+\tau_1)^{-1} \cdot H_1(p) + \min\{(1+\tau_1)^{-1}, \log_{m_2} m_1\} \cdot H_2(p)$
	2	$(1+\tau_1)^{-1} \cdot H_1(p) + (\log_{m_2} m_1 - (1+\tau_1)^{-1}\tau_2) \cdot H_2(p)$

Lemma 3.3. *Let $F(\mathtt{x}, k)$ be defined in (3.5). Then* $\liminf_{k\to\infty} \mathbb{E}_p\left[\frac{-\log_{m_1} F(\mathtt{x},k)}{k}\right] = \min\{t_1, t_2\}$, *where*

$$t_1 = \begin{cases} \frac{(1-(1-\gamma)\tau_1 \log_{m_2} m_1)\cdot H_1(p) + \log_{m_2} m_1 \cdot H_2(p)}{1+\tau_2} & \text{if } 1 + (1-\gamma)\tau_1 \le \log_{m_1} m_2, \\ \frac{\log_{m_2} m_1 \cdot H_1(p) + \log_{m_2} m_1 \cdot H_2(p)}{1+\tau_2} & \text{if } 1 + (1-\gamma)\tau_1 > \log_{m_1} m_2. \end{cases}$$

$$t_2 = \begin{cases} \frac{H_1(p)}{1+\tau_1} + \log_{m_2} m_1 \cdot H_2(p) & \text{if } 1 + \tau_1 \le \log_{m_1} m_2, \\ \frac{H_1(p)+H_2(p)}{1+\tau_1} & \text{if } 1 + (1-\gamma)\tau_1 \le \log_{m_1} m_2 < 1 + \tau_1, \\ \frac{H_1(p) + \log_{m_2} m_1 \cdot (1+(1-\gamma)\tau_1)\cdot H_2(p)}{1+\tau_1} & \text{if } 1 + (1-\gamma)\tau_1 > \log_{m_1} m_2. \end{cases}$$

Proof. Recall that $H(p) = H_1(p) + H_2(p)$. Calculating expectation of (3.10) and (3.12) respectively yields that

Case 1: if $n_i < k \le L_i \log_{m_1} m_2$, then

$$\mathbb{E}_p\left[\frac{\log_{m_1} F(\mathtt{x},k)}{k}\right] = \frac{1}{k}\left[(n_i - R_{i-1} + \max\{k - R_i, 0\})H_1(p)\right.$$
$$\left. + (n_i - R_{i-1} + \min\{0, k\log_{m_2} m_1 - n_i\})H_2(p)\right] + o(1);$$

Case 2: if $L_i \log_{m_1} m_2 < k \le n_i$, then

$$\mathbb{E}_p\left[\frac{\log_{m_1} F(\mathtt{x},k)}{k}\right] = \frac{1}{k}\left[(n_i - R_{i-1} + \max\{k - R_i, 0\})H_1(p)\right.$$
$$\left. + (n_i - R_{i-1} + k\log_{m_2} m_1 - L_i)H_2(p)\right] + o(1).$$

From these expressions, we observe that the interval $(n_i, n_{i+1}]$ can be divided by $\{L_i, R_i, n_i \log_{m_1} m_2, L_i \log_{m_1} m_2\}$ into at most five subintervals so that on each subinterval the function $k \mapsto \mathbb{E}_p\left[\frac{\log_{m_1} F(\mathtt{x},k)}{k}\right]$, by neglecting the $o(1)$ terms, is a polynomial of k of the form $ak^{-1} + b$ and hence monotone. This simplifies the calculation of the desired lower limit down to evaluating the minimum over the numbers $\{S_1, S_2, S_3, S_4, S_5\}$, which are defined to be $\liminf_{i\to\infty} \mathbb{E}_p\left[\frac{\log_{m_1} F(\mathtt{x},k_i)}{k_i}\right]$ with k_i taken as n_i, L_i, R_i, $n_i \log_{m_1} m_2$, and $L_i \log_{m_1} m_2$, respectively. These values could be straightforwardly calculated and are summarized in Table 1. Since $H_1(p), H_2(p), \tau_1, \tau_2 \ge 0$, we observe the following order among these values.

- $S_1 \ge S_3 \ge S_5$. The inequality $S_1 \ge S_3$ is clear by comparing the coefficients of $H_1(p)$ and of $H_2(p)$. As for $S_3 \ge S_5$, it is done by comparing the coefficients as follows.

$$\max\{1-\tau_1\log_{m_2} m_1, \log_{m_2} m_1\} = \begin{cases} \log_{m_2} m_1 \geq (1+\tau_1)^{-1} & \text{if } \log_{m_1} m_2 \leq 1+\tau_1, \\ 1-\tau_1\log_{m_2} m_1 < (1+\tau_1)^{-1} & \text{if } \log_{m_1} m_2 \geq 1+\tau_1, \end{cases}$$

$$\log_{m_2} m_1 \geq \min\{(1+\tau_1)^{-1}, \log_{m_2} m_1\} \quad \textit{and} \quad \log_{m_2} m_1 - (1+\tau_1)^{-1}\tau_2.$$

- $S_2 \geq S_4$ if $\tau_1 \geq \tau_2$. Note that the coefficient of $H_1(p)$ in S_4 could be viewed as the maximum of two linear functions of $\log_{m_2} m_1$, which yields

$$\max\{1-(1+\tau_2)^{-1}\tau_1\log_{m_2} m_1, (1+\tau_2)^{-1}\log_{m_2} m_1\} \leq (1+\tau_1)^{-1}.$$

  Under our assumption, this is less than the corresponding coefficient $(1+\tau_2)^{-1}$ in S_2. The order between the coefficients of $H_2(p)$ is clear.

Now we plug $\tau_2 = \gamma\log_{m_2} m_1 \cdot \tau_1$ into S_4 and S_5 to deduce

$$S_4 = \begin{cases} \frac{(1-(1-\gamma)\tau_1\log_{m_2} m_1)\cdot H_1(p)+\log_{m_2} m_1\cdot H_2(p)}{1+\tau_2} & \text{if } 1+(1-\gamma)\tau_1 \leq \log_{m_1} m_2, \\ \frac{\log_{m_2} m_1\cdot H_1(p)+\log_{m_2} m_1\cdot H_2(p)}{1+\tau_2} & \text{if } 1+(1-\gamma)\tau_1 > \log_{m_1} m_2. \end{cases}$$

$$S_5 = \begin{cases} \frac{H_1(p)}{1+\tau_1} + \log_{m_2} m_1\cdot H_2(p) & \text{if } 1+\tau_1 \leq \log_{m_1} m_2, \\ \frac{H_1(p)+H_2(p)}{1+\tau_1} & \text{if } 1+(1-\gamma)\tau_1 \leq \log_{m_1} m_2 < 1+\tau_1, \\ \frac{H_1(p)+\log_{m_2} m_1\cdot(1+(1-\gamma)\tau_1)\cdot H_2(p)}{1+\tau_1} & \text{if } 1+(1-\gamma)\tau_1 > \log_{m_1} m_2. \end{cases}$$

The proof is then completed. □

Remark 3.1. The essential difference between the measure μ constructed in this section and its counterpart in [4] lies in the subwords $\underline{\mathtt{u}_j}$ and $\underline{\mathtt{v}_j}$ $(1 \leq j \leq i)$. Specifically, these subwords satisfy (3.3) in our work, while they fulfill a priori assumptions imposed on the shrinking target problem in [4]. The nuance urges necessary modifications in the proof of constant lower limit (Lemma 3.2) and is reflected in the expression (3.13). However, the subtlety is subsumed in $\mathbb{E}_p\left[\frac{-\log_{m_1} F(\mathtt{x},k)}{k}\right]$ and in the homogeneity assumption of the Bedford-McMullen carpet; therefore, it is barely visible beyond the scope of the proof of Lemma 3.2.

Finally, we wrap up this section by noting that the lower bound of $\dim_{\mathrm{H}} W_\gamma(K,T,\psi)$ in Theorem 1.1 holds as a consequence of Lemma 3.2 and 3.3, for which we simply take $p_{\mathtt{a}} = \frac{1}{\#A}$ for all $\mathtt{a} \in A$ so that $H_1(p) = \log_{m_1} M$ and $H_2(p) = \log_{m_1} N$.

4. Proof of the upper bound

In this section, we give the rest of the proof of Theorem 1.1. We note that it suffices to prove the case where $\lim_{n\to\infty} -\frac{\log\psi(n)}{n}$ exists, since for those $\psi(n)$ without this property, we may consider $\phi(n) := \max\{\psi(n), m_1^{-\tau_1\cdot n}\}$ so that $\dim_{\mathrm{H}} W_\gamma(K,T,\psi) \leq \dim_{\mathrm{H}} W_\gamma(K,T,\phi)$ and we may apply the aforementioned result for the case where $\lim_{n\to\infty} -\frac{\log\psi(n)}{n}$ exists.

Note that the set $W_\gamma(K,T,\psi)$ can be written as a limsup set

$$W_\gamma(K,T,\psi) = \limsup_{n\to\infty} W_n^\gamma(K,T,\psi),$$

where

$$W_n^\gamma(K,T,\psi) := \left\{ x \in K : \begin{cases} |x^{(1)} - T_{m_1}^n(x^{(1)})| < \psi(n) \\ |x^{(2)} - T_{m_2}^n(x^{(2)})| < \psi(n)^\gamma \end{cases} \right\}.$$

We note that the set above can further be written as the following union:

$$W_n^\gamma(K,T,\psi) = \bigcup_{\mathtt{w}\in A^n} J_n(\mathtt{w}) := \bigcup_{\mathtt{w}\in A^n} W_n^\gamma(K,T,\psi)\cap I(\mathtt{w}).$$

For sufficiently large n, the set $J_n(\mathtt{w})$ is contained in the interior of a rectangle $R^{(1)}(\mathtt{w})\times R^{(2)}(\mathtt{w})$ of width $4\psi(n)m_1^{-n}$ and height $4\psi(n)^\gamma m_2^{-n}$, since for any $\pi(\mathtt{x})\in J_n(\mathtt{w})$,

$$\begin{aligned}|\pi(\mathtt{x})^{(i)} - \pi(\mathtt{w}^\infty)^{(i)}| &\geq |T^n(\pi(\mathtt{x}))^{(i)} - T^n(\pi(\mathtt{w}^\infty))^{(i)}| - |\pi(\mathtt{x})^{(i)} - T^n(\pi(\mathtt{x})^{(i)})| \\ &\geq m_i^n|\pi(\mathtt{x})^{(i)} - \pi(\mathtt{w}^\infty)^{(i)}| - |\pi(\mathtt{x})^{(i)} - T^n(\pi(\mathtt{x})^{(i)})|,\end{aligned} \tag{4.1}$$

$|\pi(\mathtt{x})^{(1)} - T^n(\pi(\mathtt{x})^{(1)})| \leq \psi(n)m_1^{-n}$ and $|\pi(\mathtt{x})^{(1)} - T^n(\pi(\mathtt{x})^{(1)})| \leq \psi(n)^\gamma m_2^{-n}$. By Remark 1.1, we know that $0<\tau_1<\infty$, then there exists $G_1\in\mathbb{N}$ such that for any $n\geq G_1$, $\psi(n)<1$, $(\frac{m_2}{m_1})^n>4$. Now for any given $\delta>0$, we can choose G large enough so that for any $n\geq\max\{G,G_1\}$, $\max\{4\psi(n)m_i^{-n}, 4\psi(n)^\gamma m_2^{-n}\}<\delta$. For any $n\geq 1$, let

$$\psi_1(n) = \psi(n) \text{ and } \psi_2(n) = \psi(n)^\gamma. \tag{4.2}$$

For $i=1,2$, we denote by $\mathcal{B}_{i,n}$ the smallest number of balls of diameter $4\psi_i(n)m_i^{-n}$ (the sidelength of the rectangles in $R^{(1)}(\mathtt{w})\times R^{(2)}(\mathtt{w})$ along the direction of the i-th axis) needed to cover $W_n^\gamma(K,T,\psi)$, the s-dimensional Hausdorff measure has the following estimate:

$$\mathcal{H}_\delta^s(W_\gamma(K,T,\psi)) \leq \sum_{n=G}^\infty \#\mathcal{B}_{i,n}\cdot(4\psi_i(n)m_i^{-n})^s.$$

We now estimate the value of $\#\mathcal{B}_{i,n}$, $i=1,2$.

For $i=2$, we need to estimate the number of balls $\mathcal{B}_{2,n}$ of diameter $4\psi(n)^\gamma m_2^{-n}$ to cover $W_n^\gamma(K,T,\psi) = \cup_{\mathtt{w}\in A^n} J_n(\mathtt{w})$. Since $J_n(\mathtt{w})$ is contained in the interior of a rectangle $R^{(1)}(\mathtt{w})\times R^{(2)}(\mathtt{w})$ of width $4\psi(n)m_1^{-n}$ and height $4\psi(n)^\gamma m_2^{-n}$. To estimate the values, we consider two possible cases depending on the values of $(1-\gamma)\tau_1$ and $\log_{m_1} m_2 - 1$.

Case a-1: $(1-\gamma)\tau_1 > \log_{m_1} m_2 - 1$. By the definition of τ_1, $4\psi(n)^\gamma m_2^{-n} > 4\psi(n)m_1^{-n}$. For any $n\geq G_1$, $m_1^{-n}\geq 4\psi(n)^\gamma m_2^{-n}$ by the fact that $\psi(n)<1$ and $m_1\leq m_2$. It is clear that $\#\mathcal{B}_{2,n}\leq (MN)^n$. Therefore,

$$\begin{aligned}\mathcal{H}_\delta^s(W_\gamma(K,T,\psi)) &\leq \sum_{n=G}^\infty (MN)^n(4\psi(n)^\gamma m_2^{-n})^s \\ &= C\sum_{n=G}^\infty m_1^{nh_n},\end{aligned} \tag{4.3}$$

where $C:=4^s$ is a constant and

$$h_n = \log_{m_1} M + \log_{m_1} N + s(n^{-1}\gamma\log_{m_1}\psi(n) - \log_{m_1} m_2).$$

Since we assume $\lim_{n\to\infty} -\frac{\log\psi(n)}{n}$ exists, it is clear that $\sum_{n=G}^\infty m_1^{n\cdot h_n}$ converges as long as $\lim_{n\to\infty} h_n<0$ and that this is equivalent to the condition that

$$s > \lim_{n\to\infty}\frac{\log_{m_1} M + \log_{m_1} N}{\log_{m_1} m_2 - n^{-1}\gamma\log_{m_1}\psi(n)} = \frac{\log_{m_2} M + \log_{m_2} N}{1+\tau_2}.$$

This, together with (4.3), yields that

$$0 \le \mathcal{H}^s(W_\gamma(K,T,\psi)) = \lim_{\delta\to 0}\mathcal{H}^s_\delta(W_\gamma(K,T,\psi)) \le \lim_{G\to\infty} C\sum_{n=G}^{\infty} m_1^{n\cdot h_n} = 0,$$

which implies that $\dim_{\mathrm{H}} W_\gamma(K,T,\psi) \le \frac{\log_{m_2} M + \log_{m_2} N}{1+\tau_2}$.

Case a-2: $(1-\gamma)\tau_1 \le \log_{m_1} m_2 - 1$. By the definition of τ_1, $4\psi(n)^\gamma m_2^{-n} \le 4\psi(n)m_1^{-n}$. Notice that just a ball of diameter $4\psi(n)^\gamma m_2^{-n}$ is needed to cover $J_n(\mathtt{w})$ along the direction of the second axis. Now let us look in the direction along the first axis. Let $u > 0$ be the unique integer such that $m_1^{-u} \le 4\psi(n)m_1^{-n} \le m_1^{-u+1}$, which implies there exist $\mathtt{w}'$, $\mathtt{w}'' \in A^{u-1}$ such that

$$J_n(\mathtt{w}) \subseteq (I^{(1)}_{u-1}(\mathtt{w}') \times R^{(2)}(\mathtt{w}) \cup I^{(1)}_{u-1}(\mathtt{w}'') \times R^{(2)}(\mathtt{w})) \cap K, \tag{4.4}$$

where $I^{(1)}_{u-1}(\mathtt{w}')$ is the projection of the rectangle $I_{u-1}(\mathtt{w}')$ onto the first axis, $I^{(1)}_{u-1}(\mathtt{w}'')$ is that of $I_{u-1}(\mathtt{w}'')$, and K is the Bedford-McMullen carpet. Since $4\psi(n)^\gamma m_2^{-n} \le 4\psi(n)m_1^{-n}$, there exists a unique integer $v \ge 0$ such that

$$m_1^{-u-v} \le 4\psi(n)^\gamma m_2^{-n} \le m_1^{-u-v+1}, \tag{4.5}$$

which yields that $v \le 1 - n + n\log_{m_1} m_2 + (1-\gamma)\log_{m_1}\psi(n)$. Then $J_n(\mathtt{w})$ can be covered by $2M^{v+1}$ balls of diameter $4\psi(n)m_2^{-n}$, which follows from that M denotes the number of columns containing at least one chosen rectangle. By (4.4), (4.5) and the definition of $\mathcal{B}_{2,n}$,

$$\#\mathcal{B}_{2,n} \le 2M^{v+1}\cdot(\#A)^n = 2M^{v+1}(MN)^n.$$

The remaining argument is just the same as **Case a-1**. We therefore conclude that

$$\dim_{\mathrm{H}} W_\gamma(K,T,\psi) \le (1 - \frac{\tau_1\log_{m_2} m_1}{1+\tau_2})\log_{m_1} M + \frac{\log_{m_2} N}{1+\tau_2}.$$

For $i = 1$, we need to estimate the number of balls $\mathcal{B}_{1,n}$ of diameter $4\psi(n)m_1^{-n}$ to cover $W_n(K,T,\psi) = \cup_{\mathtt{w}\in A^n} J_n(\mathtt{w})$. To estimate the values, we consider two possible cases depending on the values of $\log_{m_1} m_2$ and $1+\tau_1$.

Case b-1: $\log_{m_1} m_2 > 1+\tau_1$. By the definition of τ_1, $4\psi(n)m_1^{-n} > m_2^{-n}$. Now observe that a ball of diameter $4\psi(n)m_1^{-n}$ will also cover other n-th level rectangles in $\bigcup_{\mathtt{w}\in A^n} R^{(1)}(\mathtt{w}) \times R^{(2)}(\mathtt{w})$ along the second axis. Let $j = \lfloor \log_{m_2}\psi(n) - n\log_{m_2} m_1 + n \rfloor - 2$, then $\#\mathcal{B}_{1,n} \le (MN)^n N^{-j}$, which yields that

$$\mathcal{H}^s_\delta(W_\gamma(K,T,\psi)) \le \sum_{n=G}^{\infty}(MN)^n N^{-j}(4\psi(n)m_1^{-n})^s.$$

Such an argument also applies to the case $i = 1$, from which we conclude that $\dim_{\mathrm{H}} W_\gamma(K,T,\psi) \le \log_{m_2} N + \frac{\log_{m_1} M}{1+\tau_1}$.

Case b-2: $\log_{m_1} m_2 \le 1+\tau_1$. By the definition of τ_1, $4\psi(n)m_1^{-n} \le m_2^{-n}$.

Case b-2-1: $(1-\gamma)\tau_1 < \log_{m_1} m_2 - 1$. By the definition of τ_1, $4\psi(n)^\gamma m_2^{-n} < 4\psi(n)m_1^{-n}$. It is clear that $\#\mathcal{B}_{1,n} \le (MN)^n$. Therefore,

$$\mathcal{H}^s_\delta(W_\gamma(K,T,\psi)) \le \sum_{n=G}^{\infty}(MN)^n(4\psi(n)m_1^{-n})^s.$$

The remaining argument is just the same as the case when $i = 1$. We therefore conclude that

$$\dim_{\mathrm{H}} W_\gamma(K,T,\psi) \le \frac{\log_{m_1} M + \log_{m_1} N}{1+\tau_1}.$$

Case b-2-2: $(1-\gamma)\tau_1 > \log_{m_1} m_2 - 1$. By the definition of τ_1, $4\psi(n)^\gamma m_2^{-n} \ge 4\psi(n) m_1^{-n}$. Let $u > 0$ be the unique integer such that $m_1^{-u} \le 4\psi(n)^\gamma m_2^{-n} \le m_1^{-u+1}$, which implies there exist $\mathtt{w}'$, $\mathtt{w}'' \in A^{u-1}$ such that

$$J_n(\mathtt{w}) \subseteq (I^{(1)}_{u-1}(\mathtt{w}') \times R^{(2)}(\mathtt{w}) \cup I^{(1)}_{u-1}(\mathtt{w}'') \times R^{(2)}(\mathtt{w})) \cap K, \tag{4.6}$$

where $I^{(1)}_{u-1}(\mathtt{w}')$ is the projection of the rectangle $I_{u-1}(\mathtt{w}')$ onto the first axis, $I^{(1)}_{u-1}(\mathtt{w}'')$ is that of $I_{u-1}(\mathtt{w}'')$, and K is the Bedford-McMullen carpet. Since $4\psi(n)^\gamma m_2^{-n} \ge 4\psi(n) m_1^{-n}$, there exists a unique integer $v \ge 0$ such that

$$m_1^{-u-v} \le 4\psi(n) m_1^{-n} \le m_1^{-u-v+1}, \tag{4.7}$$

which yields that $v \le 1 - n + n\log_{m_2} m_1 + (\gamma-1)\log_{m_2}\psi(n)$. Then $J_n(\mathtt{w})$ can be covered by $2N^{v+1}$ balls of diameter $4\psi(n)m_1^{-n}$. By (4.6), (4.7) and the definition of $\mathcal{B}_{1,n}$,

$$\#\mathcal{B}_{1,n} \le 2N^{v+1}\cdot(\#A)^n = 2N^{v+1}(MN)^n.$$

The remaining argument is just the same as **Case a-1**. We therefore conclude that

$$\dim_{\mathrm{H}} W_\gamma(K,T,\psi) \le \frac{(1+(1-\gamma)\tau_1)\log_{m_2} N + \log_{m_1} M}{1+\tau_1}.$$

From the discussion above, we obtain the upper bound for Theorem 1.1.

5. Applications

In this section, we present some examples. Example 5.1 is a special case of Theorem 1.1 in which $m_1 = M$ and $m_2 = N$.

Example 5.1. Let $T : [0,1]^2 \to [0,1]^2$ be an integer diagonal matrix transformation of $[0,1]^2$, i.e., $T(x) = (T_{m_1}(x^{(1)}), T_{m_2}(x^{(2)})) := \big(m_1 x^{(1)} \pmod 1, m_2 x^{(2)} \pmod 1\big)$, with $2 \le m_1 \le m_2$. Suppose $\psi : \mathbb{N}^+ \to \mathbb{R}^+$ be a real positive function. Let

$$W(T,\psi) := \big\{\mathtt{x} \in [0,1]^2 : T^n(x) \in B(x,\psi(n)) \text{ for infinitely many } n.\big\}$$

Then,

$$\dim_{\mathrm{H}} W(T,\psi) = \begin{cases} \min\left\{\frac{2}{1+\tau_2}, \frac{1}{1+\tau_1}+1\right\} & \text{if } \log_{m_1} m_2 > 1+\tau_1, \\ \min\left\{\frac{2}{1+\tau_2}, \frac{1+\log_{m_1} m_2}{1+\tau_1}\right\} & \text{if } \log_{m_1} m_2 \le 1+\tau_1, \end{cases}$$

where τ_i, $i = 1,2$, is defined in (1.4).

The following example illustrates the case where K is a product of Cantor sets.

Example 5.2. Let $\mathcal{C}_{\frac{1}{3}}$ denote the middle third Cantor set and $\mathcal{C}_{\frac{1}{4}}$ be the attractor of the iterated function system $\{f_1, f_2, f_3\}$ on $[0,1]$, where $f_1(x) = \frac{1}{4}x$, $f_2(x) = \frac{1}{4}x + \frac{1}{4}$ and $f_3(x) = \frac{1}{4}x + \frac{3}{4}$. Define $T : \mathcal{C}_{\frac{1}{3}} \times \mathcal{C}_{\frac{1}{4}} \to \mathcal{C}_{\frac{1}{3}} \times \mathcal{C}_{\frac{1}{4}}$ as

$$T(x) = \left(T_3(x^{(1)}), T_4(x^{(2)})\right) := \left(3x^{(1)} \pmod 1, 4x^{(2)} \pmod 1\right).$$

Suppose $\psi : \mathbb{N}^+ \to \mathbb{R}^+$ is a real positive function. Let

$$W_1(T,\psi) := \left\{ \mathrm{x} \in \mathcal{C}_{\frac{1}{3}} \times \mathcal{C}_{\frac{1}{4}} : T^n(x) \in B(x, \psi(n)) \text{ for infinitely many } n \right\}.$$

Then,

$$\dim_{\mathrm{H}} W_1(T,\psi) = \begin{cases} \min\left\{ \frac{\log_3 2 + \log_4 3}{1+\tau_2}, \frac{\log_3 2}{1+\tau_1} + \log_4 3 \right\} & \text{if } \log_3 4 > 1 + \tau_1, \\ \min\left\{ \frac{\log_3 2 + \log_4 3}{1+\tau_2}, \frac{\log_3 2 + 1}{1+\tau_1} \right\} & \text{if } \log_3 4 \le 1 + \tau_1, \end{cases}$$

where τ_i, $i = 1, 2$, is defined in (1.4).

Acknowledgments

The authors would like to thank Professor Bing Li for his guidance and suggestions. The authors would like to thank Professor Meng Wu for drawing our attention to these problems and reading the early draft of the manuscript. The authors thank the referees for their valuable and very detailed suggestions.

References

[1] D. Allen, B. Bárány, On the Hausdorff measure of shrinking target sets on self-conformal sets, Mathematika 67 (2021) 807–839.
[2] S. Baker, M. Farmer, Quantitative recurrence properties for self-conformal sets, Proc. Am. Math. Soc. 149 (2021) 1127–1138.
[3] S. Baker, H. Koivusalo, Quantitative recurrence and the shrinking target problem for overlapping iterated function systems, Preprint, https://arxiv.org/abs/2305.01489.
[4] B. Bárány, M. Rams, Shrinking targets on Bedford–McMullen carpets, Proc. Lond. Math. Soc. 117 (2018) 951–995.
[5] B. Bárány, S. Troscheit, Dynamically defined subsets of generic self-affine sets, Nonlinearity 35 (2022) 4986–5013.
[6] L. Barreira, B. Saussol, Hausdorff dimension of measures via Poincaré recurrence, Commun. Math. Phys. 219 (2001) 443–463.
[7] T. Bedford, Crinkly curves, Markov partitions and box dimensions in self-similar sets, Thesis (Ph. D.), The University of Warwick, 1984.
[8] M.D. Boshernitzan, Quantitative recurrence results, Invent. Math. 113 (1993) 617–631.
[9] Y. Bugeaud, A note on inhomogeneous Diophantine approximation, Glasg. Math. J. 45 (2003) 105–110.
[10] Y.-Y. Chang, M. Wu, W. Wu, Quantitative recurrence properties and homogeneous self-similar sets, Proc. Am. Math. Soc. 147 (2019) 1453–1465.
[11] K.J. Falconer, Techniques in Fractal Geometry, vol. 3, Wiley, Chichester, 1997.
[12] A.-H. Fan, J. Schmeling, S. Troubetzkoy, A multifractal mass transference principle for Gibbs measures with applications to dynamical Diophantine approximation, Proc. Lond. Math. Soc. 107 (2013) 1173–1219.
[13] R. Hill, S.L. Velani, The ergodic theory of shrinking targets, Invent. Math. 119 (1995) 175–198.
[14] R. Hill, S.L. Velani, A zero–infinity law for well-approximable points in Julia sets, Ergod. Theory Dyn. Syst. 22 (2002) 1773–1782.
[15] M. Hussain, B. Li, D. Simmons, B.-W. Wang, Dynamical Borel-Cantelli lemma for recurrence theory, Ergod. Theory Dyn. Syst. 42 (2022) 1994–2008.
[16] M.S. Kirsebom, P. Kunde, T. Persson, Dynamical Borel-Cantelli lemma for recurrence theory, Ann. Sc. Norm. Super. Pisa, Cl. Sci. (2023) 1499–1535.
[17] D. Kleinbock, J.-J. Zheng, Dynamical Borel Cantelli lemma for recurrence under Lipschitz twists, Nonlinearity 36 (2023) 1434–1460.
[18] H. Koivusalo, F.A. Ramírez, Recurrence to shrinking targets on typical self-affine fractals, Proc. Edinb. Math. Soc. 61 (2018) 387–400.
[19] A. Käenmäki, T. Rajala, V. Suomala, Local multifractal analysis in metric spaces, Nonlinearity 26 (2013) 2157–2173.
[20] B. Li, L.-M. Liao, S. Velani, E. Zorin, The shrinking target problem for matrix transformations of tori: revisiting the standard problem, Adv. Math. 421 (2023) 108994.
[21] B. Li, B.-W. Wang, J. Wu, J. Xu, The shrinking target problem for matrix transformations of tori: revisiting the standard problem, Proc. Lond. Math. Soc. 108 (2014) 159–186.
[22] L.-M. Liao, S. Seuret, Diophantine approximation by orbits of expanding Markov maps, Ergod. Theory Dyn. Syst. 33 (1984) 585–608.
[23] C. McMullen, The Hausdorff dimension of general Sierpiński carpets, Nagoya Math. J. 96 (2013) 1–9.

[24] T. Persson, A strong Borel-Cantelli lemma for recurrence, Stud. Math. 268 (2023) 75–89.
[25] T. Persson, M. Rams, On shrinking targets for piecewise expanding interval maps, Ergod. Theory Dyn. Syst. 37 (2017) 646–663.
[26] S. Seuret, B.-W. Wang, Quantitative recurrence properties in conformal iterated function systems, Adv. Math. 280 (2015) 472–505.
[27] L.-M. Shen, B.-W. Wang, Shrinking target problems for beta-dynamical system, Sci. China Math. 56 (2013) 91–104.
[28] B. Tan, B.-W. Wang, Quantitative recurrence properties in conformal iterated function systems, Adv. Math. 228 (2011) 2071–2097.
[29] J. Tseng, On circle rotations and the shrinking target properties, Discrete Contin. Dyn. Syst. 20 (2008) 1111–1122.

II

Stochastics and Dynamics
Vol. 25, No. 1 (2025) 2550007 (26 pages)

DOI: 10.1142/S0219493725500078

On the topological pressure of axial product on trees

Jung-Chao Ban

Department of Mathematical Sciences,
National Chengchi University,
Taipei 11605, Taiwan, ROC
Mathematics Division,
National Center for Theoretical Science,
National Taiwan University,
Taipei 10617, Taiwan, ROC
jcban@nccu.edu.tw

Yu-Liang Wu*

Department of Mathematical Sciences,
University of Oulu, P. O. Box 3000,
Oulu 90014, Finland
yu-liang.wu@oulu.fi

Received 20 January 2023
Revised 9 October 2024
Accepted 11 February 2025
Published 18 March 2025

This paper investigates the topological pressure of isotropic axial products of Markov subshifts on the d-tree. We show that the quantity increases with dimension d. To achieve this, we introduce the pattern distribution vectors and the associated transition matrices and partially transplant the large deviation theory to tree-shifts. Additionally, we apply our main result to a broader class of shift spaces, accompanied by numerical experiments for verification.

Keywords: Tree-shifts; monotonicity; pressure.

AMS Subject Classification: 37B10, 60C05, 37N40, 37D35

1. Introduction

This paper investigates the topological pressure of the axial product of a Markov subshift on the d-tree, inspired by recent studies on the limiting entropy of the axial product of $\mathbb{N}^d$ [11, 12] and the asymptotic pressure of the axial product on d-tree [15]. Before presenting the main results, the motivation behind the study is outlined below.

*Corresponding author.

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

$$\bigotimes_{i=1}^{d} X_i = \{x \in \mathcal{A}^{\mathbb{N}^d} : \forall\, g \in \mathbb{N}^d\ \forall\, i \in \{1, \ldots, d\}, x_{g+\mathbb{Z}_+\mathsf{e}_i} \in X_i\}, \tag{1.1}$$

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

$$\times_{i=1}^{d} X_i = \{x \in \mathcal{A}^{\mathcal{T}^{(d)}} : \forall\, g \in \mathcal{T}^{(d)}\ \forall\, i \in \{1, \ldots, d\}, (x_{gf_i^n})_{n\in\mathbb{Z}_+} \in X_i\}. \tag{1.2}$$

An axial product $\bigotimes_{i=1}^d X_i$ (or $\times_{i=1}^d X_i$) is said to be *isotropic* if $X_i = X_j$ for all $1 \leq i \neq j \leq d$, and *anisotropic* if otherwise. Isotropic axial products of shifts on $\mathbb{N}^d$ were first introduced in [11], and many important physical systems, such as the hard square model on $\mathbb{N}^2$ or $\mathbb{Z}^2$, belong to this class. In the context of tree-shifts, a notable subclass of isotropic axial products is the family of *hom tree-shifts*,[a] where X_i are one-sided Markov subshifts and identical. This family is widely studied in the literature [1–3, 7, 13–15] and has attracted growing attention since the subshifts defined on it exhibit rich and diverse phenomena in topological (cf. [2]) and statistical contexts (cf. [4, 5]).

One of the key quantities that characterize such symbolic dynamical systems is topological entropy, which, in this context, measures the asymptotic growth in the number of admissible patterns with finite supports in the following manner: for a subshift $X \subseteq \mathcal{A}^G$ ($G = \mathbb{N}^d$ or $\mathbb{Z}^d$), its *topological entropy* is defined as

$$h(X) = \lim_{n\to\infty} \frac{\log |\pi([1,n]^d, X)|}{|[1,n]^d|}, \tag{1.3}$$

where $\pi : G \times \mathcal{A}^G \to \mathcal{A}^F$ is the *canonical projection* onto F defined by $(\pi(F,x))_g = x_g$, $|F|$ indicates the number of elements in a set F, and the limit is known to exist because $\mathbb{N}^d/\mathbb{Z}^d$ is an amenable semigroup/group (cf. [9]) and $\{[1,n]^d\}_{n=1}^{\infty}$ is a *Følner sequence*. Analogously, we define the *topological entropy* of a tree-shift $T^{(d)} \subseteq \mathcal{A}^{\mathcal{T}^{(d)}}$ to be

$$h(T^{(d)}) = \lim_{n\to\infty} \frac{\log |\pi(\Delta_n^{(d)}, T^{(d)})|}{|\Delta_n^{(d)}|}, \tag{1.4}$$

where $\Xi_n^{(d)} = \{f_1, f_2, \ldots, f_d\}^n$, $\Delta_n^{(d)} = \bigcup_{i=0}^n \Xi_i^{(d)}$, and the existence of the limit (1.4) was proved in [13]. For more on the existence of the limit for shifts defined on a large class of trees, see [6]. It is important to note that despite similar

[a]Such shifts are also called the associated tree shifts in [13, 14].

definitions of topological entropy for these two types of systems, the resulting quantities differ significantly in various ways. For instance, there are notable differences (see [4]) in the structures of $\{h(T) : T \text{ is a Markov tree-shift on } \mathcal{T}^{(d)}\}$ and $\{h(X) : X \text{ is a Markov shift on } \mathbb{N}^d\}$.

The theme of this paper centers on the behavior of the topological entropy as the dimension of the underlying lattice grows. This concept was introduced by Louidor *et al.* [11], leading to the study of *limiting entropy* defined as

$$h^{(\infty)}(X) = \lim_{d\to\infty} h\left(\bigotimes_{i=1}^{d} X\right),$$

where the limit exists since $h(\bigotimes_{i=1}^{d} X)$ is non-increasing in d. Meyerovitch and Pavlov [12] later proved that this asymptotic quantity coincides with the independence entropy (see the references for a definition). A similar investigation on the limiting entropy of $\times_{i=1}^{d} X$ was initiated by Petersen and Salama. In [14], they focused on X_G, the golden-mean[b] subshift and employed the method of *site strip approximation* to prove that $h(X_G^{\times d})$ is strictly increasing in d [14, Theorem 3.7].

Recently, Petersen and Salama [15] extended the notion of limiting entropy to *topological pressure*.[c] Under a similar setting to [8], they introduced the idea of topological pressure to isotropic axial products on a broad class of trees called generalized Fibonacci trees.[d] In this paper, we focus on the cases where the underlying lattices are d-trees. Under the circumstances, one may consider a physical system of particles sitting on the d-tree so that the energy assigned to a particle of type $a \in \mathcal{A}$ comes solely from the influence of the ambient field $\log w_a$ ($w \in \mathbb{R}^{\mathcal{A}}_{\geq 0}$) and the pairwise interaction $\log A_{a,b}$ ($A \in \mathbb{R}^{\mathcal{A}\times\mathcal{A}}_{\geq 0}$) with neighboring particle of type $b \in \mathcal{A}$. Collectively, such information is embedded in a single matrix $E \in \mathbb{R}^{\mathcal{A}\times\mathcal{A}}_{\geq 0}$ defined as $E_{a,b} = w_a A_{a,b}$. Under these assumptions, the configurations adhering to the pairwise interaction constraint A naturally form an isotropic axial product $X_E^{\times d}$, where

$$X_E = \{x \in \mathbb{Z}_+ : E_{x_{n+1},x_n} > 0, \forall\, n \in \mathbb{Z}_+\}$$

and one can take the same route as in the study of thermodynamic formalism to define the *partition function* $|Z_n(X_E^{\times d}, E)|$ on the finite subsystem Δ_n as

$$|Z_n(X_E^{\times d}, E)| = \sum_{u\in\pi(\Delta_n^{(d)}, X_E^{\times d})} w_{u_\epsilon} \prod_{g\in\Delta_{n-1}^{(d)}} \prod_{i=1}^{d} E_{u_{gf_i},u_g} \tag{1.5}$$

and the *topological pressure* as

$$\mathbf{P}(X_E^{\times d}, E) = \lim_{n\to\infty} \frac{\log |Z_n(X_E^{\times d}, E)|}{|\Delta_n^{(d)}|}, \tag{1.6}$$

[b]A shift space with two symbols $0, 1$ that preclude the existence of neighboring 1s.
[c]They refer to it as *asymptotic pressure*.
[d]See [15] for a definition.

where the existence of the limit was proved in [15, Theorem 3.10], and $\mathbf{P}(X_E^{\times d}, E) = h(X_E^{\times d})$ if E is a 0-1 matrix. Moreover, the authors showed under the assumption

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

that the topological pressure $\mathbf{P}(X_E^{\times d}, E)$ converges asymptotically, as $d \to \infty$, to

$$\log r_E := \log \max\left\{\sum_{a\in\mathcal{A}} E_{a,b} : b \in \mathcal{A}\right\}, \tag{1.7}$$

generalizing the limiting entropy to limiting pressure. Building upon this, this paper further establishes that $\mathbf{P}(X_d^{\times E}, E)$ is increasing in d with the aid of *pattern distribution*. Specifically, by letting $\Gamma_{\mathcal{A}}$ represent the set of all probability vectors indexed by $\mathcal{A}$ and $\Upsilon_{\mathcal{A}}$ be the set of left stochastic matrices acting on $\Gamma_{\mathcal{A}}$, we successfully relate the topological pressure to the maximum $P^{(\infty)}(d, E)$ of the following optimization problem:

$$\begin{cases} \text{maximize} & \displaystyle\sum_{j=0}^{\infty} \frac{d-1}{d^{j+1}} \sum_{a\in\mathcal{A}} \sum_{b\in\mathcal{A}} \mathsf{p}_a^{(j+1)} \mathsf{P}_{b,a}^{(j)} \log \frac{E_{b,a}}{\mathsf{P}_{b,a}^{(j)}}, \\ \text{subject to} & \mathsf{p} \in \Gamma_{\mathcal{A}}^{\mathbb{Z}_+}, \quad \mathsf{P} \in \Upsilon_{\mathcal{A}}^{\mathbb{Z}_+}, \\ & \mathsf{P}_{a,b}^{(j)} = 0 \quad \text{if } E_{a,b} = 0, \\ & \mathsf{p}^{(j)} = \mathsf{P}^{(j)} \mathsf{p}^{(j+1)}, \quad 0 \le j < \infty, \end{cases} \tag{Problem 1}$$

obtaining our main result.

Theorem 1.1. *Suppose that the interaction matrix E satisfies* (A). *Then, the following assertions hold*:

(a) $P^{(\infty)}(d, E)$ *is continuous and increasing in d on* $(1, \infty)$.
(b) $P^{(\infty)}(d, E) = \mathbf{P}(X_E^{\times d}, E)$ *for all* $d \in \mathbb{N}\backslash\{1\}$.
(c) *Let $\rho(E)$ be the spectral radius of E and r_E be as in* (1.7). *We have*

$$\lim_{d\to 1+} P^{(\infty)}(d, E) = \log \rho(E) \quad \text{and} \quad \lim_{d\to\infty} P^{(\infty)}(d, E) = \log r_E. \tag{1.8}$$

Heuristically speaking, the maximizing probability vectors p and stochastic matrices P essentially capture the characteristics of a "typical configuration" in $X_E^{\times d}$ as discussed in Sec. 3. Applying this theorem, the following corollary partially generalizes a result discussed in [16] regarding the monotone convergence of entropy for subshifts on free groups.

Corollary 1.1. *Let $\mathcal{G}^{(d)}$ be a free group with d generators, $E = E^T$ satisfy assumption* (A), *and $Y_E^{(d)}$ be a Markov shift space over $\mathcal{G}^{(d)}$. Then, the entropy $h(Y_E^{(d)})$ is increasing in d.*

This follows from the fact that $h(Y_E^{(d)}) = h(X_E^{\times(2d-1)})$, as is proved in [6, Proposition 4.5].

A few remarks could be made at this point.

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

(2) From Theorem 1.1, if E is a binary matrix, then $\mathbf{P}(X_E^{\times d}, E) = h(X_E^{\times d})$ is increasing in d, as opposed to $h(\bigotimes_{i=1}^d X_E)$, which is decreasing in d.

(3) Aside from its capability to demonstrate that $\mathbf{P}(X_E^{\times d}, E)$ is increasing in $d \geq 2$, the function $d \mapsto P^{(\infty)}(d, E)$ is also interesting in that it interpolates topological pressure between tree-shifts $X_E^{\times d}$ and one-sided subshift $X_E = X_E^{\times 1}$, given that $P^{(\infty)}(1+, E) = \log \rho(E)$ coincides with the topological pressure of the potential $t \mapsto \log E_{t_1,t_0}$ defined for $t \in X_E$.

The paper is organized as follows. Section 2 provides background on tree shifts and introduces the key concept of pattern distribution, which is central to our exposition. Section 3 relates the topological pressure to the intermediate quantity $P^{(\infty)}(d, E)$ and its finite approximation $P^{(k)}(d, E)$ (maximum of (Problem 5)), for which the discussions are summarized in Theorem 3.1. The proofs of Theorems 3.1 and 1.1 are postponed until Secs. 3.4.1 and 3.4.2, respectively. Finally, Sec. 4 contains two examples with their $P^{(\infty)}(d, E)$ plotted for verification purposes (Figs. 1 and 2).

2. Preliminaries

2.1. *Markov tree-shifts and topological pressure*

To introduce the idea of pattern distribution, we should generalize topological entropy and topological pressure in a way that instead of being confined to initial n-subtree $\Delta_n^{(d)}$, we consider projections onto the set

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

Bearing this in mind, we describe the set of *blocks*, in terms of the projection map, as

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

The *weight* of $u \in B_{n:m}(X_E^{\times d})$ on $\Delta_{n:m}^{(d)}$ is then defined as

$$Z_{n:m}[u, E] = \begin{cases} w[u|_{\Xi_n^{(d)}}] \displaystyle\prod_{g\in\Delta_{n:m-1}^{(d)}} \prod_{i=1}^{d} E_{u_{gf_i},u_g} & \text{if } n < m, \\ w[u|_{\Xi_n^{(d)}}] & \text{otherwise,} \end{cases}$$

where notation w is slightly abused to express the weight of a block $u \in B_{n:n}(X_E^{\times d})$:

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

We then can introduce an analog of the partition function (1.5) defined as

$$|Z_{n:m}(X_E^{\times d}, E)| = \sum_{u \in B_{n:m}(X_E^{\times d})} Z_{n:m}[u, E],$$

defined on $\Delta_{n:m}^{(d)}$ so that $|Z_n(X_E^{\times d}, E)| = |Z_{0:n}(X_E^{\times d}, E)|$. It is noteworthy that we suppress the notation $w \in \mathbb{R}$ in all of the notations above since the limit (1.6) is independent of the vector. For convenience, we usually further omit the dependency on $X_E^{\times d}$, E and d for the above notations as long as their meanings are clear from the context.

Our conventions and notation for matrices and vectors are as follows. Recall that $\Gamma_{\mathcal{A}}$ is the set of all probability vectors indexed by $\mathcal{A}$ and $\Upsilon_{\mathcal{A}}$ is the set of left stochastic matrices acting on $\Gamma_{\mathcal{A}}$. The notation for vectors in $\Gamma_{\mathcal{A}}$ will be in a lowercase sans serif font, such as p, q, while the matrices in $\Upsilon_{\mathcal{A}}$ will be denoted in uppercase, such as P, Q. A left stochastic matrix P has each of its column summing to 1, namely, $\sum_{a \in \mathcal{A}} \mathsf{P}_{a,b} = 1$ for all $b \in \mathcal{A}$. The transpose of a matrix M is denoted by M^T, and for a vector v, its transpose is written as v^T. The product of two square matrices M and N (or the product of a matrix M and a vector v) of the same dimension is denoted by MN (or Mv). For conciseness, we denote $\prod_{i=1}^k M_i = M_1 M_2 \cdots M_k$ for any k square matrices of the same dimension. Finally, the standard unit vector associated with the symbol a is denoted by e_a, and both the all-one vector and the all-one matrix are denoted by $\mathbb{1}$.

The spaces $\Gamma_{\mathcal{A}}$ and $\Upsilon_{\mathcal{A}}$ in this paper are implicitly endowed with the variational distance defined as follows.

Definition 2.1. Let $\mathsf{p}, \mathsf{q} \in \Gamma_{\mathcal{A}}$ and $\mathsf{P}, \mathsf{Q} \in \Upsilon_{\mathcal{A}}$. The variational distance between two vectors is given by

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

and the variational distance between two matrices is defined as

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

Additionally, for any pairs (p, P) and (q, Q), their variational distance is defined as

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

For convenience, the product spaces $\Gamma_{\mathcal{A}}^n$, $\Upsilon_{\mathcal{A}}^n$ and $\Gamma_{\mathcal{A}}^n \times \Upsilon_{\mathcal{A}}^n$ are assumed throughout to be equipped with the maximum metrics, which, with a slight abuse of notation, are also denoted by d_v, d_V and $\mathsf{d}_{v,V}$, respectively.

Proposition 2.1. *Suppose* $\mathsf{p}, \mathsf{q} \in \Gamma_{\mathcal{A}}$ *and* $\mathsf{P}, \mathsf{Q}, \mathsf{R} \in \Upsilon_{\mathcal{A}}$. *Then, the following hold*:

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

Proof. It is clear that

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

that

$$\mathsf{d}_v(\mathsf{Pp}, \mathsf{Pq}) = \frac{1}{2} \sum_{a \in \mathcal{A}} \left| \sum_{b \in \mathcal{A}} \mathsf{P}_{a,b} \mathsf{p}_b - \mathsf{P}_{a,b} \mathsf{q}_b \right| = \frac{1}{2} \sum_{b \in \mathcal{A}} |\mathsf{p}_b - \mathsf{q}_b| \sum_{a \in \mathcal{A}} \mathsf{P}_{a,b} \leq \mathsf{d}_v(\mathsf{p}, \mathsf{q})$$

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

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

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

2.2. *Pattern distribution*

The primary goal of this subsection is to introduce the distribution vectors and the transition matrices, which, in essence, capture the frequency of symbol occurrence in a block, and to prepare necessary combinatorial estimates for later discussions.

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

$$\tau_n(t) = \left(\frac{\sum_{g \in \Xi_n^{(d)}} \chi_a(t_g)}{|\Xi_n^{(d)}|} \right)_{a \in \mathcal{A}} \in \Gamma_{\mathcal{A}} \tag{2.1}$$

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

$$D_n(X_E^{\times d}) = \{\tau_n(t) : t \in X_E^{\times d}\}. \tag{2.2}$$

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

$$\eta_n(t)_{a,b} = \begin{cases} \left(\dfrac{\sum_{g \in \Xi_n^{(d)}, t_g = b} \sum_{h \in \sigma(g)} \chi_a(t_h)}{\sum_{g \in \Xi_n^{(d)}, t_g = b} |\sigma(g)|} \right) & \text{if} \displaystyle\sum_{g \in \Xi_n^{(d)}, t_g = b} |\sigma(g)| > 0, \\ \dfrac{E_{a,b}}{\sum_{a \in \mathcal{A}} E_{c,b}} & \text{otherwise} \end{cases} \tag{2.3}$$

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

$$S_n(X_E^{\times d}) = \{\eta_n(t) : t \in X_E^{\times d}\}. \tag{2.4}$$

For the sake of convenience, we similarly generalize the notations:

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

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

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

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

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

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

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

and the partition functions with given distribution vectors and transition matrices are defined as

$$|Z_{n:m}(X_E^{\times d}, E; \mathsf{q})| = \sum_{u \in B_{n:m}(X_E^{\times d}; \mathsf{q})} Z_{n:m}[u, E],$$

$$|Z_{n:m}(X_E^{\times d}, E; \mathsf{q}, \mathsf{Q})| = \sum_{u \in B_{n:m}(X_E^{\times d}; \mathsf{q}, \mathsf{Q})} Z_{n:m}[u, E].$$

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

Proposition 2.2. *For any* $n \leq m$,

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

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

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

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

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

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

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

The rest of the argument follows a similar line of reasoning.

Finally, the limit of the ratio holds true as a consequence of the exponential growth of $|\Xi_n|$. □

3. Topological Pressure and Pattern Distribution

This section explains the relationship between topological pressure and pattern distribution, laying the groundwork for the proof of Theorem 1.1. Before diving into the exposition, we first outline the general idea. The core concept behind distribution vectors and transition matrices is simple: any two blocks $u, v \in B_{n:m}(\mathsf{q}, \mathsf{Q})$ have the same weight. Thus, by keeping track of $|B_{n:m}(\mathsf{q}, \mathsf{Q})|$, we gain a clearer understanding of the partition function. The road map of our approach is as below.

Step 1. Applying Lemma 2.2, we establish

$$\lim_{n\to\infty} \frac{\log |Z_n(X_E^{\times d}, E)|}{|\Delta_n^{(d)}|} = \lim_{k\to\infty} \lim_{n\to\infty} \max_{(\mathsf{q},\mathsf{Q})\in W_{n:n+k}} \frac{\log |Z_{n:n+k}(X_E^{\times d}, E; \mathsf{q}, \mathsf{Q})|}{|\Delta_{n:n+k}^{(d)}|}.$$

Step 2. Instead of solving the maximization problem on the right-hand side, we relate it to a more tractable optimization problem (Problem 5), whose maximum $P^{(k)}(d, E)$ satisfies

$$\lim_{k\to\infty} \lim_{n\to\infty} \max_{(\mathsf{q},\mathsf{Q})\in W_{n:n+k}} \frac{\log |Z_{n:n+k}(X_E^{\times d}, E; \mathsf{q}, \mathsf{Q})|}{|\Delta_{n:n+k}^{(d)}|} = \lim_{k\to\infty} P^{(k)}(d, E).$$

Step 3. Determine $P^{(k)}(d, E)$ for $d \in (1, \infty)$.

Step 4. Prove that $P^{(\infty)}(d, E) = \lim_{k\to\infty} P^{(k)}(d, E)$ and show that the function, defined for $d \in (1, \infty)$, is continuous and increasing.

Our argument is deeply inspired by the combinatorial proof of Cramér's theorem in large deviation theory (see, for example, [10, Chap. 2]), allowing us to apply the well-established techniques therein.

3.1. *Step* 1

This subsection studies the partition function as a function of distribution vectors and transition matrices. Noting that $B_{n:m}$ is a disjoint union of $B_{n:m}(\mathsf{q},\mathsf{Q})$ ($(\mathsf{q},\mathsf{Q}) \in W_{n:m}$), we may bound the partition function by

$$\max_{\mathsf{q}\in D_{n:m}} |Z_{n:m}(\mathsf{q},\mathsf{Q})| \le |Z_{n:m}| \le |D_{n:m}| \cdot \max_{\mathsf{q}\in D_{n:m}} |Z_{n:m}(\mathsf{q},\mathsf{Q})|. \tag{3.1}$$

Therefore, for any unbounded increasing sequence $(\Delta_{n_i:m_i})_{i\in\mathbb{N}}$, i.e. a sequence satisfying that $|\Delta_{n_i:m_i}| \le |\Delta_{n_{i+1}:m_{i+1}}|$ and that $\lim_{i\to\infty} |\Delta_{n_i:m_i}| = \infty$, it follows from Proposition 2.2 that

$$\liminf_{i\to\infty} \frac{\log|Z_{n_i:m_i}|}{|\Delta_{n_i:m_i}|} = \liminf_{i\to\infty} \max_{\mathsf{q}\in D_{n_i:m_i}} \frac{\log|Z_{n_i:m_i}(\mathsf{q},\mathsf{Q})|}{|\Delta_{n_i:m_i}|}, \tag{3.2}$$

$$\limsup_{i\to\infty} \frac{\log|Z_{n_i:m_i}|}{|\Delta_{n_i:m_i}|} = \limsup_{i\to\infty} \max_{\mathsf{q}\in D_{n_i:m_i}} \frac{\log|Z_{n_i:m_i}(\mathsf{q},\mathsf{Q})|}{|\Delta_{n_i:m_i}|}. \tag{3.3}$$

Following from a similar argument to [13, Theorem 2.1], the upper and lower limits are known to coincide in the following two essential cases.

Lemma 3.1. *Let E satisfy assumption* (A) *and $(\Delta_{n_i:m_i})_{i\in\mathbb{N}}$ be an unbounded increasing sequence. Then,*

- *If $m_i - n_i = k$ for all i, then* (3.2) *equals* (3.3).
- *If $\lim_{i\to\infty} m_i - n_i = \infty$, then both* (3.2) *and* (3.3) *coincide with $\mathbf{P}$.*

Proof. For short, we denote $a_i = \log|Z_{n_i:m_i}|/|\Delta_{n_i:m_i}|$.

To prove the first claim, we may assume without loss of generality that $m_i = n_i + k = i + k$, since the remaining cases follow naturally from this special case since $n_i \to \infty$. Under this assumption, it is not hard to see from the definition that

$$|Z_{i+1:i+k+1}| \le |Z_{i:i+k}|^d \quad \text{and} \quad |\Delta_{i+1:i+k+1}| = d\cdot|\Delta_{i:i+k}| \quad \text{for all } i. \tag{3.4}$$

Hence, a_i is decreasing and in particular, convergent.

For the second claim, it is readily checked, following from (3.4), that

$$\limsup_{i\to\infty} a_i \le \limsup_{i\to\infty} \frac{\log|Z_{m_i-n_i}|}{|\Delta_{m_i-n_i}|} = \mathbf{P}$$

and thus it remains to prove $\liminf_{i\to\infty} a_i = \mathbf{P}$. To this end, we show a reversed inequality of (3.4):

$$\max_{(\mathsf{q},\mathsf{Q})\in W_{i:i+k}} |Z_{i:i+k}(\mathsf{q},\mathsf{Q})| \ge \max_{(\mathsf{q},\mathsf{Q})\in W_{0:k}} |Z_{0:k}(\mathsf{q},\mathsf{Q})|^{d^i} \quad \text{for all } i,$$

immediately yielding

$$\liminf_{i\to\infty} a_i \ge \liminf_{i\to\infty} \frac{\log|Z_{m_i-n_i}(\mathsf{q},\mathsf{Q})|}{|\Delta_{m_i-n_i}|} = \mathbf{P}.$$

To prove inequality, assume $(\mathsf{q}, \mathsf{Q}) \in W_{0:k}$ is a maximizer of the right-hand side and $u^{(g)} \in W_{0:k}(\mathsf{q}, \mathsf{Q})$, $g \in \Xi_i$. Under the circumstances, $\mathsf{q}^{(0)} = \mathsf{e}_a$ for some $a \in \mathcal{A}$, implying that $u_\epsilon^{(j)} = a$ for all $g \in \Xi_i$. Due to assumption (A), there exist $b \in \mathcal{A}^{i+1}$ with $E_{b_{\ell+1},b_\ell} > 0$ for $0 \leq \ell < i$ and $b_i = a$, and thus a block $v \in W_{i:i+k}(\mathsf{q}, \mathsf{Q})$ defined as

$$v_g = \begin{cases} b_\ell & \text{if } g \in \Xi_\ell \text{ for } \ell \leq i, \\ u_{g''}^{(g')} & \text{if } g = g'g'',\ g' \in \Xi_i. \end{cases}$$

As a result,

$$\begin{aligned} |Z_{i:i+k}(\mathsf{q}, \mathsf{Q})| &= \sum_{v \in W_{i:i+k}(\mathsf{q},\mathsf{Q})} Z[v, E] \geq \prod_{g \in \Xi_i} \sum_{u^{(g)} \in W_{0:k}(\mathsf{q},\mathsf{Q})} Z[u^{(g)}, E] \\ &= |Z_{0:k}(\mathsf{q}, \mathsf{Q})|^{d^i} \end{aligned}$$

proving the desired result. □

An immediate consequence of the lemma above is

$$\mathbf{P} = \lim_{k\to\infty} \lim_{n\to\infty} \max_{\mathsf{q} \in W_{n:n+k}} \frac{\log |Z_{n:n+k}(\mathsf{q})|}{|\Delta_{n:n+k}|}, \tag{3.5}$$

which leads to the investigation of the following optimization problem:

$$\begin{cases} \text{maximize} & \dfrac{\log |Z_{n:n+k}(\mathsf{q}, \mathsf{Q})|}{|\Delta_{n:n+k}|}, \\ \text{subject to} & (\mathsf{q}, \mathsf{Q}) \in W_{n:n+k}. \end{cases} \tag{Problem 2}$$

3.2. *Step 2*

Solving the original optimization problem (Problem 2) is challenging due to its combinatorial nature. Thus, we take the following steps to transform the problem into a regular optimization problem without changing the limiting behavior of the maximum:

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

To achieve it, we first approximate the objective function using Stirling's approximation (Problem 2). Before we demonstrate this, we first define the "Kullback–Leibler divergence" vector. Let $\mathsf{Q} \in \Upsilon_{\mathcal{A}}$ and M be a non-negative matrix of the same dimension and $\mathsf{Q}_{a,b} = 0$ if $M_{a,b} = 0$. We define such a vector as

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

where $0 \log \frac{0}{0}$ is interpreted as 0. We should stress that M is not necessarily a stochastic matrix, and so $D_{\mathrm{KL}}(\mathsf{Q}\|M)$ need not be non-negative.

Proposition 3.1. *Let* $(\mathsf{q}, \mathsf{Q}) \in \Omega_{n:m}$. *Then,*

$$\frac{\log |Z_{n:m}(\mathsf{q}, \mathsf{Q})|}{|\Delta_{n:m}|} = \frac{\log |Z_{n:n}(\mathsf{q}^{(0)})|}{|\Delta_{n:m}|} - \sum_{j=0}^{m-n-1} \frac{|\Xi_{n+j+1}|}{|\Delta_{n:m_i}|} D_{\mathrm{KL}}(\mathsf{Q}^{(j)} \| E)^T \mathsf{q}^{(j)} + O\left(\frac{\log |\Delta_{n:m}|}{|\Delta_{n:m}|}\right).$$

Proof. Recall that Stirling's approximation gives

$$\ln(n!) = n \log n - n + O(\ln n).$$

A combinatorial argument shows that

$$|Z_{n:m}(\mathsf{q}, \mathsf{Q})| = |Z_{n:n}(\mathsf{q}^{(0)})| \prod_{j=0}^{m-n-1} \prod_{b \in \mathcal{A}} \left[\binom{|\Xi_{n+j+1}| \cdot \mathsf{q}_b^{(j)}}{|\Xi_{n+j+1}| \cdot \mathsf{q}_b^{(j)} \cdot \mathsf{Q}_{a,b}^{(j)}} \cdot \prod_{a \in \mathcal{A}} E_{a,b}^{|\Xi_{n+j+1}| \cdot \mathsf{q}_b^{(j)} \cdot \mathsf{Q}_{a,b}^{(j)}} \right],$$

where $\binom{a}{b_1, b_2, \ldots, b_k} = \frac{a!}{b_1! b_2! \cdots b_k!}$ is the multinomial coefficient, which according to Stirling's approximation can be expressed as

$$\log \binom{a}{b_1, b_2, \ldots, b_k} = -a \sum_{i=1}^{k} \frac{b_i}{a} \log \frac{b_i}{a} + O(\log a).$$

Essentially, the multinomial coefficient in the expression of $|Z_{n:m}(\mathsf{q}, \mathsf{Q})|$ simply counts the blocks which have $|\Xi_{n+j+1}| \cdot \mathsf{q}_b^{(j)}$ type b particles on level $n+j$ that produce $|\Xi_{n+j+1}| \cdot \mathsf{q}_b^{(j)} \cdot \mathsf{Q}_{a,b}^{(j)}$ type a particles on level $n+j+1$. Therefore,

$$\begin{aligned}
\frac{\log |Z_{n:m}(\mathsf{q}, \mathsf{Q})|}{|\Delta_{n:m}|} &= \frac{\log |Z_n(\mathsf{q}^{(0)})|}{|\Delta_{n:m}|} + \sum_{j=0}^{m-n-1} \frac{|\Xi_{n+j+1}|}{|\Delta_{n:m}|} \\
&\quad \times \sum_{b \in \mathcal{A}} \left[- \sum_{a \in \mathcal{A}} \mathsf{q}_b^{(j)} \mathsf{Q}_{a,b}^{(j)} \log \left(\frac{\mathsf{Q}_{a,b}^{(j)}}{E_{a,b}} \right) + O\left(\frac{\log |\Xi_{n+j+1}|}{|\Xi_{n+j+1}|} \right) \right] \\
&= \frac{\log |Z_{n_i}(\mathsf{q}^{(0)})|}{|\Delta_{n:m}|} - \sum_{j=0}^{m-n-1} \frac{|\Xi_{n+j+1}|}{|\Delta_{n:m}|} D_{\mathrm{KL}}(\mathsf{Q}^{(j)} || E)^T \mathsf{q}^{(j)} \\
&\quad + O\left(\frac{\log |\Delta_{n:m}|}{|\Delta_{n:m}|} \right),
\end{aligned}$$

where the last equality follows from the concavity of $x \mapsto -x \log x$. □

As a corollary of the proposition above, we transform (Problem 2) into the following:

$$\begin{cases}\text{maximize} & \dfrac{\log|Z_{n:n}(\mathsf{q}^{(0)})|}{|\Delta_{n:n+k}|} - \displaystyle\sum_{j=0}^{k-1} \frac{|\Xi_{n+j+1}|}{|\Delta_{n:n+k}|} D_{\mathrm{KL}}(\mathsf{Q}^{(j)}||E)^T \mathsf{q}^{(j)}, \\ \text{subject to} & (\mathsf{q},\mathsf{Q}) \in W_{n:n+k}. \end{cases} \tag{Problem 3}$$

The trickiest part of (Problem 3) lies in the first term involving $|Z_{n:n}(\mathsf{q}^{(0)})|$, for which, to the authors' knowledge, no good estimate is available. Nevertheless, our main interest lies in the case $k \to \infty$, where the term converges uniformly (with respect to q) to zero and hence could be safely dropped. Thus, we shall continue our discussion by rephrasing the problem, by writing $\mathsf{p} = \overleftarrow{\mathsf{q}}$ and $\mathsf{P} = \overleftarrow{\mathsf{Q}}$, as

$$\begin{cases}\text{maximize} & -\displaystyle\sum_{j=0}^{k-1} \frac{d-1}{d^{j+1} - d^{j-k}} D_{\mathrm{KL}}(\mathsf{P}^{(j)}||E)^T \mathsf{p}^{(j+1)}, \\ \text{subject to} & (\mathsf{p},\mathsf{P}) \in \overleftarrow{W_{n:n+k}} := \{(\overleftarrow{\mathsf{q}}, \overleftarrow{\mathsf{Q}}) : (\mathsf{q},\mathsf{Q}) \in W_{n:n+k}\}. \end{cases} \tag{Problem 4}$$

We remark that the bottom-up convention of (Problem 4) turns out to be more convenient for our later discussion.

In what follows, we will further show that, as $n \to \infty$, the maximum of (Problem 4) converges to that with the feasible domain replaced by

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

which is a set containing $\overleftarrow{W_{n:n+k}}$ for all n. It is noteworthy that Ω_k is a compact set on which the objective function is continuous, guaranteeing a maximizer for the problem defined on $\overleftarrow{W_{n:n+k}}$. Given these, our aforementioned aim could be reached if we were able to show that $\overleftarrow{W_{k:k+n}}$, as $n \to \infty$, is asymptotically dense in Ω_k. Unfortunately, it is rarely the case, since for $(\mathsf{q},\mathsf{Q}) \in W_{k:k+n}$, $\mathsf{q}_b^{(i)} = 0$ forces $\mathsf{Q}_{a,b}^{(i)} = \frac{E_{a,b}}{\sum_{c\in\mathcal{A}} E_{c,b}}$, whereas no such constraint is seen in Ω_k. However, it is obvious that the maximum is the same with or without the constraint, and thus we may further impose the following restrictions.

Definition 3.1. Let E be the restriction matrix. A pair $(\mathsf{q},\mathsf{Q}) \in \Gamma_{\mathcal{A}} \times \Upsilon_{\mathcal{A}}$ is called *typical* if the following hold:

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

Furthermore, for any fixed $\mathsf{P} \in \Upsilon_{\mathcal{A}}^{k+1}$, the objective function is just an affine function of $\mathsf{p}^{(0)} \in \Gamma_{\mathcal{A}}$. Therefore, we can assume such maximizer, denoted by $(\mathsf{p},\mathsf{P}) \in \Omega_k$, has its last vector $\mathsf{p}^{(k)}$ taking the form of $\mathsf{p}^{(k)} = \mathsf{e}_a$ ($a \in \mathcal{A}$) and lies in

$$\Omega'_k = \{(\mathsf{p},\mathsf{P}) \in (\Gamma^k_{\mathcal{A}} \times \{\mathsf{e}_a\}_{a\in\mathcal{A}}) \times \Upsilon^k_{\mathcal{A}} : \mathsf{p}^{(j)} = \mathsf{P}^{(j)}\mathsf{p}^{(j+1)}, \\ (\mathsf{P}^{(j)}, \mathsf{p}^{(j+1)}) \text{ is typical}, 0 \le j < k\}.$$

We then show in the following lemma that the set $\overleftarrow{W_{n:n+k}}$ is asymptotically dense in Ω'_k as n tends to infinity.

Lemma 3.2. *Let* $(\mathsf{q}',\mathsf{Q}) \in D_n \times \Upsilon_{\mathcal{A}}$ *be typical. Then, there exists* $((\mathsf{q}',\mathsf{Q}'\mathsf{q}'),\mathsf{Q}') \in W_{n:n+1}$ *such that*

$$|\mathsf{Q}_{a,b} - \mathsf{Q}'_{a,b}| \le \frac{1}{|\Xi_{n+1}|\mathsf{q}'_b} \quad \textit{for all } a,b \in \mathcal{A}.$$

Moreover, Q' *can be chosen so as to satisfy that* $\mathsf{Q}_{a,b}{}' = 0$ *if* $\mathsf{Q}_{a,b} = 0$.

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

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

It is not hard to see that these criteria, together with the additional properties stated in the lemma, can always be satisfied simultaneously by some $\mathsf{Q}' \in \Upsilon_{\mathcal{A}}$. □

Proposition 3.2. $\lim_{n\to\infty} \sup_{(\mathsf{p},\mathsf{P})\in\Omega'_k} \mathsf{d}_{v,V}(\overleftarrow{W_{n:n+k}}, (\mathsf{p},\mathsf{P})) = 0.$

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

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

We take n sufficiently large such that

$$\delta := \max_{i,a:\mathsf{q}^{(i)}_a>0} (|\Xi_{n+i+1}|\mathsf{q}^{(i)}_a)^{-1} \quad \text{satisfying } \frac{k|\mathcal{A}|\delta}{\zeta} < \min\left\{\frac{1}{2}, \epsilon\right\}.$$

We then construct a pair $(\mathsf{q}',\mathsf{Q}') \in W_{n:n+k}$ such that $\mathsf{d}_{v,V}((\overleftarrow{\mathsf{q}'}, \overleftarrow{\mathsf{Q}'}), (\mathsf{q},\mathsf{Q})) < \epsilon$ through the following process. To begin with, take $\mathsf{q}^{(0)'} = \mathsf{q}^{(0)} \in D_n$ and choose $\mathsf{Q}^{(0)'}$ as in Lemma 3.2 to define $\mathsf{q}^{(1)'} = \mathsf{Q}^{(0)'}\mathsf{q}^{(0)'} \in D_{n+1}$. Now suppose $((\mathsf{q}^{(i)'})_{0\le i\le j}, (\mathsf{Q}^{(i)'})_{0\le i\le j-1})$ is found and $\mathsf{q}^{(j)'} = 0$ if and only if $\mathsf{q}^{(j)} = 0$, then $(\mathsf{q}^{(j)'}, \mathsf{Q}^{(j)}) \in D_{n+j} \times \Upsilon_{\mathcal{A}}$ is typical, so we again choose $\mathsf{Q}^{(j)'}$ as in Lemma 3.2 and set $\mathsf{q}^{(j+1)'} = \mathsf{Q}^{(j)'}\mathsf{q}^{(j)'} \in D_{n+j+1}$. Otherwise, the process terminates. We claim that the process terminates only after step k. To see this, we make use of the following properties, which are proved in the next paragraph, for the well-defined

$((\mathsf{q}^{(i)'})_{0\le i\le j}, (\mathsf{Q}^{(i)'})_{0\le i\le j-1})$: for all $0 \le i \le j-1$:

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

If the listed properties hold and if we suppose toward a contradiction that $j < k$, then we see that ${\mathsf{q}^{(j)}}'_a = (\mathsf{Q}^{(j-1)'}\mathsf{q}^{(j-1)'})_a = 0$ if and only if $\mathsf{q}_a^{(j)} = (\mathsf{Q}^{(j-1)}\mathsf{q}^{(j-1)})_a = 0$, and thus the process should proceed, which is a contradiction. Now that $((\mathsf{q}^{(j)'}, \mathsf{Q}^{(j)'}\mathsf{q}^{(j)'}), \mathsf{Q}^{(j)'}) \in D_{n+j:n+j+1}$ for all $0 \le j < k$, which implies $(\mathsf{q}, \mathsf{Q}) \in W_{n:n+k}$ and $\mathtt{d}_{v,V}(\overleftarrow{W_{n:n+k}}, (\mathsf{p}, \mathsf{P})) < k|\mathcal{A}|\delta < \epsilon$ as desired.

We prove the aforementioned claim by induction on i. When $i = 0$, since $\mathsf{q}^{(0)} = \mathsf{q}^{(0)'}$, property (a) is automatic. For (b), if $\mathsf{q}_b^{(0)} = 0$, then

$$\mathsf{Q}_{a,b}^{(0)'} = \mathsf{Q}_{a,b}^{(0)} = E_{a,b} \cdot (\sum_{a\in\mathcal{A}} E_{c,b})^{-1}.$$

If $\mathsf{q}_b^{(0)} > 0$, then by Lemma 3.2,

$$\mathsf{Q}_{a,b}^{(0)'} \ge \mathsf{Q}_{a,b}^{(0)} - \tfrac{1}{|\Xi_{n+1}|\mathsf{q}'_b} \ge \zeta - \delta > 0.$$

In addition, we may apply Lemma 3.2 again to deduce $\mathtt{d}_V(\mathsf{Q}^{(0)'}, \mathsf{Q}^{(0)}) \le |\mathcal{A}|\delta$.

For the induction step, we assume the hypotheses hold for $i-1 < j-1$. To prove (a) for the index i, note that $\mathsf{q}^{(i)'} = \mathsf{Q}^{(i-1)'}\mathsf{q}^{(i-1)'} = 0$ if and only if $\mathsf{q}^{(i-1)} = \mathsf{Q}^{(i-1)}\mathsf{q}^{(i-1)} = 0$ is straightforward from the induction hypotheses, and a simple calculation shows that

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

For (b), we note that if $\mathsf{q}_a^{(i)'} > 0$,

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

We apply Lemma 3.2 for the last time to obtain $\mathtt{d}_V(\mathsf{Q}^{(i)'}, \mathsf{Q}^{(i)}) \le |\mathcal{A}|\delta$. Furthermore, $\mathsf{Q}_{a,b}^{(i)'} = \mathsf{Q}_{a,b}^{(i)}$ if $\mathsf{q}_{a,b}^{(i)} = 0$, and if $\mathsf{q}_{a,b}^{(i)} \ne 0$, then

$$\mathsf{Q}_{a,b}^{(i)'} \ge \mathsf{Q}_{a,b}^{(i)} - \mathtt{d}_V(\mathsf{Q}^{(i)'}, \mathsf{Q}^{(i)}) \ge \zeta - |\mathcal{A}|\delta > 0.$$

Our proof by induction is therefore completed. □

As the last step of this subsection, we would like to once again reformulate the problem (Problem 4) to eliminate its dependence on k in the objective function. We note that the function $D_{\mathrm{KL}}(\mathsf{P}^{(j)}\|E)^T\mathsf{p}^{(j+1)}$ in (Problem 4) is uniformly bounded for all j, k and $\mathsf{P}^{(j)}$, $\mathsf{p}^{(j+1)}$. Furthermore, its associated coefficient $(d-1)/(d^{j+1}-d^{j-k})$ admits a limit $(d-1)/d^{j+1}$ and the ratio between the two coefficients is uniform for all j:

$$\frac{d-1}{d^{j+1}-d^{j-k}}\bigg/\frac{d-1}{d^{j+1}}=\frac{1}{1-d^{-k-1}}\to 1.$$

As a consequence, we arrive at the following optimization problem:

$$\begin{cases}\text{maximize} \quad -\displaystyle\sum_{j=0}^{k-1}\frac{d-1}{d^{j+1}}D_{\mathrm{KL}}(\mathsf{P}^{(j)}\|E)^T\mathsf{p}^{(j+1)}, \\ \\ \text{subject to} \quad (\mathsf{p},\mathsf{P})\in\Omega_k.\end{cases} \qquad \text{(Problem 5)}$$

whose maximum, denoted $P^{(k)}(d,E)$, converges, due to Proposition 3.2 and the arguments above, to the topological pressure:

$$\mathbf{P}(X_E^{\times d},E)=\lim_{k\to\infty}P^{(k)}(d,E).$$

It is noteworthy that the optimization problem above is closely related to (Problem 1), whose the maximum $P^{(\infty)}(d,E)$ is well-defined since its feasible domain is compact with respect to the metric $\mathsf{d}_{v,V}^{\infty}$ defined as

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

and the objective function is continuous on the feasible domain. Our goal is thus to demonstrate that $P^{(k)}(d,E)\to P^{(\infty)}(d,E)$, which, as a function of $d\in(1,\infty)$, can be shown to be continuous and increasing using uniform convergence of $P^{(k)}(d,E)$ on compact subintervals. However, the proof for this requires knowledge about the solutions of (Problem 5) and is on hold for the time being.

3.3. *Step 3*

As pointed out in the last subsection, the objective function of (Problem 5) is affine in $\mathsf{p}^{(k)}$, and thus $\mathsf{p}=\mathsf{e}_a$ for some $a\in\mathcal{A}$ can be assumed since $\Gamma_{\mathcal{A}}$ is the convex hull of $\{\mathsf{e}_a\}_{a\in\mathcal{A}}$. The problem simplifies to the following:

$$\begin{cases}\text{maximize} \quad -\displaystyle\sum_{j=0}^{k-1}\frac{d-1}{d^{j+1}}D_{\mathrm{KL}}(\mathsf{P}^{(j)}\|E)^T\prod_{\ell=j+1}^{k-1}\mathsf{P}^{(\ell)}\mathsf{p}, \\ \\ \text{subject to} \quad \mathsf{p}=\mathsf{e}_a, \quad \mathsf{P}\in\Upsilon_{\mathcal{A}}^k, \\ \qquad\qquad\quad \mathsf{P}_{a,b}^{(j)}=0 \quad \text{if } E_{a,b}=0,\ 0\le j<k.\end{cases} \qquad (3.6)$$

We denote by $F_k(\mathsf{p},\mathsf{P};1/d,E)$ the objective function of (3.6) and by $F_\infty(\mathsf{p},\mathsf{P};1/d,E)$ the objective function of (Problem 1), and we find a maximizer as follows.

For conciseness, we suppress E and $1/d$ if doing so should not occasion any confusion.

Proposition 3.3. *The optimization problem* (3.6) *admits an optimal transition* $\mathsf{P}^{(i)}$ $(i = 0, \dots, k-1)$ *independent of* p *that is defined as*

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

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

$$\lambda^{(i)}_a = \frac{d-1}{d^i} \log \sum_{b:E_{b,a}>0} e^{\frac{d^i}{d-1}\lambda^{(i-1)}_b} E_{b,a} \quad \text{for } i = 1, \dots, k.$$

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

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

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

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

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

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

$$\begin{aligned} \lambda^{(i)} &:= -\frac{d-1}{d^i} D_{\mathrm{KL}}(\mathsf{P}^{(i-1)} \| E) + {\mathsf{P}^{(i-1)}}^T \lambda^{(i-1)} \\ &= -\sum_{j=0}^{i-1} \frac{d-1}{d^{j+1}} \left(\prod_{\ell=j+1}^{i-1} \mathsf{P}^{(\ell)} \right)^T D_{\mathrm{KL}}(\mathsf{P}^{(j)} \| E) \end{aligned} \tag{3.7}$$

and express $F_{k;\mathsf{Q}^{(i+1)},\dots,\mathsf{Q}^{(k-1)}}$ as

$$F_{k;\mathsf{Q}^{(i+1)},\dots,\mathsf{Q}^{(k-1)}}(\mathsf{Z})$$
$$= -\sum_{j=0}^{i-1} \frac{d-1}{d^{j+1}} D_{\mathrm{KL}}(\mathsf{P}^{(j)} \| E)^T \left(\prod_{\ell=j+1}^{i-1} \mathsf{P}^{(\ell)} \right) Z \left(\prod_{\ell=i+1}^{k-1} \mathsf{Q}^{(\ell)} \right) \mathsf{p}$$

$$
\begin{aligned}
&-\frac{d-1}{d^{i+1}}D_{\mathrm{KL}}(\mathsf{Z}\|E)^T\prod_{\ell=i+1}^{k-1}\mathsf{Q}^{(\ell)}\mathsf{p}-\sum_{j=i+1}^{k-1}\frac{d-1}{d^{j+1}}D_{\mathrm{KL}}(\mathsf{Q}^{(j)}\|E)^T\prod_{\ell=j+1}^{k-1}\mathsf{Q}^{(\ell)}\mathsf{p}\\
&=\left({\lambda^{(i)}}^T\mathsf{Z}-\frac{d-1}{d^{i+1}}D_{\mathrm{KL}}(\mathsf{Z}\|E)^T\right)\prod_{\ell=i+1}^{k-1}\mathsf{Q}^{(\ell)}\mathsf{p}\\
&\quad-\sum_{j=i+1}^{k-1}\frac{d-1}{d^{j+1}}D_{\mathrm{KL}}(\mathsf{Q}^{(j)}\|E)^T\prod_{\ell=j+1}^{k-1}\mathsf{Q}^{(\ell)}\mathsf{p},
\end{aligned}
\tag{3.8}
$$

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

$$
\mathsf{P}^{(i)}_{a,b}=\begin{cases}\dfrac{e^{\frac{d^{i+1}}{d-1}\lambda^{(i)}_a}E_{a,b}}{\sum_{c:E_{c,b}>0}e^{\frac{d^{i+1}}{d-1}\lambda^{(i)}_c}E_{c,b}} & \text{if } E_{a,b}>0,\\ 0 & \text{otherwise.}\end{cases}
\tag{3.9}
$$

In this manner, we successfully construct the desired sequence and prove the optimality. Moreover, if we plug (3.9) into (3.7), we deduce the recursive relation of λ, and the proof is complete. □

Remark 3.1. In Proposition 3.3, the vector $e^{d^{i+1}/(d-1)\lambda^{(i)}}$ satisfies $e^{d/(d-1)\lambda^{(0)}}=\mathbb{1}$ and the relation

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

which is essentially a generalization of the formula used in [6, Algorithm 1] for the computation of entropy.

3.4. *Step 4*

3.4.1. *Convergence of $P^{(k)}(d,E)$*

We first prove the following theorem regarding the convergence of the optimization problems, preparing us for the proof of Theorem 1.1. The maximizer P^* satisfies the following properties.

Theorem 3.1. *Let $d\in(1,\infty)$, E satisfy assumption* (A),

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

and

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

Then, the maximizer found in Proposition 3.3 *satisfies the following properties*:

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

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

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

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

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

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

is non-negative and increasing.

(c) *There exists* $\mathsf{p}^* \in \Gamma_{\mathcal{A}}^{\mathbb{Z}_+}$ *such that* $(\mathsf{p}^*, \mathsf{P}^*)$ *is a maximizer of* (Problem 1). *In particular,*

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

Proof. (a) We show by induction that the map $\lambda^{(i)} \overset{f_i}{\longmapsto} \lambda^{(i+1)}$ in Proposition 3.3 satisfies that

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

which follows from the monotonicity of f_i: $f_i(\lambda') \geq f_i(\lambda'')$ if $\lambda' \geq \lambda''$. When $i = 0$,

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

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

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

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

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

$$(\xi_\ell)_{\ell \geq 0} = (\xi_\ell^{a,b,i})_{\ell \geq 0} \quad \text{such that } \xi_0 = a,\ \xi_i = b,\ E_{\xi_\ell,\ \xi_{\ell+1}} > 0,\ \forall\, \ell \geq 0. \quad (3.10)$$

Then, we take transition matrices $\mathsf{P}^{(k+\ell)'}$ such that $\mathsf{P}^{(k+\ell)'}_{\xi_\ell, \xi_{\ell+1}} = 1$. Under the circumstances, we have $(\prod_{\ell=0}^{k+1} \mathsf{P}^{(k+j)'})_{a,b} \mathsf{e}_b = \mathsf{e}_a$. Therefore, according to (3.6),

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

This finishes the proof.

(c) We start by noting that the existence of optimizer $(\mathsf{p}', \mathsf{P}')$ is guaranteed by compactness of the feasible domain and continuity of the objective function. Next, by recursively replacing P' by P^* according to Proposition 3.3, the compactness again asserts that there exists a maximizer of the form $(\mathsf{p}^*, \mathsf{P}^*)$ (and thus the first equality).

The third equality and the existence of the limits therein follow from (b). Precisely, for each $a \in \mathcal{A}$ and each $0 \leq i < L$, the following relations:

$$\lim_{n \geq 0} \max_{a \in \mathcal{A}_\infty} \cdot = \sup_{n \geq 0} \max_{a \in \mathcal{A}_\infty} \cdot = \max_{a \in \mathcal{A}_\infty} \sup_{n \geq 0} \cdot = \max_{a \in \mathcal{A}_\infty} \lim_{n \to \infty} \cdot$$

coincide on the sequence

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

To finish the proof of this part, we note the following fact due to assumption (A):

$$\text{For each } a \in \mathcal{A} \text{ and } k \geq |\mathcal{A}|,\ \text{there exists } b \in \mathcal{A}_\infty \text{ such that } (E^k)_{a,b} > 0. \quad (3.11)$$

Immediately this yields that

$$\lim_{n \to \infty} \max_{a \in \mathcal{A}_\infty} F_{Ln+i}(\mathsf{e}_a, \mathsf{P}^{(0:Ln+i-1)^*}) = \lim_{n \to \infty} \max_{a \in \mathcal{A}_\infty} F_{Ln}(\mathsf{e}_a, \mathsf{P}^{(0:Ln-1)^*}) \quad \text{for all } i,$$

justifying the existence of the limits and their coincidence.

To prove the second equality, for each $a \in \mathcal{A}_\infty$ and each $k \in \mathbb{Z}_+$, there exists an extension $(\mathsf{p}(a,k), \mathsf{P}(a,k)) \in \Gamma_{\mathcal{A}}^{\mathbb{Z}_+} \times \Upsilon_{\mathcal{A}}^{\mathbb{Z}_+}$ of $((\prod_{j=\ell}^{k-1} \mathsf{P}^{(\ell)^*} \mathsf{e}_a)_{\ell=0}^k, \mathsf{P}^{(0:k-1)^*})$ (using the sequence $(\xi_\ell^{a,a,L})_\ell$ in (3.10)) such that

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

For the other inequality, we note that due to the affinity of $\mathsf{p}\mapsto F_k(\mathsf{p},\mathsf{P}^{(0:k-1)^*})$,

$$F_\infty(\mathsf{p}^*,\mathsf{P}^*)=\lim_{k\to\infty}F_k(\mathsf{p}^{(0:k)^*},\mathsf{P}^{(0:k-1)^*})\leq\liminf_{k\to\infty}\max_{a\in\mathcal{A}}F_k(\mathsf{e}_a,\mathsf{P}^{(0:k-1)^*}). \quad (3.12)$$

As a result, due to (3.11), we may apply (b) to deduce

$$\begin{aligned}&\lim_{k\to\infty}\max_{a\in\mathcal{A}_\infty}F_k(\mathsf{e}_a,\mathsf{P}^{(0:k-1)^*})\\ &\quad=\lim_{k\to\infty}\max_{a\in\mathcal{A}_\infty}F_k(\mathsf{e}_a,\mathsf{P}^{(0:k-1)^*})-\sum_{\ell=k-|\mathcal{A}|}^{k-1}\frac{d-1}{d^{\ell+1}}\cdot\gamma\\ &\quad\geq\limsup_{k\to\infty}\max_{a\in\mathcal{A}}F_{k-|\mathcal{A}|}(\mathsf{e}_a,\mathsf{P}^{(0:k-|\mathcal{A}|-1)^*})\geq F_\infty(\mathsf{p}^*,\mathsf{P}^*). \quad (3.13)\end{aligned}$$

Combining the inequalities above and the third equality proves the second equality.

Finally, we note that the last equality, together with the existence of the limit, also follows from (3.12) and (3.13), given that $P^{(k)}(d,E)=\max_{a\in\mathcal{A}}F_k(\mathsf{e}_a,\mathsf{P}^{(0:k-1)^*})$. □

3.4.2. *Proof of Theorem* 1.1

Let β and γ be the constants defined in Theorem 3.1, which are independent of d. We show that $P^{(k)}$ converges to $P^{(\infty)}$ in a strong sense.

Proof. (a) For our convenience, we denote $q_k(d)=d^{-k}-d^{-k-1}$ throughout the rest of the discussion. In addition, we denote by $\mathsf{P}^*(d)$ the optimal matrices given in Proposition 3.3 with the associated $\lambda^{(k)}$ denoted by $\lambda^{(k)^*}(d)$.

Our proof of continuity and increasing property relies heavily on the uniform convergence of $\lambda^{(Ln+i)^*}(d)_a$: for every $a\in\mathcal{A}_\infty$ and every $0\leq i\leq L-1$, the function $\Lambda^{(i)^*}(d)_a:=\lim_{n\to\infty}\lambda^{(Ln+i)^*}(d)_a$, which is well-defined for $d\in(1,\infty)$ according to Theorem 3.1(b) is continuous with the convergence uniform on every compact subinterval of $I\subset(1,\infty)$. To begin with, we note $\lambda^{(k)^*}(d)$ is obviously a continuously differentiable function from its definition in Proposition 3.3. To prove the uniform convergence on compact subinterval, we verify that it is an equicontinuous sequence by computing the derivative of $\lambda^{(k)^*}(d)$:

$$\begin{aligned}(\lambda^{(k)^*})'(d)&=(q_k(d)\log(E^Te^{q_k(d)^{-1}\lambda^{(k-1)^*}(d)}))'\\ &=q_k'(d)\log(E^Te^{q_k(d)^{-1}\lambda^{(k-1)^*}(d)})+\frac{q_k(d)}{(E^Te^{q_k(d)^{-1}\lambda^{(k-1)^*}(d)})}\end{aligned}$$

$$\cdot\left[E^T \mathrm{diag}\left(-\frac{q_k'(d)}{q_k(d)^2}\lambda^{(k-1)^*}(d) + \frac{(\lambda^{(k-1)^*})'(d)}{q_k(d)}\right) e^{q_k(d)^{-1}\lambda^{(k-1)^*}(d)}\right]$$

$$= -q_k'(d) D_{\mathrm{KL}}(\mathsf{P}^{(k-1)^*}(d)\|E) + \mathsf{P}^{(k-1)^*}(d)^T (\lambda^{(k-1)^*})'(d)$$

$$= \sum_{j=0}^{k-1}(jd^{-j-1} - (j+1)d^{-j-2}) \left(\prod_{\ell=j+1}^{k-1} \mathsf{P}^{(\ell)^*}(d)\right)^T D_{\mathrm{KL}}(\mathsf{P}^{(j)^*}(d)|E) \tag{3.14}$$

and note that the derivative is uniformly bounded on the interval I:

$$|(\lambda^{(Ln+i)^*})'(d)| \le \sup_{d\in I} \sum_{j=0}^{\infty} |jd^{-j-1} - (j+1)d^{-j-2}|(|\beta| + |\gamma|) < \infty.$$

The equicontinuity then follows from the estimate and the mean value theorem. Now that the functions, when restricted to I, are clearly uniformly bounded, we apply the Arzelà-Ascoli theorem to obtain uniform convergence.

For the continuity of $P^{(\infty)}(d, E)$, we apply our claim to the function

$$\max_{a\in\mathcal{A}_\infty} \lambda^{(Ln+i)^*}(d)_a = \max_{a\in\mathcal{A}_\infty} F_{Ln+i}(\mathsf{e}_a, \mathsf{P}^{(0:Ln+i-1)^*}(d)).$$

to deduce uniform convergence on compact subinterval, which, together with Theorem 3.1(c), yields the desired result.

To prove that $P^{(\infty)}(d, E)$ is increasing in d, we first associate with each $a \in \mathcal{A}_\infty$ an open set I_a such that

$$\Lambda^{(Ln)^*}(d)_a \ge \Lambda^{(Ln)^*}(d)_b, \quad \text{for all } b \in \mathcal{A}$$

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

$$\{x \in U : f_1(x) > f_2(x)\} \cup \{x \in U : f_2(x) > f_1(x)\}$$
$$\cup\, \mathrm{int}(\{x \in U : f_2(x) = f_1(x)\})$$

is dense in U. We then consider an alternative form of (3.14):

$$(\mathsf{e}_a^T \lambda^{(Ln)^*})'(d) = \sum_{j=1}^{Ln-1} \sum_{i=1}^{j} (d^{-j-1} - d^{-j-2}) \mathsf{e}_a^T \left(\prod_{\ell=j+1}^{Ln-1} \mathsf{P}^{(\ell)^*}(d)\right)^T D_{\mathrm{KL}}(\mathsf{P}^{(j)^*}(d)\|E)$$
$$- \sum_{j=0}^{Ln-1} d^{-j-2} \mathsf{e}_a^T \left(\prod_{\ell=j+1}^{Ln-1} \mathsf{P}^{(\ell)^*}(d)\right)^T D_{\mathrm{KL}}(\mathsf{P}^{(j)^*}(d)\|E)$$
$$= -\sum_{i=1}^{Ln-1} d^{-i-1} F_{Ln-i}(\mathsf{e}_a, \mathsf{P}^{(i:Ln-1)^*}(d); d, E) + \frac{d^{-2}}{1-d^{-1}} \lambda_a^{(Ln)^*}(d), \tag{3.15}$$

where the last equality follows from swapping the order of summation. If we further express $d^{-2} \cdot (1-d^{-1})^{-1} = \sum_{i=1}^{\infty} d^{-i-1}$, then the above becomes

$$
\begin{aligned}
(\mathsf{e}_a^T \lambda^{(Ln)^*})'(d) = &- \sum_{i=1}^{Ln-1} d^{-i-1}[F_{Ln-i}(\mathsf{e}_a, \mathsf{P}^{(i:Ln-1)^*}(d); d, E) - \lambda^{(Ln)^*}(d)_a] \\
&+ \sum_{i=Ln}^{\infty} d^{-i-1} \lambda^{(Ln)^*}(d)_a, \\
&\geq \sum_{i=1}^{Ln-1} d^{-i-1}[\lambda^{(Ln)^*}(d)_a - \lambda^{(Ln-i)^*}(d)_a] + \sum_{i=Ln}^{\infty} d^{-i-1} \lambda^{(Ln)^*}(d)_a,
\end{aligned}
$$

where the last inequality follows again from Theorem 3.1(b) and the last function, due to our claim, converges uniformly on every compact subinterval to the following function:

$$
\sum_{i=1}^{L} \frac{d^L - 1}{d^{L+i+1} - d^{L+i}} (\Lambda^{(0)}(d)_a^* - \Lambda^{(L-i)}(d)_a^*),
$$

which is non-negative for all $x \in I_a$. By the mean value theorem, this provides the following bound for $[d', d''] \subset I_a$:

$$
P^{(\infty)}(d'', E) - P^{(\infty)}(d', E) \geq \limsup_{n\to\infty} \inf_{d \in [d', d'']} (\mathsf{e}_a^T \lambda^{(Ln)^*})'(d) \cdot (d'' - d') = 0.
$$

The proof is thus finished by noting that $\bigcup_{a \in \mathcal{A}} I_a$ is open and dense in $(1, \infty)$ and $P^{(\infty)}(d, E)$ is continuous.

(b) It follows essentially from Theorem 3.1(c) that $P^{(\infty)}(d, E) = \lim_{k\to\infty} P^{(k)}(d, E)$, where the limit is known to converge to $\mathbf{P}(X_E^{\times d}, E)$ as discussed in Sec. 3.2.

(c) On the one hand, Theorem 3.1(a) guarantees $\lim_{d\to\infty} P^{(\infty)}(d, E) \leq \beta$. On the other hand, Theorem 3.1(b) together with assumption (A) implies, by taking $k = 1$, $j = 0$ and letting $i \to \infty$, that

$$
\begin{aligned}
\lim_{d\to\infty} P^{(\infty)}(d, E) &= \lim_{d\to\infty} \lim_{i\to\infty} \max_{b \in \mathcal{A}} F_{i+1}(\mathsf{e}_b, \mathsf{P}^{(0:i)^*}; d, E) \\
&\geq \lim_{d\to\infty} \lim_{i\to\infty} \max_{a \in \mathcal{A}} F_1(\mathsf{e}_a, \mathsf{P}^{(0:0)^*}; d, E) + \sum_{\ell=1}^{i} (d^{-k} - d^{-k-1}) \cdot \gamma \\
&= \lim_{d\to\infty} (1 - d^{-1})\beta + d^{-1} \cdot \gamma = \beta.
\end{aligned}
$$

For $\lim_{d\to 1+} P^{(\infty)}(d, E)$, we plug the Parry measure

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

into (Problem 5) to deduce $\lim_{d\to 1+} P^{(\infty)}(d,E) \geq \log\rho(E)$. To prove the other inequality, for every $\epsilon > 0$ we first construct a new interaction matrix

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

Since $P^{(\infty)}(d,E) \leq P^{(\infty)}(d,E')$ for and $d > 1$ according to (Problem 1), it is sufficient to show that $\lim_{d\to 1+} P^{(\infty)}(d,E') \leq \log\rho(E') = \log\rho(E)+\epsilon$. Let v and w be the left and right eigenvector, respectively, associated with $\rho(E')$ such that $w^T v = 1$. According to the spectral decomposition of E', we know $\lim_{n\to\infty} \rho(E')^{-n}(E')^n = vw^T$ is non-negative, and thus we can assume without loss of generality that $v = \rho(E')^{-1}E'v$ is a positive probability vector. We see from Remark 3.1 that for the system defined by E', there exists $C_d = \max_{a\in\mathcal{A}'} v_a^{1-d} \geq 1$ such that

$$\begin{aligned} v^T e^{\frac{d^{k+1}}{d-1}\lambda^{(k)}} &= v^T((E')^T e^{\frac{d^k}{d-1}\lambda^{(k-1)}})^d \\ &\leq C_d \cdot (v^T (E')^T e^{\frac{d^k}{d-1}\lambda^{(k-1)}})^d \\ &= C_d \cdot \rho(E')^d (v^T e^{\frac{d^k}{d-1}\lambda^{(k-1)}})^d. \end{aligned} \tag{3.16}$$

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

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

This provides a uniform bound for $\lim_{d\to 1+} P^{(\infty)}(d,E')$, and the claim is proved by estimating the pressure using (3.16) and letting $d \to 1+$. More specifically,

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

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

4. Experiments

In this section, we present two examples related to Theorem 1.1. Consider the golden-mean tree-shifts $X_G^{\times d}$ with

$$G = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

We claim that the function $P^{(\infty)}(d,E)$ can be approximated by $P^{(k)}(d,E)$ with a crude estimate of error

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

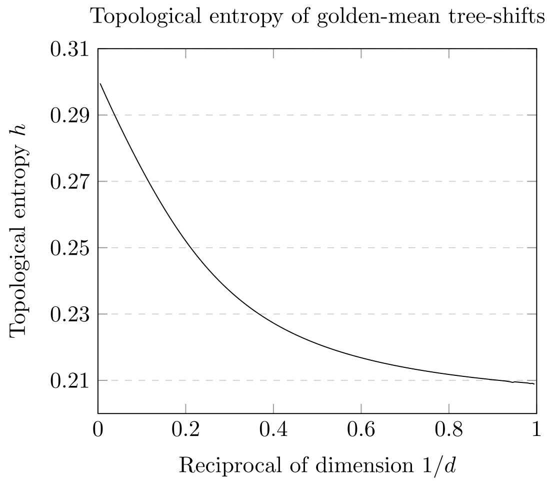

Fig. 1. Topological entropy of golden-mean tree-shifts.

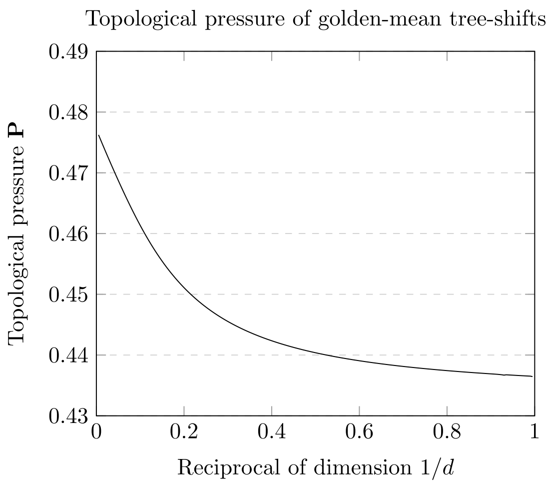

Fig. 2. Topological pressure of golden-mean tree-shifts.

Indeed, the first inequality follows from Theorem 3.1(b), and the second follows, by comparing (Problem 5) and (Problem 1), from the fact that each term in (Problem 1) admits an upper bound $-\frac{d-1}{d^{j+1}} D_{\mathrm{KL}}(\mathsf{P}^{(j)}|E)\mathsf{p}^{(j+1)} \leq \frac{d-1}{d^{j+1}}\beta$. The figure of the topological entropy is given in Fig. 1. For the purpose of demonstration of Theorem 1.1 for general interaction matrices, we include Fig. 2 to show the increasing property of pressure when

$$E = \begin{bmatrix} 2 & 2 \\ 1 & 0 \end{bmatrix}.$$

Both of the figures turn out to be consistent with Theorem 1.1 in the sense that both functions are continuous, $\lim_{d\to\infty} P^{(\infty)}(d, G) = \log_{10} 2 \approx 0.3010$,

$\lim_{d\to\infty} P^{(\infty)}(d,E) = \log_{10} 3 \approx 0.4771$, $\lim_{d\to 1+} P^{(\infty)}(d,G) = \log_{10} \frac{1+\sqrt{5}}{2} \approx 0.2090$ and $\lim_{d\to 1+} P^{(\infty)}(d,E) = \log_{10}(1+\sqrt{3}) \approx 0.4365$.

Acknowledgments

The first author was supported by the Ministry of Science and Technology (Contract No. MOST 111-2115-M-004-005-MY3). We would like to thank the anonymous referee for the helpful comments that greatly improve the quality of the paper.

ORCID

Jung-Chao Ban https://orcid.org/0000-0002-4920-6945
Yu-Liang Wu https://orcid.org/0000-0002-5181-056X

References

1. N. Aubrun and M.-P. Beal, Tree-shifts of finite type, *Theor. Comput. Sci.* **459** (2012) 16–25.
2. J.-C. Ban and C.-H. Chang, Tree-shifts: Irreducibility, mixing, and chaos of tree-shifts, *Trans. Amer. Math. Soc.* **369** (2017) 8389–8407.
3. J.-C. Ban and C.-H. Chang, Tree-shifts: The entropy of tree-shifts of finite type, *Nonlinearity* **30** (2017) 2785–2804.
4. J.-C. Ban, C.-H. Chang, W.-G. Hu and Y.-L. Wu, On structure of topological entropy for tree-shift of finite type, *J. Differ. Equ.* **292** (2021) 325–353.
5. J.-C. Ban, C.-H. Chang, W.-G. Hu and Y.-L. Wu, Topological entropy for shifts of finite type over $\mathbb{Z}$ and trees, *Theor. Comput. Sci.* **930** (2022) 24–32.
6. J.-C. Ban, C.-H. Chang, Y.-L. Wu and Y.-Y. Wu, Stem and topological entropy on cayley trees, *Math. Phys. Anal. Geom.* **25** (2022) 1.
7. I. Benjamini and Y. Peres, Markov chains indexed by trees, *Ann. Probab.* **22** (1994) 219–243.
8. R. M. Burton, C.-E. Pfister and J. E. Steif, The variational principle for Gibbs states fails on trees, *Markov Process. Relat. Fields* **3** (1995) 387–406.
9. T. Ceccherini-Silberstein and M. Coornaert, *Cellular Automata and Groups* (Springer Science & Business Media, 2010).
10. A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (Springer, 2010).
11. E. Louidor, B. Marcus and R. Pavlov, Independence entropy of $\mathbb{Z}^d$-shift spaces, *Acta Appl. Math.* **126** (2013) 297–317.
12. T. Meyerovitch and R. Pavlov, On independence and entropy for high-dimensional isotropic subshifts, *Proc. Lond. Math. Soc.* **109** (2014) 921–945.
13. K. Petersen and I. Salama, Tree shift topological entropy, *Theor. Comput. Sci.* **743** (2018) 64–71.
14. K. Petersen and I. Salama, Entropy on regular trees, *Discrete Contin. Dyn. Syst.* **40** (2020) 4453–4477.
15. K. Petersen and I. Salama, Asymptotic pressure on some self-similar trees, *Stoch. Dyn.* **23** (2022) 2350009.
16. S. T. Piantadosi, Symbolic dynamics on free groups, *Discrete Contin. Dyn. Syst.* **20** (2008) 725–738.

III

Received: 23 December 2023 | Revised: 26 March 2025 | Accepted: 5 May 2025

DOI: 10.1112/jlms.70198

RESEARCH ARTICLE

Journal of the London Mathematical Society

Hausdorff dimensions of irreducible Markov hom tree-shifts

Jung-Chao Ban[1,2] | **Guan-Yu Lai**[1] | **Yu-Liang Wu**[3]

[1]Department of Mathematical Sciences, National Chengchi University, Taipei, Taiwan, ROC

[2]Math. Division, National Center for Theoretical Science, National Taiwan University, Taipei, Taiwan, ROC

[3]Department of Mathematical Sciences, University of Oulu, Oulu, Finland

Correspondence
Yu-Liang Wu, Department of Mathematical Sciences, P.O. Box 3000, 90014 University of Oulu, Finland.
Email: Yu-Liang.Wu@oulu.fi

Funding information
National Science and Technology Council, Grant/Award Numbers: NSTC 111-2115-M-004-005-MY3, NSTC 111-2811-M-004-002-MY2; National Center for Theoretical Science

Abstract
This paper features a Cramér'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 means and a formula for the Hausdorff dimension of the symbolic space associated with the irreducible Markov chain.

1 | INTRODUCTION

This paper presents a version of Cramér's theorem for finite-state Markov chains indexed by d-trees, accompanied by a number of folklore theorems and an exploration of their connections to dimension theory.

Let $\mathcal{A}$ be a state space that is at most countable. Recall that a classical Markov chain is an $\mathcal{A}$-valued stochastic process $(X_n)_{n\in\mathbb{N}}$ satisfying the *Markov property*: For any $n \in \mathbb{N}$ and $a, b \in \mathcal{A}$, the finite-dimensional distributions of the process satisfy

$$\mathbb{P}(X_n = b | X_{n-1} = a, X_{n-2}, \dots, X_0) = \mathbb{P}(X_n = b | X_{n-1} = a) = M_{a,b}$$

with the *transition probabilities* $\{M_{a,b} : a, b \in \mathcal{A}\}$ independent of n. In this context, the transition probabilities satisfy $\sum_{b\in\mathcal{A}} M_{a,b} = 1$ for all $a \in \mathcal{A}$, forming a stochastic matrix $M = (M_{a,b})_{a,b\in\mathcal{A}}$ known as the *transition matrix*. Additionally, the law of the initial state X_0 can be represented as a probability vector $\pi = (\pi_a)_{a\in\mathcal{A}} = (\mathbb{P}(X_0 = a))_{a\in\mathcal{A}}$, referred to as the *initial distribution*. For example, a lattice random walk can be modeled as a Markov chain whose state space $\mathcal{A}$ consists of all nodes in the lattice and X_n is the position at time n with the transition probability $M_{a,b}$ specifying the probability to move from position a to b. Extending this paradigm, Benjamini and Peres in [5] introduced the concept of generalized Markov chains as $\mathcal{A}$-valued processes $(X_g)_{g\in T}$ indexed by some generalized "time" T. In this framework, T is a *rooted tree*, that is, a locally finite, connected acyclic graph that features a distinguished vertex ϵ called *root*. For $g, h \in T$, we write $g \leqslant h$ if g lies on the (unique) path from the root ϵ to h. We denote by $\tilde{g}$ the parent of g with $\tilde{g} \leqslant g$, and define $g \wedge h$ as the farthest vertex on both paths from ϵ to h and from ϵ to g. A tree-indexed Markov chain is then defined as follows.

Definition 1.1. Let $\mathcal{A}$ denote the state space, T a rooted tree, $M \in \mathbb{R}_{\geqslant 0}^{\mathcal{A}\times\mathcal{A}}$ a stochastic matrix, and $\pi \in \mathbb{R}_{\geqslant 0}^{\mathcal{A}}$ a probability vector. A *Markov chain indexed by T* with transition matrix M and initial distribution π is an $\mathcal{A}$-valued stochastic process $(X_g)_{g\in T}$ satisfying $\mathbb{P}(X_\epsilon = a) = \pi_a$ for all $a \in \mathcal{A}$ and the Markov property: For all $a, b \in \mathcal{A}$,

$$\mathbb{P}(X_g = b | X_{\tilde{g}} = a, X_h \text{ for } g \wedge h \leqslant \tilde{g}) = \mathbb{P}(X_g = b | X_{\tilde{g}} = a) = M_{a,b}, \tag{1}$$

Notably, the underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$ of a T-indexed Markov chain can be realized using a standard Borel space. Specifically, by defining $\Omega = \mathcal{A}^T$, $\mathcal{F}$ as the Borel σ-algebra, and $X_g(t) = t_g$, the Markov property together with the assumption of initial distribution uniquely determines a Borel probability measure (denoted also by $\mathbb{P}$ by a slight abuse of notation). This measure, referred to as a *Markov measure*, is supported within

$$\mathcal{T}_{\mathbf{M}} = \{t \in \mathcal{A}^T : \mathbf{M}_{t_{\tilde{g}}, t_g} = 1 \text{ for all } g \in T \setminus \{\epsilon\}\}, \tag{2}$$

where $\mathbf{M}$ is the *incidence matrix* associated with M, a matrix satisfying $\mathbf{M}_{a,b} = 0$ or 1 depending on $M_{a,b} = 0$ or > 0, respectively. In particular, when $T = \mathbb{Z}_+$, the measure $\mathbb{P}$ reduces to a classical Markov measure and the space $\mathcal{T}_{\mathbf{M}}$ to a Markov subshift, making them special cases of the broader framework. Motivated by this observation, the present study examines and compares tree-indexed and conventional Markov chains through the lenses of large deviation theory and dimension theory, and our focus is on the class of finite-state Markov chains indexed by rooted d-trees ($d \geqslant 2$), each corresponding to the Cayley graph of the free semigroup generated by $\Sigma = \{1, 2, \dots, d\}$:

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

with $\Sigma^0 = \{\epsilon\}$ being part of our convention.

Despite their introduction in [5] in a general form, special cases of tree-indexed Markov chains had appeared in earlier studies, typically as special cases of i.i.d. tree-indexed processes. For example, first-passage percolation on trees (cf. [21]) is modeled by tree-indexed i.i.d. random variables where each variable represents the "passage time" to traverse between vertices in T, and the primary focus is on the asymptotic passage time to reach a vertex from the root. For developments in this area, see, for example, [6] and see [14] for a multifractal analysis of first-passage percolation. Further discussion on tree-indexed i.i.d. variables and some related fields can be found in [6, Section 1] and [7].

Beyond the i.i.d. setting, the tree-indexed Markov chains, as explored in [6], are closely related to tree-indexed random walks on groups or general countable graphs, where the rooted tree plays the role of “generalized time” as illustrated in the previous paragraph. This approach addresses a key challenge in the analysis of random walks on large groups or graphs, where a single sample path visits only a relatively small portion of the space and thus provides limited information. With the abundance of sample paths offered by the generalized time, tree-indexed random walks provide a more comprehensive view of their underlying structure.

Another frequently explored aspect of nonconventional Markov chains is their entropy, which motivates the first part of the present work. This line of research originates from the study of statistical mechanics on trees, as initiated in [27] and [30]. For related studies, see also [9, 10, 13, 15]. Continuing in this direction, the article focuses on both the typical and large deviation behavior of configurations. Specifically, typical behavior is characterized by an ergodic theorem (also known as the Shannon–McMillan–Breiman theorem or the law of large numbers, depending on the context), while large deviation behavior is addressed through a Cramér’s theorem. This focus is influenced by [8] and its successors, where the researchers examined the entropy of the stationary random fields on both rooted and unrooted d-trees. Subsequent works by the same authors [35, 36], building upon Pemantle’s combinatorial approach [22], established a Shannon–McMillan–Breiman theorem with convergence in probability for ergodic random fields on homogeneous trees. Further developments can be found in [17, 29, 31, 33, 34] and references therein, with noteworthy contributions regarding deviation inequalities provided by [23, 28, 32].

Broadly speaking, the previously mentioned studies center on the asymptotic behavior of the following empirical average over the initial n-subtrees $\Delta_n := \cup_{i=0}^{n} \Sigma^i$:

$$Y_n = \frac{1}{|\Delta_n|} \sum_{g \in \Delta_n \setminus \{\epsilon\}} \log A_{X_{\tilde{g}}, X_g}, \tag{3}$$

where $A \in \mathbb{R}_{>0}^{\mathcal{A}\times\mathcal{A}}$ is the *weight* matrix and $|\cdot|$ denotes the cardinality of a set, with $|\Sigma^n| = d^n$ and $|\Delta_n| = \sum_{i=0}^{n} |\Sigma^i| = \frac{d^{n+1}-1}{d-1}$, in particular. As it turns out, the empirical averages Y_n exhibit an asymptotically periodic structure whose period is closely linked to that of the initial state. Recall that the *period* of a state a in a Markov chain is defined as $p = \gcd\{n \in \mathbb{N} : (M^n)_{a,a} > 0\}$, with $p = \infty$ if no such n exists. Recall also that a nonnegative matrix $A \in \mathbb{R}_{\geqslant 0}^{\mathcal{A}\times\mathcal{A}}$ is *irreducible* if for all $a, b \in \mathcal{A}$, there exists $n \in \mathbb{N}$ such that $(A^n)_{a,b} > 0$, and is said to be *primitive* if the condition holds for all large n. A Markov chain is said to be *irreducible* (respectively, *ergodic*) if it has an irreducible (respectively, primitive) transition matrix. As readily checked facts, all states in an irreducible Markov chain share a common period, which is referred to as the period of the Markov chain, and the period of an ergodic Markov chain is 1. In summary, the works listed in the previous paragraph successfully recover several classical theorems for ergodic Markov chains, but leave open the question of how the results extend explicitly to more general systems, in which the behavior of Y_n differs significantly from its conventional counterpart. Particularly, unlike classical Markov chains, almost sure convergence for (3) cannot be guaranteed even for irreducible Markov chains (see Example 3.12). In light of this, the conditional averages $(Y_{pn+j} | X_\epsilon = a_0)$ $(n \in \mathbb{N})$ with initial state a_0 of finite period p will be studied in place of Y_n throughout the work. Agreeing with this philosophy, our first result is an analog of Cramér’s theorem, which is presented after the introduction of necessary definitions and notations below.

Recall that given a positive real sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ tending to zero and a lower semicontinuous function $I : \mathbb{R} \to [0, \infty]$, a sequence of real-valued random variables $(X_n)_{n\in\mathbb{N}}$ is said to satisfy the *large deviation principle* with speed $(\varepsilon_n)_{n\in\mathbb{N}}$ and rate I if for all Borel set $S \subset \mathbb{R}$,

$$\sup_{\alpha\in\mathring{S}} -I(\alpha) \leqslant \liminf_{n\to\infty} \varepsilon_n \log \mathbb{P}(X_n \in S) \leqslant \limsup_{n\to\infty} \varepsilon_n \log \mathbb{P}(X_n \in S) \leqslant \sup_{\alpha\in\overline{S}} -I(\alpha).$$

For matrices $A, B \in \mathbb{R}^{\mathcal{A}\times\mathcal{A}}$ (and similarly $\mathbb{R}^{\mathcal{A}}$) and $r > 0$, we denote by $A \odot B := (A_{a,b}B_{a,b})_{a,b\in\mathcal{A}}$ the Hadamard product of A and B, and by $A^{\odot r} := (A^r_{a,b})_{a,b\in\mathcal{A}}$ the element-wise power of A whenever well defined. With these notations, we define the operator $\Psi_{A,r} : \mathbb{R}^{\mathcal{A}}_{\geqslant 0} \to \mathbb{R}^{\mathcal{A}}_{\geqslant 0}$ by

$$\Psi_{A,r}(x) = (Ax)^{\odot r}. \tag{4}$$

Theorem 1.2 (Cramér). *Let T be a rooted d-tree, $(X_g)_{g\in T}$ be a finite-state Markov chain with transition matrix M, and $(Y_n)_{n\in\mathbb{N}}$ be the empirical average with respect to the weight $A \in \mathbb{R}^{\mathcal{A}\times\mathcal{A}}_{>0}$ as in (3). If $a_0 \in \mathcal{A}$ is a state of finite period p, then $(Y_{pn+j}|X_\epsilon = a_0)_{n\in\mathbb{N}}$ satisfy the large deviation principle with speed $(|\Delta_{pn+j}|^{-1})_{n\in\mathbb{N}}$ and rate*

$$\Lambda^*_j(\alpha) = \sup_{\mu\in\mathbb{R}}[\mu\alpha - \Lambda_j(\mu)],$$

where, with $\mathbb{1} \in \mathbb{R}^{\mathcal{A}}$ denoting the all-one vector,

$$\Lambda_j(\mu) = \lim_{n\to\infty} \frac{\log((\Psi_{M\odot A^{\odot\mu},d})^{pn+j}(\mathbb{1}))_{a_0}}{|\Delta_{pn+j}|}.$$

A few comments about the theorem should be mentioned. First of all, we should emphasize that assuming the initial state has finite period involves no loss generality in the characterization of Y_n. Precisely, we observe that the random tree

$$\tilde{T} = \{g \in T : X_h \text{ has infinite period for all } h \leqslant g, h \neq g\}$$

is almost surely a subtree of $\Delta_{|\mathcal{A}|}$, and its leaf vertices $\partial\tilde{T} = \{h \in \tilde{T} : hi \notin \tilde{T}, \forall i \in \Sigma\}$ are associated only with states of finite period. This allows us to pass Theorem 1.2 to $(Y_n|X_g, g \in \partial\tilde{T})$. Second, a key challenge in proving Theorem 1.2 lies in the uncertainty of differentiability of Λ_j, and hence more sophisticated arguments, such as Gärtner–Ellis theorem, do not apply. This forces us to resort to a combinatorial reasoning. Finally, results on large deviations of tree-indexed Markov chains are also considered in [11] for trees generated by a Galton–Watson process with a positive probability of producing zero offspring, which differs from our setting.

Despite its tediousness, an advantage of our combinatorial approach lies in that it almost immediately gives the following version of ergodic theorem for irreducible Markov chains.

Theorem 1.3. *Under the assumptions of Theorem 1.2, if the Markov chain is irreducible of period p, then for any state $a_0 \in \mathcal{A}$,*

$$\lim_{n\to\infty}(Y_{pn+j}|X_\epsilon = a_0) = \sum_{i=0}^{p-1} \frac{d^{-i}}{\sum_{\ell=0}^{p-1} d^{-\ell}} \sum_{a,b\in\mathcal{A}} \pi_a^{(i+1-j)} M_{a,b} \log A_{a,b} \qquad \mathbb{P}\text{-a.s.},$$

where $(\pi^{(i)})_{i\in\mathbb{Z}}$ *are the unique left probability eigenvectors of* M^p *satisfying* $\pi^{(i)} = \pi^{(i+1)}M$ *and* $\pi^{(i)}_{a_0} > 0$ *if and only if* $i \equiv 0 \pmod p$*.*

This confirms the cyclic structure of convergence of Y_n and complements the state-of-the-art results by extending the limit theorems from primitive transition matrices to irreducible ones. This periodic pointwise convergence will be utilized in the subsequent sections.

The second part of the article is motivated by another intriguing aspect of the tree-indexed Markov chains: the "size" of the outcome space of labeled trees $\mathcal{T}_{\mathbf{M}}$ (recall (2)). This outcome space, also known as a Markov hom tree-shift, represents a special family of symbolic systems introduced by Aubrun and Béal [1] that can be viewed as a generalization of the one-sided Markov subshifts, considering that the latter could be defined using the former with the underlying tree $T = \mathbb{Z}_+$. Building upon this concept, Petersen and Salama [24, 25] introduced the topological entropy for such spaces as an analog to topological entropy for traditional shift spaces to investigate their complexity. Results therein were later generalized to asymptotic pressure for a broad class of systems [26]. Notably, the topological entropy could also be interpreted as the box-counting dimension under a proper metric (see (19)). Motivated by this observation, the present paper aims to investigate the Hausdorff dimension under the specified metric and relate such a quantity to the eigenvalues of certain nonlinear transfer operators: Drawing inspiration from the classical thermodynamic formalisms, the candidates of transfer operator $\mathcal{L}_{A,r} : \mathbb{R}^{\mathcal{A}}_{\geqslant 0} \to \mathbb{R}^{\mathcal{A}}_{\geqslant 0}$ for $r = (r_0, r_1, \dots, r_{p-1}) \in \mathbb{R}^p_{>0}$ and $A \in \mathbb{R}^{\mathcal{A}\times\mathcal{A}}_{\geqslant 0}$ are defined as

$$\mathcal{L}_{A,r}(x) = \Psi_{A,r_{p-1}} \circ \Psi_{A,r_{p-2}} \circ \cdots \circ \Psi_{A,r_0}(x), \tag{5}$$

with each $\Psi_{A,r_i}(x)$ defined in (4). The Hausdorff dimension is then related to the eigenvalues of the nonlinear operator, as is seen in the following theorem.

Theorem 1.4. *Let* $\mathcal{T}_{\mathbf{M}}$ *be a Markov hom tree-shift with an irreducible incidence matrix* $\mathbf{M}$ *of period* p*. Then,*

$$\dim_H \mathcal{T}_{\mathbf{M}} = \min_{r\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1} \prod_{i=0}^{\ell} r_i^{-1} \right)^{-1} \cdot \log \rho_*(\mathcal{L}_{\mathbf{M},r}), \tag{6}$$

where

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

$$\rho_*(\mathcal{L}_{\mathbf{M},r}) = \inf\{\alpha > 0 : \mathcal{L}_{\mathbf{M},r}(u) = \alpha \cdot u \in \mathbb{R}^{\mathcal{A}} \setminus \{0\}\}.$$

In particular, if $\mathbf{M}$ *is a primitive matrix, then* $p = 1$*,* $\mathcal{L}_{\mathbf{M},r} = \mathbf{M}$ *for* $r \in \mathcal{R}_{1,d} = \{1\}$*, and* $\dim_H \mathcal{T}_{\mathbf{M}} = \log \rho_*(\mathbf{M}) = \log \rho(\mathbf{M})$*, where* $\rho(\mathbf{M})$ *is the spectral radius of* $\mathbf{M}$*.*

It is noteworthy that the theorem formally generalizes the classical thermodynamic formalisms. Specifically, when $d = 1$, the set $\mathcal{R}_{p,d}$ simplifies to $= \{(1, 1, \dots, 1)\}$, leading formally to the conclusion that $\dim_H \mathcal{T}_{\mathbf{M}} = p^{-1} \log \rho(\mathbf{M}^p) = \log \rho(\mathbf{M})$, a known result for irreducible Markov subshifts.

At this stage, two main concerns arise regarding Theorem 1.4: the existence of nonnegative eigenvectors and the attainment of $\min_{r\in\mathcal{R}_{p,d}}$. To address the former issue, the proof of the theorem is deferred to Section 4.1, where a brief overview of nonlinear Perron–Frobenius theory is provided to establish the required existence. The latter issue is discussed in Lemma 4.4, which forms the cornerstone in the proof of Theorem 1.4. Notably, the proposed formula of the Hausdorff dimension can be further adapted to serve as an upper bound for any Markov hom tree-shifts, leading to the following corollary.

Corollary 1.5. *Let $\mathcal{T}_{\mathbf{M}}$ be a Markov hom tree-shift. Then, $\dim_H \mathcal{T}_{\mathbf{M}} \leqslant \log \rho(\mathbf{M})$. In particular, for any irreducible $\mathbf{M}$, the equality holds if and only if $\mathbf{M}$ has uniform row sums.*

To provide an overview of our proof strategy, it is worth mentioning that the upper bound of the Hausdorff dimension is derived via constructing efficient covers of $\mathcal{T}_{\mathbf{M}}$, 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 not only in providing a formula for Hausdorff dimension but also in establishing an analog of the variational principle for the tree-shifts.

The organization of the paper is as follows. In Section 2, we introduce preliminary definitions and establish the general conventions. Section 3 is devoted to the proofs of Theorems 1.2 and 1.3. Section 4 presents the proposed variational formula for Hausdorff dimensions (Theorem 1.4), which involves a recapitulation of the nonlinear Perron–Frobenius theory (Section 4.1), followed by necessary lemmas on duality (Section 4.2) and lower and upper bounds for Hausdorff dimensions (Sections 4.3 and 4.4). Section 4.5 concludes with a discussion of a general upper bound for the Hausdorff dimension and an illustrative example validating our dimension formula. Finally, the paper concludes with a brief discussion in Section 5.

2 | PRELIMINARIES

We let T, Σ, Δ_n, and so on, be as defined and use the alias $\Xi_n = \Sigma^n$ to avoid possible confusion. We will consistently denote the transition matrix as M and its incidence matrix as $\mathbf{M}$. To simplify our discussion, two assumptions will be made throughout unless otherwise mentioned. The first one is rather fundamental, which states as follows:

$$\textstyle\sum_b \mathbf{M}_{a,b} > 0 \text{ for all } a \in \mathcal{A}. \tag{A0}$$

By assuming so, we avoid "nonessential" states in our discussion, as any a failing the criterion would appear in neither $\mathcal{T}_{\mathbf{M}}$ nor the Markov chain, and can be easily excluded by restricting to a subset $\mathcal{A}' = \mathcal{A} \setminus \{a\}$ of the state space without actually altering the systems. The second assumption owes to our exclusive attention to states of finite period as in Theorem 1.2, and it requires the following:

$$\text{There exists } a_0 \in \mathcal{A} \text{ such that every } a \in \mathcal{A} \text{ admits } n \in \mathbb{N} \text{ satisfying } (\mathbf{M}^n)_{a_0,a} > 0. \tag{A1}$$

Though such a_0 might not be unique, the choice will not affect our discussion and will henceforth be fixed. Notably, this assumption becomes transparent if the incidence matrix $\mathbf{M}$ is irreducible, as it is automatically satisfied under the circumstances. It plays a role only when $\mathbf{M}$ is not irreducible, in which case states are pruned if they would not be seen given the initial state a_0, simplifying discussion of Theorem 1.2. We should also mention that such requirement will only be dropped in Section 4.5 when deriving an upper bound of Hausdorff dimensions for general Markov hom tree-shift. By writing $[k] = \{0, 1, \dots, k-1\}$ for $k \in \mathbb{N}$, the assumption (A1) naturally induces a collection $\mathcal{P}(\mathbf{M}) = \{\mathcal{A}_j : j \in [p]\}$ of subsets of $\mathcal{A}$ such that

$$a \in \mathcal{A}_j \text{ if and only if } (\mathbf{M}^n)_{a_0,a} > 0 \text{ for some } n \equiv j \pmod{p}, \tag{7}$$

where we set $\mathcal{A}_{i+p} = \mathcal{A}_i$ for all $i \in \mathbb{Z}$. Although technical, $\mathcal{P}(\mathbf{M})$ is essential for characterizing the cyclic structure of convergence in Theorems 1.2 and 1.3. In particular, $\mathcal{P}(\mathbf{M})$ forms a partition of $\mathcal{A}$ if $\mathbf{M}$ is irreducible.

Some frequently used symbols are borrowed from the work of the first and the third authors [3]. 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 $\Gamma_{\mathcal{A}}$; explicitly, this means that given any $\mathsf{p} \in \Gamma_{\mathcal{A}}$ and any $\mathsf{P} \in \Upsilon_{\mathcal{A}}$, p, as a row vector, can always be multiplied by P from the right so that their product pP is still a probability vector. For conciseness, $\mathcal{A}$ is suppressed from $\Gamma_{\mathcal{A}}$ and $\Upsilon_{\mathcal{A}}$, unless other index sets are involved. We define the *variational distances* d_v and d_V on Γ and Υ, respectively, as

$$\mathsf{d}_v(\mathsf{p}, \mathsf{q}) = \frac{1}{2} \sum_{a \in S} |\mathsf{p}_a - \mathsf{q}_a| \text{ and } \mathsf{d}_V(\mathsf{P}, \mathsf{Q}) = \max_{a \in \mathcal{A}} \mathsf{d}_v((\mathsf{P}_{a,b})_{b \in \mathcal{A}}, (\mathsf{Q}_{a,b})_{b \in \mathcal{A}}).$$

Associated with the metrics is a sup metric $\mathsf{d}_{v,V}((\mathsf{p}, \mathsf{P}), (\mathsf{q}, \mathsf{Q}))$ on the product space $\Gamma \times \Upsilon$ defined as

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

For a labeled tree $t \in \mathcal{A}^T$, the *nth-level distribution* of t is denoted as

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

In addition, the *nth-level transition* of t is written as

$$\eta_n(t)_{a,b} = \begin{cases} \left(\frac{|\{g \in \Xi_{n+1} : t_{\tilde{g}} = a, t_g = b\}|}{|\{g \in \Xi_{n+1} : t_{\tilde{g}} = a\}|} \right) & \text{if } |\{g \in \Xi_{n+1} : t_{\tilde{g}} = a\}| > 0, \\ \frac{\mathbf{M}_{a,b}}{\sum_{c \in \mathcal{A}} \mathbf{M}_{c,b}} & \text{otherwise.} \end{cases}$$

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

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

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

$$S_{n:m} = \{\eta_{n:m-1}(t) : t \in \mathcal{T}_{\mathbf{M}}\}$$

is the set of all admissible transitions. Also, let

$$W_{n:m} = \{(\tau_{n:m}(t), \eta_{n:m-1}(t)) : t \in \mathcal{T}_{\mathbf{M}}\}$$

be the collective admissible sets of distributions and transitions. In a similar fashion, we may extend the definition of a subtree Δ_n to a subgraph consisting of vertices from level n to level m:

$$\Delta_{n:m} = \cup_{i=n}^{m} \Xi_i,$$

so the set of admissible blocks by prescribed distributions $\mathsf{q} = (\mathsf{q}^{(0)}, \dots, \mathsf{q}^{(m-n)})$ and transitions $\mathsf{Q} = (\mathsf{Q}^{(0)}, \dots, \mathsf{Q}^{(m-n-1)})$ is defined as

$$B_{n:m}(\mathsf{q}, \mathsf{Q}) = \{t|_{\Delta_{n:m}} \in \mathcal{A}^{\Delta_{n:m}} : t \in \mathcal{T}_{\mathbf{M}}, \tau_{m:n}(t) = \mathsf{q}, \eta_{n:m-1}(t) = \mathsf{Q}\}.$$

In particular, the set of all admissible blocks $B_{n:m}$ is written as the union of $B_{n:m}(\mathsf{q}, \mathsf{Q})$:

$$B_{n:m} = \cup_{(\mathsf{q},\mathsf{Q}) \in W_{n:m}} B_{n:m}(\mathsf{q}, \mathsf{Q}).$$

For simplicity, "n :" is suppressed in the notation of $D_{n:m}$, $S_{n:m}$, $W_{n:m}$, and $B_{n:m}$ if $n = 0$. Notably, for $(\mathsf{q}, \mathsf{Q}) \in W_{n:m}$, it is necessary that (a) $\mathsf{q}^{(i+1)} = \mathsf{q}^{(i)}\mathsf{Q}^{(i)}$, and (b) $\mathsf{Q}^{(i)}_{a,b} = 0$ if $\mathbf{M}_{a,b} = 0$ for every $a, b \in \mathcal{A}$ and all $0 \leqslant i < m - n$. For brevity, We write $\mathsf{Q}^{(i)} \prec \mathbf{M}$ whenever (b) is satisfied and write $\mathsf{Q}^{(i)} \sim \mathbf{M}$ when both $\mathsf{Q}^{(i)} \prec \mathbf{M}$ and $\mathbf{M} \prec \mathsf{Q}^{(i)}$. For convenience's sake, we denote by $(\mathsf{p}, \mathsf{P}) = (\overleftarrow{\mathsf{q}}, \overleftarrow{\mathsf{Q}})$ the reversed sequence of (q, Q), (i.e., $(\mathsf{p}^{(0)}, \dots, \mathsf{p}^{(n-m)}) = (\mathsf{q}^{(n-m)}, \dots, \mathsf{q}^{(0)})$ and $(\mathsf{P}^{(0)}, \dots, \mathsf{P}^{(n-m-1)}) = (\mathsf{Q}^{(n-m-1)}, \dots, \mathsf{Q}^{(0)})$) and consider the following set of all extended reversed sequences:

$$Z := \{(\mathsf{p}, \mathsf{P}) \in \Gamma^{\mathbb{Z}_+} \times \Upsilon^{\mathbb{Z}_+} : \mathsf{p}^{(i)} = \mathsf{p}^{(i+1)}\mathsf{P}^{(i)} \text{ and } \mathsf{P}^{(i)} \prec \mathbf{M} \text{ for all } i \in \mathbb{Z}_+\}, \tag{8}$$

together with which we assign

$$C_j = \{x \in \mathbb{R}^{\mathcal{A}}_{\geqslant 0} : x_a = 0 \text{ if } a \notin \mathcal{A}_j\} \tag{9}$$

and

$$Z_j = \{(\mathsf{p}, \mathsf{P}) \in Z : \mathsf{p}^{(i)} \in C_{j-i} \text{ for all } i \in \mathbb{Z}_+\}. \tag{10}$$

Additionally, for any sequence of probability vectors (and similarly for stochastic matrices), say $\mathsf{p} \in \Gamma^{\mathbb{Z}_+}$, we write

$$\mathsf{p}^{(i:j)} = (\mathsf{p}^{(i)}, \mathsf{p}^{(i+1)}, \dots, \mathsf{p}^{(j)}), \quad 0 \leqslant i \leqslant j \leqslant \infty.$$

Finally, the convention of matrices and vectors is manifested as follows. 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. In particular, $\{\mathsf{e}_a : a \in \mathcal{A}\}$ is the standard basis of $\mathbb{R}^{\mathcal{A}}$.

TABLE 1 Summary of notation.

Σ	generating set of the semigroup
T	free semigroup generated by Σ
Ξ_n	alias of Σ^n
Δ_n	$\cup_{i=0}^{n} \Xi_i$
$\mathcal{A}$	state space
$\mathbf{M}$	incidence matrix
$\mathcal{T}_{\mathbf{M}}$	Markov hom tree-shift associated with $\mathbf{M}$
$[k]$	$\{0, 1, \dots, k-1\}$
$\tilde{g}$	parent of $g \in T$
a_0	label satisfying (A1)
p	period of state a_0
$\mathcal{P}(\mathbf{M})$	$\{\mathcal{A}_j : j \in [p]\}$
$\Gamma_{\mathcal{A}}, \Gamma$	set of probability vectors indexed by $\mathcal{A}$
$\Upsilon_{\mathcal{A}}, \Upsilon$	set of stochastic matrix indexed by $\mathcal{A}$
p, q	vectors in $\Gamma_{\mathcal{A}}$
$M, \mathsf{P}, \mathsf{Q}$	matrices in $\Upsilon_{\mathcal{A}}$
$\tau_n(t), \tau_{n:m}(t)$	distribution vectors associated with $t \in \mathcal{T}_{\mathbf{M}}$
$\eta_n(t), \eta_{n:m}(t)$	transition matrices associated with $t \in \mathcal{T}_{\mathbf{M}}$
$\mathsf{d}_v, \mathsf{d}_V, \mathsf{d}_{v,V}$	variational distances
$D_{n:m}, D_n$	set of feasible distributions by $\mathcal{T}_{\mathbf{M}}$
$S_{n:m}, S_n$	set of feasible transitions by $\mathcal{T}_{\mathbf{M}}$
$W_{n:m}, W_n$	set of collectively feasible distributions and transitions by $\mathcal{T}_{\mathbf{M}}$
$B_{n:m}, B_n$	feasible patterns of $\mathcal{T}_{\mathbf{M}}$
$B_{n:m}(\mathsf{q}, \mathsf{Q}), B_n(\mathsf{q}, \mathsf{Q})$	feasible patterns of $\mathcal{T}_{\mathbf{M}}$ with distributions q and transitions Q
$(\mathsf{p}, \mathsf{P}) = (\overleftarrow{\mathsf{q}}, \overleftarrow{\mathsf{Q}})$	reversed sequence of (q, Q)
Z, Z_j	feasible set of extended reversed sequences
C_j	vectors supported on $\mathcal{A}_j$
$D_{\mathrm{KL}}(\mathsf{P}\|\|A)$	relative entropy of P and A
$A \odot B$	Hadamard product of A and B
$A^{\odot r}$	matrix or vector A raised to power r
$(X_g)_{g \in T}$	Markov chain indexed by T

Except for probability vectors in Γ, every vector in the article is by default a column vector. In addition, for any $\mathsf{P} \in \Upsilon$ and any nonnegative matrix $A \in \mathbb{R}_{\geqslant 0}^{\mathcal{A} \times \mathcal{A}}$ satisfying $\mathsf{P} \prec A$, we define their "Kullback–Leibler divergence" as

$$D_{\mathrm{KL}}(\mathsf{P}||A)_b := \sum_{b \in \mathcal{A}} \mathsf{P}_{a,b} \log \left(\frac{\mathsf{P}_{a,b}}{A_{a,b}} \right),$$

where $0 \log \frac{0}{A_{a,b}}$ is interpreted as 0 for $A_{a,b} \geqslant 0$. Also, when functions, such as division /, exponential function e, and logarithm log, are acting on matrices or vectors, the actions are by default entrywise unless stated otherwise. The notations are summarized in Table 1.

3 | LARGE DEVIATIONS OF MARKOV CHAINS ON *d*-TREES

The aim of this section is to carry out the proof of Theorem 1.2, which states the large deviation principle for the conditional sample mean.

To illustrate our proof strategy, we present a heuristic proof of the following version of Cramér's theorem (see, for instance, [12, Theorem 2.1.24] for a similar version).

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 $\{a_1, a_2, \dots, a_k\}$. Then,*

$$\frac{1}{n}\log\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}X_i \geqslant \alpha\right) \to \sup_{\mu\in\mathbb{R}}\,[\mu\alpha - \Lambda(\mu)] \textit{ for } \alpha \geqslant \mathbb{E}[X_1],$$

where

$$\Lambda(\mu) = \log\mathbb{E}[e^{\mu X_1}] = \log\sum_{j=1}^{k} p_j \cdot e^{\mu a_j} \textit{ with } (p_j)_{j=1}^k = (\mathbb{P}(X_1 = a_j))_{j=1}^k.$$

By virtue of Stirling's approximation, one can reformulate the probability in terms of an optimization problem:

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

where χ_a denotes the characteristic function of the state a and the first equality holds since the set

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

has its cardinality bounded from above by $\binom{n}{k-1}$, and the second from the fact that L_n is asymptotically dense in the simplex of probability vectors. The theorem is ultimately proved by deriving the rate function $\sup_{\mu\in\mathbb{R}}[\mu\alpha - \Lambda(\mu)]$, as a dual problem, utilizing standard techniques from optimization theory. Inspired by this argument, the authors analogously reproduced the necessary approximations and estimations in [3], based upon which the method of types (see, e.g., [12, Chapter 2.1.1]) can be adapted for Theorem 1.2.

To prove Theorem 1.2, 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 $\Gamma^{\mathbb{Z}_+} \times \mathrm{Y}^{\mathbb{Z}_+}$, which is defined as

$$\mathsf{d}_{v,V}^{\infty}((\mathsf{p},\mathsf{P}),(\mathsf{q},\mathsf{Q})) := \sum_{i=0}^{\infty} \frac{d-1}{d^{i+1}} \cdot \mathsf{d}_{v,V}((\mathsf{p}^{(i)},\mathsf{P}^{(i)}),(\mathsf{q}^{(i)},\mathsf{Q}^{(i)})) \qquad (d \in \mathbb{N} \setminus \{1\}). \tag{11}$$

We note that the product topology is compatible with this metric, rendering $\Gamma^{\mathbb{Z}_+} \times \mathrm{Y}^{\mathbb{Z}_+}$ a compact metric space. For the scope of this article, the following two families of functions on Z are of interest. For $0 \leqslant n \leqslant m \leqslant \infty$ and $A \in \mathbb{R}_{\geqslant 0}^{\mathcal{A}\times\mathcal{A}}$ satisfying $\mathbf{M} \prec A$, define $\mathcal{F}_{n:m}, \mathcal{G}_{n:m} : Z \to \mathbb{R}$ (with parameter A) by

$$\mathcal{F}_{n:m}(\mathsf{p},\mathsf{P};A) = -\sum_{i=n}^{m} \frac{d-1}{d^{i-n+1}} \mathsf{p}^{(i+1)} D_{\mathrm{KL}}(\mathsf{P}^{(i)}||A),$$

$$\mathcal{G}_{n:m}(\mathsf{p},\mathsf{P};A) = \sum_{i=n}^{m} \frac{d-1}{d^{i-n+1}} \mathsf{p}^{(i+1)}(\mathsf{P}^{(i)} \odot \log A)\mathbb{1},$$

where, as previously, $A \odot B$ denotes the Hadamard product of matrices A and B, the logarithm is applied entrywise, and $0 \log 0 = 0$. For conciseness, we suppress "n :" when $n = 0$ and "m" when $m = \infty$. The functions $\mathcal{F}$ and $\mathcal{G}$ turn out to be continuous, as shown in the following lemma.

Lemma 3.2. *Suppose $A \in \mathbb{R}_{\geqslant 0}^{\mathcal{A}\times\mathcal{A}}$ satisfying $\mathbf{M} \prec A$. The functions $\mathcal{F}(\cdot;A)$, $\mathcal{G}(\cdot;A) : Z \to \mathbb{R}$ are uniformly convergent and thus continuous.*

Proof. Indeed, it is because the functions $(\mathsf{p},\mathsf{P}) \mapsto \mathsf{p}^{(i+1)} D_{\mathrm{KL}}(\mathsf{P}||A)$ and $(\mathsf{p},\mathsf{P}) \mapsto \mathsf{p}^{(i+1)}(\mathsf{P}^{(i)} \odot \log A)\mathbb{1}$ are continuous and uniformly bounded over Z. □

The continuity and compactness are crucial to the rest of our discussion. In particular, they guarantee that for $\alpha \in \mathbb{R}$ and all $n \in \mathbb{N} \cup \{\infty\}$, the domains

$$Z^{(n)}(\alpha;A) = \{(\mathsf{p},\mathsf{P}) \in Z : \mathcal{G}_n(\mathsf{p},\mathsf{P};A) = \alpha\} \text{ and } Z_j^{(n)}(\alpha;A) = Z^{(n)}(\alpha;A) \cap Z_j$$

are compact, of which we omit the superscript "(n)" if $n = \infty$.

Our next step is to discuss combinatorial approximations of the empirical distribution of patterns, which are manifested in the following series of lemmas.

Lemma 3.3 Proposition 2.2, [3]. *For any $n \leqslant m$,*

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

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

where $\Delta_{n:m} = \cup_{i=n}^{m} \Xi_i$ has cardinality $\sum_{i=n}^{m} = d^i$.

The lemma indicates that the growth of admissible distributions and transitions is subexponential with respect to $\Delta_{n:m}$, which is markedly smaller than the exponential growth rate of the cardinality of the set $B_{n:m}$. The next two lemmas, on the other hand, demonstrate that the set $W_{n:m}$ is, in some sense, asymptotically dense in Z, and Stirling's approximations can be applied as in our heuristic proof of Cramér's theorem. Indeed, if we define

$$W'_k = \left\{ (\mathsf{p}, \mathsf{P}) \in \left(\Gamma^k \times \{\mathsf{e}_a\}_{a\in\mathcal{A}}\right) \times \Upsilon^k \,:\, \mathsf{P}^{(i)} \prec \mathbf{M}, \mathsf{p}^{(i+1)} = \mathsf{p}^{(i)}\mathsf{P}^{(i)}, \right.$$
$$\left. \mathsf{P}^{(i)}_{a,b} = \frac{\mathbf{M}_{a,b}}{\sum_{c\in\mathcal{A}} \mathbf{M}_{c,b}} \text{ if } \mathsf{p}^{(i+1)}_b = 0 \text{ for all } i \in [k] \right\},$$

we have the following lemmas.

Lemma 3.4 [3, Proposition 3.2]. *Let* $\mathsf{d}^k_{\upsilon,V}$ *be any product metric on the space* $\Gamma^{k+1} \times \Upsilon^k$*, say, for example,*

$$\mathsf{d}^k_{\upsilon,V}((\mathsf{p},\mathsf{P}),(\mathsf{q},\mathsf{Q})) := \max\left\{ \max_{0\leqslant i\leqslant k} \mathsf{d}_\upsilon(\mathsf{p}^{(i)},\mathsf{q}^{(i)}), \max_{0\leqslant i< k} \mathsf{d}_V(\mathsf{P}^{(i)},\mathsf{Q}^{(i)}) \right\}.$$

Then, $\lim_{n\to\infty} sup_{(\mathsf{p},\mathsf{P})\in W'_k} d^k_{\upsilon,V}(W_{n:n+k}, (\overleftarrow{\mathsf{p}}, \overleftarrow{\mathsf{P}})) = 0.$

Notably, given any $(\mathsf{p},\mathsf{P}) \in Z$, the values of $\mathcal{F}_n(\mathsf{p},\mathsf{P};A)$ and $\mathcal{G}_n(\mathsf{p},\mathsf{P};A)$ are completely independent of $\mathsf{P}^{(i)}_{a,b}$ whenever $\mathsf{p}^{(i+1)}_a = 0$, which renders the corresponding condition in the definition of W'_k transparent.

Lemma 3.5. *For any* $(\mathsf{q},\mathsf{Q}) \in W_n$ *and any* $(\mathsf{p},\mathsf{P}) \in Z$ *satisfying* $(\mathsf{p}^{(0:n)}, \mathsf{P}^{(0:n-1)}) = (\overleftarrow{\mathsf{q}}, \overleftarrow{\mathsf{Q}})$,

$$\frac{1}{|\Delta_n|} \sum_{u\in B_n(\mathsf{q},\mathsf{Q})} \sum_{g\in\Delta_n\setminus\{\epsilon\}} \log A_{u_{\tilde{g}},u_g} = (1-d^{-n-1}) \cdot \mathcal{F}_n(\mathsf{p},\mathsf{P};A) + O\left(\frac{\log|\Delta_n|}{|\Delta_n|}\right),$$

where the implicit constant in the big O notation is independent of (q,Q) *and* (p,P).

Proof. The equality follows from [3, Proposition 3.1] together with the observation that $\frac{d-1}{d^{n-i+1}} / \frac{|\Xi_i|}{|\Delta_n|} = 1 - d^{-n-1}$. □

Similar to the classical arguments using the method of types, the theory of convex optimization plays a key role. In the present work, we resort to Sion's minimax theorem (see, e.g., [18] for a proof). To apply it, we need to verify *a priori* the concavity/convexity of the objective function, and this is the moment when the following lemma is apropos. Throughout the discussion, the convexity of a vector-valued function is defined in an entrywise sense, as stated below.

Definition 3.6. Let V be a convex subset of a vector space. An extended vector-valued function $(f_a)_{a\in\mathcal{A}} : V \to (\mathbb{R} \cup \{\pm\infty\})^{\mathcal{A}}$ is said to be *convex* if each f_a is convex over the domain V.

Lemma 3.7. *Let V be a convex set in a vector space. Suppose that*

- $E : V \to \mathbb{R}_{\geqslant 0}^{\mathcal{A}\times\mathcal{A}}$ *satisfies* $E(x) \sim \mathbf{M}$ *for all* $x \in V$ *and* $\log E$ *(entrywise logarithm) is convex,*
- $\lambda : V \to \mathbb{R}^{\mathcal{A}}$ *is convex, and*
- $q : V \to \mathbb{R}_{>0}$ *is affine.*

If E or q is constant, then the following map is convex:

$$x \mapsto q(x)\log(E(x)e^{q(x)^{-1}\lambda(x)}),$$

where exponentiation and logarithm are both applied entrywise.

Proof. If $E(x) = E$ is constant, then for nonnegative α, α' satisfying $\alpha + \alpha' = 1$, we have

$$\begin{aligned}
&(q(\alpha x + \alpha' x'))\log\left(Ee^{\frac{\lambda(\alpha x+\alpha' x')}{q(\alpha x+\alpha' x')}}\right) = \log\left(\left(Ee^{\frac{\lambda(\alpha x+\alpha' x')}{\alpha q(x)+\alpha' q(x')}}\right)^{\alpha q(x)+\alpha' q(x')}\right)\\
&\leqslant \log\left(\left(Ee^{\frac{\alpha q(x)}{\alpha q(x)+\alpha' q(x')}\frac{\lambda(x)}{q(x)}+\frac{\alpha' q(x')}{\alpha q(x)+\alpha' q(x')}\frac{\lambda(x')}{q(x')}}\right)^{\alpha q(x)+\alpha' q(x')}\right)\\
&\leqslant \log\left(\left(Ee^{\frac{\lambda(x)}{q(x)}}\right)^{\alpha q(x)}\right) + \log\left(\left(Ee^{\frac{\lambda(x')}{q(x')}}\right)^{\alpha' q(x')}\right)\\
&= \alpha q(x)\log(Ee^{q(x)^{-1}\lambda(x)}) + \alpha' q(x')\log(Ee^{q(x')^{-1}\lambda(x')}),
\end{aligned}$$

where the first inequality results from convexity and the second from Hölder's inequality. Similarly, if, without loss of generality, $q(x) = 1$ is constant, then

$$\begin{aligned}
&\log\left(E(\alpha x + \alpha' x')e^{\lambda(\alpha x+\alpha' x')}\right) = \log\left(e^{\log E(\alpha x+\alpha' x')}e^{\lambda(\alpha x+\alpha' x')}\right)\\
&\leqslant \log\left(e^{\alpha\log E(x)+\alpha'\log E(x')}e^{\alpha\lambda(x)+\alpha'\lambda(x')}\right)\\
&\leqslant \alpha\log\left(E(x)e^{\lambda(x)}\right) + \alpha'\log\left(E(x')e^{\lambda(x')}\right),
\end{aligned}$$

by Hölder's inequality. □

Lemma 3.8. *Suppose that assumption (A1) holds. Then, for all sufficiently large $n_0 \in \mathbb{N}$, $\min_{a\in\mathcal{A}_0}(\mathbf{M}^{pn_0})_{a_0,a} > 0$. In addition, for every $\mathsf{q} \in C_0 \cap \Gamma$, there exists $(\mathsf{p}, \mathsf{P}) \in Z_0$ such that $\mathsf{p}^{(pn_0)} = \mathsf{e}_{a_0}$ and that $\mathsf{p}^{(0)} = \mathsf{q}$.*

Proof. The existence of n_0 satisfying $\min_{a\in\mathcal{A}_0}(\mathbf{M}^{pn_0})_{a_0,a} > 0$ follows from (A1) and that a_0 has finite period p. To verify the additional properties are also satisfied, we first construct, for every $a \in \mathcal{A}_0$, a sequence $(b^{a,i})_{i\in\mathbb{Z}_+}$ satisfying that $b^{a,0} = a$, that $b^{a,pn_0} = a_0$, and that $\mathbf{M}_{b^{a,i+1},b^{a,i}} = 1$ for all $i \in \mathbb{Z}_+$. Additionally, we assume that for every $i = 0, 1, \ldots, pn$, the condition $b^{a,i} = b^{a',i}$ implies $(b^{a,j})_{j\geqslant i} = (b^{a',j})_{j\geqslant i}$. To achieve this, if $b^{a,1} = b^{a',1}$ but $(b^{a,j})_{j\geqslant 1} \neq (b^{a',j})_{j\geqslant 1}$, then $(b^{a',j})_{j\geqslant 1}$ can be replaced by $(b^{a,j})_{j\geqslant 1}$, and the replacement continues until the property is satisfied for all $b^{a,1}$,

$a \in \mathcal{A}_0$. Iterating process for $i = 1, \ldots, pn_0$, the property is further satisfied by $b^{a,i}$ for i and all a. The lemma is then concluded by choosing

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

and

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

□

Lemma 3.9. *Suppose that assumption (A1) holds. Let $A \in \mathbb{R}^{\mathcal{A}\times\mathcal{A}}_{\geqslant 0}$ satisfy $\mathbf{M} \prec A$ and define $E(\mu) = M \odot A^{\odot\mu}$ for every $\mu \in \mathbb{R}$. Then, the following hold.*

(1) *Let $(\lambda^{(n)}(\mu))_{n\in\mathbb{Z}_+}$ be a sequence of $\mathbb{R}^{\mathcal{A}}$-valued functions iteratively defined by*

$$\lambda^{(n+1)}(\mu) = \frac{d-1}{d^{i+1}} \log\left(E(\mu) e^{\frac{d^{n+1}}{d-1}\lambda^{(n)}(\mu)}\right) \text{ with } \lambda^{(0)}(\mu) = 0. \tag{12}$$

Then, $\mu \mapsto \lambda^{(n)}(\mu)$ is convex for all $n \in \mathbb{Z}_+$.

(2) *For all $a \in \mathcal{A}$ and $n \in \mathbb{N}$,*

$$\lambda^{(n)}(\mu)_a = \max_{(\mathsf{p},\mathsf{P})\in Z, \mathsf{p}^{(n)} = \mathsf{e}_a} \mathcal{F}_n(\mathsf{p}, \mathsf{P}; E(\mu)),$$

$$\inf_{\mu\in\mathbb{R}} [-\mu\alpha + \lambda^{(n)}(\mu)_a] = \sup_{(\mathsf{p},\mathsf{P})\in Z^{(n)}(\alpha;A), \mathsf{p}^{(n)} = \mathsf{e}_a} \mathcal{F}_n(\mathsf{p}, \mathsf{P}; M).$$

(3) *Let $n_0 \in \mathbb{N}$ be a constant given in Lemma 3.8. Then, there exists an absolute constant $C > 0$ such that for all $n \in \mathbb{N}$,*

$$|\lambda^{(n+pn_0)}(\mu)_{a_0} - \max_{a\in\mathcal{A}_0} \lambda^{(n)}(\mu)_a| \leqslant Cd^{-n}(|\mu| + 1).$$

(4) *Let Λ_j, Λ_j^* be as defined in Theorem 1.2. Then, Λ_j is well defined and convex, satisfying*

$$\Lambda_j(\mu) = \max_{(\mathsf{p},\mathsf{P})\in Z_j} \mathcal{F}(\mathsf{p}, \mathsf{P}; E(\mu)).$$

Also, Λ_j^* is convex and lower semicontinuous, satisfying

$$-\Lambda_j^*(\alpha) = \sup_{(\mathsf{p},\mathsf{P})\in Z_j(\alpha;A)} \mathcal{F}(\mathsf{p}, \mathsf{P}; M).$$

In particular, $(\Lambda_j^*)^{-1}[0, \infty) = [\alpha_1, \alpha_2]$ is a nonempty compact interval.

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

(1) The convexity is secured by Lemma 3.7.

(2) The first equality is given in [3, Proposition 11]. To prove the second, note that

$$\sup_{(\mathsf{p},\mathsf{P})\in Z_j}\left\{\mathcal{F}_n(\mathsf{p},\mathsf{P};M)\,:\,\mathcal{G}_n(\mathsf{p},\mathsf{P};A)=\alpha,\mathsf{p}^{(n)}=\mathsf{e}_a\right\}$$

$$=\sup_{\substack{\mathsf{P}^{(0:n-1)}\in \Upsilon^n\\ \forall i,\mathsf{P}^{(i)}\prec\mathbf{M}}}\inf_{\mu\in\mathbb{R}}\left[-\mu\alpha-\sum_{i=0}^{n-1}\frac{d-1}{d^{i+1}}\mathsf{e}_a\left(\prod_{\ell=1}^{n-i-1}\mathsf{P}^{(n-\ell)}\right)D_{\mathrm{KL}}(\mathsf{P}^{(i)}||E)\right],$$

where $\prod_{i=n}^{m}\mathsf{P}^{(i)}=\mathsf{P}^{(n)}\mathsf{P}^{(n+1)}\cdots\mathsf{P}^{(m)}$ is set to identity matrix if $m<n$. By fixing $\mathsf{P}^{(1:n-1)}$ and $\mathsf{p}^{(n)}$, we may apply the minimax theorem to swap "$\sup_{\mathsf{P}^{(0)}\in\Upsilon,\mathsf{P}^{(0)}\prec\mathbf{M}}$" and "$\inf_{\mu\in\mathbb{R}}$" and simplify the expression:

$$\sup_{(\mathsf{p},\mathsf{P})\in Z_j}\left\{\mathcal{F}_n(\mathsf{p},\mathsf{P};M)\,:\,\mathcal{G}_n(\mathsf{p},\mathsf{P};A)=\alpha,\mathsf{p}^{(n)}=\mathsf{e}_a\right\}$$

$$=\sup_{\substack{\mathsf{P}^{(1:n-1)}\in \Upsilon^{n-1}\\ \forall i,\mathsf{P}^{(i)}\prec\mathbf{M}}}\inf_{\mu\in\mathbb{R}}\Bigg[-\mu\alpha+\mathsf{e}_a\left(\prod_{\ell=1}^{n-1}\mathsf{P}^{(n-\ell)}\right)\lambda^{(1)}(\mu)$$
$$-\sum_{i=1}^{n-1}\frac{d-1}{d^{i+1}}\mathsf{e}_a\left(\prod_{\ell=1}^{n-i-1}\mathsf{P}^{(n-\ell)}\right)D_{\mathrm{KL}}(\mathsf{P}^{(i)}||E)\Bigg].$$

We then proceed, by applying the minimax theorem recursively, to swap $\sup_{\mathsf{P}^{(i)}\in\Upsilon,\mathsf{P}^{(i)}\prec\mathbf{M}}$ and $\inf_{\mu\in\mathbb{R}}$ as before to derive the second equality.

(3) The proof relies on the following estimate: For any $\beta\in\mathbb{R}^{\mathcal{A}}_{>0}$,

$$\max_{\mathsf{q}\in\Gamma}\left|\sum_{a\in\mathcal{A}}\mathsf{q}_a\log\frac{\beta_a}{\mathsf{q}_a}\right|\leqslant\log|\mathcal{A}|+\max_a|\log\beta_a|,$$

which yields an absolute constant $C>0$ such that $|D_{\mathrm{KL}}(\mathsf{P}^{(i)}||E(\mu))|\leqslant C\cdot(|\mu|+1)$ and that, in particular, $\mathcal{F}_{n:m}(\mathsf{p},\mathsf{P};M)=\mathcal{F}_{n:m}(\mathsf{p},\mathsf{P};E(0))\leqslant Cd^{-n}$. Now given any $a\in\mathcal{A}_0$, express $\lambda^{(n)}(\mu)_a=\mathcal{F}_n(\mathsf{p},\mathsf{P};M)$ for some $(\mathsf{p},\mathsf{P})\in Z_n$ satisfying $\mathsf{p}^{(n)}=\mathsf{e}_a$. We then apply Lemma 3.8 to find some $(\mathsf{p}',\mathsf{P}')\in Z_n$ such that $(\mathsf{p}'^{(0:n)},\mathsf{P}'^{(0:n-1)})=(\mathsf{p}^{(0:n)},\mathsf{P}^{(0:n-1)})$ and $\mathsf{p}^{(n+pn_0)}=\mathsf{e}_{a_0}$. This implies

$$\lambda^{(n+pn_0)}(\mu)_{a_0}\geqslant\lambda^{(n)}(\mu)_a+d^{-n}\mathcal{F}_{n:n+pn_0}(\mathsf{p}',\mathsf{P}';M)\geqslant\lambda^{(n-pn_0)}(\mu)_{a_0}-Cd^{-n}.$$

For the other inequality, supposing $\lambda^{(n+pn_0)}(\mu)_{a_0}=\mathcal{F}_{n+pn_0}(\mathsf{p},\mathsf{P};M)$ for some $(\mathsf{p},\mathsf{P})\in Z_n$ with $\mathsf{p}^{(n+pn_0)}=\mathsf{e}_{a_0}$ yields

$$\max_{a\in\mathcal{A}_0}\lambda^{(n)}_a\geqslant\mathcal{F}_n(\mathsf{p},\mathsf{P};M)+\mathcal{F}_{n:n+pn_0}(\mathsf{p},\mathsf{P};M)-Cd^{-n}=\lambda^{(n+pn_0)}_{a_0}-Cd^{-n}.$$

Combining the above proves the proposed bound.

(4) The existence of the limit in Λ_j as well as the proposed equality is justified by (3), while its convexity follows from (1).

For Λ_j^*, we note that its convexity and lower semicontinuity are consequences of the Legendre transform of Λ_j (see, e.g., [12, Lemma 2.2.5 (a)]). We proceed to prove the first identity.

Note that by definition,

$$\sup_{(\mathsf{p},\mathsf{P})\in Z_j}\{\mathcal{F}(\mathsf{p},\mathsf{P};M) : \mathcal{G}(\mathsf{p},\mathsf{P};A)=\alpha\} = \sup_{(\mathsf{p},\mathsf{P})\in Z_j}\inf_{\mu\in\mathbb{R}}\left[-\mu\alpha+\Lambda_j(\mu)\right] \leqslant -\Lambda_j^*(\mu), \tag{13}$$

which is also known as the *weak duality* in the context of optimization theory. Hence, it remains to prove the other inequality. For simplicity, we write

$$f_j^{(n)}(\alpha) = \sup_{(\mathsf{p},\mathsf{P})\in Z_j(\alpha;A),\mathsf{p}^{(n)}=\mathsf{e}_{a_0}} \mathcal{F}_n(\mathsf{p},\mathsf{P};M) \quad \text{and} \quad f_j(\alpha) = \sup_{(\mathsf{p},\mathsf{P})\in Z_j(\alpha;A)} \mathcal{F}(\mathsf{p},\mathsf{P};M),$$

and take n_0 to be an integer as in Lemma 3.8.

First, we claim that there is a compact interval $\emptyset \neq [\alpha_1',\alpha_2'] \subseteq [\alpha_1,\alpha_2]$ such that $\sup_{(\mathsf{p},\mathsf{P})\in Z_j(\alpha;A)} \mathcal{F}(\mathsf{p},\mathsf{P};M) > -\infty$ if and only if $\alpha \in [\alpha_1',\alpha_2']$. The nonemptiness and compactness of the domain of f_j follows from the compactness of Z_j and the continuity of $\mathcal{F}$ and $\mathcal{G}$. Based on the observation, if

$$\alpha_1' := \mathcal{G}(\mathsf{p}',\mathsf{P}';A) < \mathcal{G}(\mathsf{p}'',\mathsf{P}'';A) =: \alpha_2'$$

are the extreme points of the domain of f_j with $(\mathsf{p}',\mathsf{P}'),(\mathsf{p}'',\mathsf{P}'') \in Z_j$, then for all $n\in\mathbb{N}$ there exist, by Lemma 3.8, approximations $(\mathsf{q}',\mathsf{Q}'),(\mathsf{q}'',\mathsf{Q}'')\in Z_j$ such that

$$\begin{cases} ((\mathsf{q}')^{(p(n+n_0)+j+1:\infty)},(\mathsf{Q}')^{(p(n+n_0)+j:\infty)}) = ((\mathsf{q}'')^{(p(n+n_0)+j+1:\infty)},(\mathsf{Q}'')^{(p(n+n_0)+j:\infty)}), \\ ((\mathsf{q}')^{(0:pn+j)},(\mathsf{Q}')^{(0:pn+j-1)}) = ((\mathsf{p}')^{(0:pn+j)},(\mathsf{P}')^{(0:pn+j-1)}), \\ ((\mathsf{q}'')^{(0:pn+j)},(\mathsf{Q}'')^{(0:pn+j-1)}) = ((\mathsf{p}'')^{(0:pn+j)},(\mathsf{P}'')^{(0:pn+j-1)}), \end{cases}$$

which, in turn, implies that for all sufficiently large n,

$$\mathcal{F}(\mathsf{q}',\mathsf{Q}';M) < \alpha_1' + Cd^{-pn-j} < \alpha_2' - Cd^{-pn-j} < \mathcal{F}(\mathsf{q}'',\mathsf{Q}'';M).$$

Since it is not hard to construct a continuous curve $\gamma : [0,1]\to Z_j$ connecting $(\mathsf{q}',\mathsf{Q}')$ and $(\mathsf{q}'',\mathsf{Q}'')\in Z_j$, it follows from the intermediate value theorem that

$$Z_j(\alpha;A)\neq\emptyset \text{ for all } \alpha\in[\alpha_1'+Cd^{-pn-j},\alpha_2'-Cd^{-pn-j}],$$

proving the claim when $n\to\infty$.

Next, we show that (13) holds for all $\alpha\notin[\alpha_1,\alpha_2]$, or equivalently, $[\alpha_1',\alpha_2'] = [\alpha_1,\alpha_2]$. Suppose $\alpha > \alpha_2'$ (similar for $\alpha<\alpha_1'$), so the uniform convergence of $\mathcal{G}_n(\mathsf{p},\mathsf{P};A)$ implies $f_j^{(pn+j)}(\alpha) = -\infty$ for all sufficiently large n. Therefore, by (2) and (3),

$$\begin{aligned} -\infty = f_j^{(p(n+n_0)+j)}(\alpha) &= \inf_{\mu\in\mathbb{R}}\left[-\mu\alpha + \lambda^{(p(n+n_0)+j)}(\mu)_{a_0}\right] \\ &\geqslant \inf_{\mu\in\mathbb{R}}\left[-\mu\alpha_0 + \max_{a\in\mathcal{A}_0}\lambda^{(pn+j)}(\mu)_a - Cd^{-(pn+j)}(|\mu|+1)\right] \\ &\geqslant \inf_{\alpha' : |\alpha'-\alpha|\leqslant 2Cd^{-(pn+j)}} -\Lambda_j^*(\alpha') - 2Cd^{-(pn+j)}. \end{aligned} \tag{14}$$

Since the estimate holds for all $\alpha>\alpha_1'$ and all large n, Λ_j^* is convex, and $\Lambda_j^*(\alpha_2')<\infty$ is finite, we deduce that $\Lambda_j^*(\alpha)=\infty$ for all $\alpha>\alpha_2'$, that is, $\alpha_2=\alpha_2'$.

Finally, it remains to show that (13) holds for $\alpha \in [\alpha_1, \alpha_2]$. The case $\alpha \in (\alpha_1, \alpha_2)$ is proved similarly as above. Let $\delta > 0$. By uniform convergence of $\mathcal{G}_n$, $f_j^{(pn+j)}(\alpha) < \infty$ for all sufficiently large n. It follows from (3) that for all large n,

$$\begin{aligned}\sup_{\alpha' : |\alpha'-\alpha|<\delta} f_j(\alpha') &\geqslant f_j^{(p(n+n_0)+j)}(\alpha) - Cd^{-(p(n+n_0)+j)} \\ &= \inf_{\mu\in\mathbb{R}} \left[-\mu\alpha + \lambda^{(p(n+n_0)+j)}(\mu)_{a_0} - Cd^{-(p(n+n_0)+j)}\right] \\ &\geqslant \inf_{\mu\in\mathbb{R}} \left[-\mu\alpha + \max_{a\in\mathcal{A}_0}\left[\lambda^{(pn+j)}(\mu)_a - 2Cd^{-pn+j}(|\mu|+1)\right]\right] \qquad (15) \\ &\geqslant \inf_{\mu\in\mathbb{R}} \left[-\mu\alpha + \Lambda_j(\mu) - 3Cd^{-(pn+j)}|\mu|\right] - 3Cd^{-(pn+j)} \\ &\geqslant \inf_{\alpha' : |\alpha'-\alpha|\leqslant 3Cd^{-(pn+j)}} -\Lambda_j^*(\alpha') - 3Cd^{-(pn+j)}.\end{aligned}$$

Since f_j and $-\Lambda_j^*$ are continuous on (α_1, α_2), the equality (13) is proved by first letting $n \to \infty$ and then $\delta \to 0$. If $\alpha \in \{\alpha_1, \alpha_2\}$, the discussion is divided into the following two cases. If $\alpha_1 = \alpha_2$, then $f_j(\alpha_1) = 0$ as $\max_{(\mathsf{p},\mathsf{P})\in Z_j} \mathcal{F}(\mathsf{p}, \mathsf{P}; M)$ has maximum 0 by nonnegativity of Kullback–Leibler divergence, and $0 = f_j(\alpha_1) \geqslant -\Lambda_j^*(\alpha_1) \geqslant \Lambda_j^*(0) = 0$. If $\alpha_1 < \alpha_2$, it suffices to show that f_j is continuous at α_1 and α_2, since under the circumstances both f_j and $-\Lambda_j^*$ are continuous at $[\alpha_1, \alpha_2]$ and therefore coincide. This follows similarly from the intermediate value theorem as before. □

Proof of Theorem 1.2. For conciseness, we prove only the case $j = 0$. Observing that the ratio $\frac{d-1}{d^{n-i+1}} / \frac{|\Xi_i|}{|\Delta_n|} = 1 - d^{-n-1}$ is uniform for $0 \leqslant i \leqslant n$, we write

$$\widetilde{\mathcal{F}}_n(\mathsf{p}, \mathsf{P}; M) := -\sum_{i=0}^{n-1} \frac{|\Xi_{n-i}|}{|\Delta_n|} \mathsf{p}^{(i+1)} D_{\mathrm{KL}}(\mathsf{P}^{(i)}||M) = \frac{\mathcal{F}_n(\mathsf{p}, \mathsf{P}; M)}{1 - d^{-n-1}},$$

$$\widetilde{\mathcal{G}}_n(\mathsf{p}, \mathsf{P}; A) := \sum_{i=0}^{n-1} \frac{|\Xi_{n-i}|}{|\Delta_n|} \mathsf{p}^{(i+1)} (\mathsf{P}^{(i)} \odot \log A)\mathbb{1} = \frac{\mathcal{G}_n(\mathsf{p}, \mathsf{P}; A)}{1 - d^{-n-1}}.$$

Notably, the sample mean $Y_{pn} = \widetilde{\mathcal{G}}_n(\mathsf{p}, \mathsf{P}; A)$ if the Markov chain $X = (X_g)_{g\in T}$ satisfies $(\tau_{0:n}(X), \eta_{0:n-1}(X)) = (\overleftarrow{\mathsf{p}^{(0:n)}}, \overleftarrow{\mathsf{P}^{(0:n-1)}})$, which follows from a combinatorial argument: In level i $(0 \leqslant i < pn)$, there exist $|\Xi_i|\tau_i(X)_a = |\Xi_i|\mathsf{p}_a^{(n-i)}$ vertices associated with state a, which are responsible for $(d \cdot |\Xi_i|\tau_i(X)_a) \cdot \eta_i(X)_{a,b} = |\Xi_{i+1}|\mathsf{p}_a^{(n-i)}\mathsf{P}_{a,b}^{(n-i-1)}$ transitions from state a to b between level i and $i + 1$.

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

$$\begin{aligned}&\frac{1}{|\Delta_{pn}|} \log \mathbb{P}\Big(Y_{pn} \in S \Big| X_\epsilon = a_0\Big) \\ &= \sup_{\alpha\in S} \left\{ \frac{1}{|\Delta_{pn}|} \log \mathbb{P}\big(\widetilde{\mathcal{G}}_{pn}(\mathsf{p}, \mathsf{P}; A) = \alpha\big) \;:\; \mathsf{p}^{(pn)} = \mathsf{e}_{a_0},\right.\end{aligned}$$

$$\left.\overleftarrow{(\mathsf{p}^{(0:pn)}}, \overleftarrow{\mathsf{P}^{(0:pn-1)}}) = (\tau_{0:pn}(X), \eta_{0:pn-1}(X))\right\} + O\left(\frac{\log|\Delta_{pn}|}{|\Delta_{pn}|}\right)$$

$$= \sup_{\alpha\in S}\left\{ \widetilde{\mathcal{F}}_n(\mathsf{p},\mathsf{P};M) \,:\, \mathsf{p}^{(pn)} = \mathsf{e}_{a_0}, (\overleftarrow{\mathsf{p}^{(0:pn)}}, \overleftarrow{\mathsf{P}^{(0:pn-1)}}) = (\tau_{0:pn}(X), \eta_{0:pn-1}(X)),\right.$$

$$\left.\widetilde{\mathcal{G}}_n(\mathsf{p},\mathsf{P};A) = \alpha\right\} + O\left(\frac{\log|\Delta_{pn}|}{|\Delta_{pn}|}\right), \tag{16}$$

To prove the upper bound of the theorem, we consider the closed $(1/k)$-neighborhood $S_k = \{\alpha' \in \mathbb{R} : \inf_{\alpha\in S}|\alpha'-\alpha| \leqslant 1/k\}$ of S. Taking advantage of the uniform convergence and the uniform boundedness of $\mathcal{F}_n$ and $\mathcal{G}_n$, we can bound (16), for sufficiently large n, from above by

$$\sup_{\alpha\in S_k}\{\mathcal{F}(\mathsf{p},\mathsf{P};M) \,:\, (\mathsf{p},\mathsf{P}) \in Z_0(\alpha;A)\} + o(1),$$

which, by Lemma 3.9 (4), coincides further with

$$\sup_{\alpha\in S_k}\inf_{\mu\in\mathbb{R}}[-\mu\alpha + \Lambda_0(\mu)] + o(1).$$

The desired inequality is obtained by first letting $n\to\infty$ and then $k\to\infty$. For the lower bound, we analogously take $S_k = \{\alpha' \in \mathbb{R} : \inf_{\alpha\notin S}|\alpha'-\alpha| \geqslant 1/k\}$. By Lemma 3.4 and the equicontinuity of $\mathcal{F}_n$ and $\mathcal{G}_n$, one derives the following uniform lower bound for (16):

$$\sup_{\alpha\in S_k}\left\{\frac{\mathcal{F}(\mathsf{p},\mathsf{P};M)}{1-d^{-pn-1}} \,:\, (\mathsf{p},\mathsf{P}) \in Z^{(n)}((1-d^{-pn-1})\alpha;A), \mathsf{p}^{(n)} = \mathsf{e}_{a_0}\right\} + o(1),$$

Letting $n\to\infty$, the lower limit of the above (possibly $-\infty$) is further bounded from below by

$$\sup_{\alpha\in S_{k+1}}\{\mathcal{F}(\mathsf{p},\mathsf{P};M) \,:\, (\mathsf{p},\mathsf{P}) \in Z_0(\alpha;A)\},$$

or, equivalently by Lemma 3.9 (4),

$$\sup_{\alpha\in S_{k+1}}\inf_{\mu\in\mathbb{R}}[-\mu\alpha + \Lambda_0(\mu)].$$

The proof is concluded by letting $k\to\infty$. □

As a special case, the following corollary simplifies the expression of Theorem 1.2 when focusing on intervals.

Corollary 3.10. *Suppose that X, Y, and $\Lambda_j^*(\alpha)$ are as in Theorem 1.2. Then,*

$$\lim_{\varepsilon\to 0^+}\lim_{n\to\infty}\frac{1}{|\Delta_{pn+j}|}\log\mathbb{P}\Big(Y_{pn+j} \in (\alpha-\varepsilon,\alpha+\varepsilon)\Big|X_\epsilon = a_0\Big) = -\Lambda_j^*(\alpha).$$

Proof. It is a consequence of Theorem 1.2 with the existence of the limit justified by continuity $\Lambda_j^*(\alpha)$, a consequence of convexity and lower semicontinuity. □

Proof of Theorem 1.3. For simplicity, we only give the proof of the case $j = 0$.

To begin, we note that the existence and uniqueness of $(\pi^{(i)})_{i\in\mathbb{Z}}$ is a consequence of the Perron–Frobenius theorem (see, e.g., [16, Theorems 8.2.8 and 8.4.4]). Precisely, according to the characterization (7), we can express M as a superdiagonal block matrix $M = (M_{i,j})_{0\leqslant i,j\leqslant p-1}$ with each $M_{i,j} \in \mathbb{R}_{\geqslant 0}^{\mathcal{A}_i\times\mathcal{A}_j}$, and M^p is a diagonal block matrix whose diagonal elements are primitive matrices. The existence and uniqueness follows immediately from the cited theorem.

We first show that any maximizer (p,P) of $\max_{(\mathsf{p},\mathsf{P})\in Z_0} \mathcal{F}(\mathsf{p},\mathsf{P};M)$ satisfies

$$\mathcal{G}(\mathsf{p},\mathsf{P};A) = \mathcal{G}(\mathsf{p}^*,\mathsf{P}^*;A) \tag{17}$$

where

$$(\mathsf{p}^*,\mathsf{P}^*) = ((\pi^{(0)},\pi^{(1)},\dots,\pi^{(p-1)})^{\mathbb{Z}_+},(M)^{\mathbb{Z}_+}) \in Z_0.$$

Notably, $(\mathsf{p}^*,\mathsf{P}^*)$ is indeed a maximizer: For every $(\mathsf{q},\mathsf{Q}) \in Z_0$,

$$D_{\mathrm{KL}}(\mathsf{Q}^{(i)}||M) \geqslant 0 \quad \text{and} \quad D_{\mathrm{KL}}(\mathsf{Q}^{(i)}||M) = 0 \text{ iff } \mathsf{Q}^{(i)} = M, \tag{18}$$

from which the maximality follows. To demonstrate (17), we apply (18) to derive that for any optimal solution (p,P),

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

Recall that M^p is a diagonal block matrix whose diagonal elements are primitive. The Perron–Frobenius theorem guarantees that

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

Since $\mathsf{p}_a^{(pn+i)} = \pi_a^{(i)}$ is positive if and only if $a \in \mathcal{A}_{-i}$ and $n, i \geqslant 0$, once again we can apply (18) to derive $\mathsf{P}_{a,b}^{(pn+i)} = M_{a,b}$ if $(a,b) \in \mathcal{A}_{-i-1}\times\mathcal{A}_{-i}$, and the claim is thus proved. Equivalently, by Lemma 3.9, our claim implies $-\Lambda_0^*(\alpha)$ admits a unique maximum point

$$\alpha = \alpha^* := \mathcal{G}(\mathsf{p}^*,\mathsf{P}^*;A) = \sum_{i=0}^{p-1} \frac{d^{-i}}{\sum_{\ell=0}^{p-1} d^{-\ell}} \sum_{a,b\in\mathcal{A}} \pi_a^{(i+1)} M_{a,b} \log A_{a,b},$$

at which $-\Lambda_0^*$ attains its maximum 0.

The rest of the proof is an application of the Borel–Cantelli lemma. Observe that $\delta(\varepsilon) = \sup_{|\alpha-\alpha^*|\geqslant\varepsilon} -\Lambda_0^*(\alpha) < 0$ for all $\varepsilon > 0$ due to the uniqueness of α^*. This secures the existence of $N \in \mathbb{N}$ such that for all $n \geqslant N$,

$$\mathbb{P}\Big(\Big|Y_{pn} - \alpha^*\Big| > \varepsilon \Big| X_\epsilon = a_0\Big) \leqslant e^{\delta(\varepsilon)|\Delta_{pn}|}$$

Summing over n, we obtain

$$\sum_{n=N}^{\infty} \mathbb{P}\Big(\Big|Y_{pn} - \alpha^*\Big| > \varepsilon \Big| X_\epsilon = a_0\Big) \leqslant \sum_{n=N}^{\infty} e^{\delta(\varepsilon)|\Delta_{pn}|} < \infty,$$

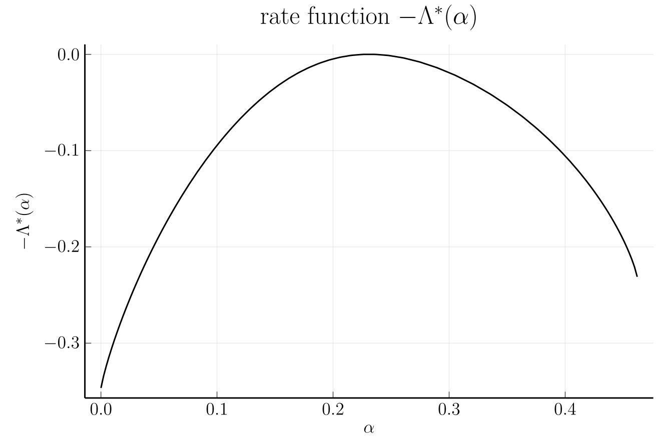

FIGURE 1 Rate function Λ_0^*.

which, by the Borel–Cantelli lemma, implies

$$\mathbb{P}\left(\limsup_{n\to\infty}\{\left|Y_{pn}-\alpha^*\right|>\varepsilon\}\middle|X_\epsilon=a_0\right)=0.$$

This completes the proof since $\varepsilon>0$ is arbitrary. □

Example 3.11. Let $d=2$ and A, M be matrices defined as

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

Since M is primitive, Theorem 1.2 ensures that the averages

$$Y_n\to\sum_{a,b\in\mathcal{A}}\pi_a M_{a,b}\log A_{a,b} \quad \mathbb{P}\text{-a.s.}$$

and satisfy the large deviation principle with rate $\Lambda_0^*(\alpha)$. Numerically, $-\Lambda_0^*(\alpha)$ is computed and depicted in 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 $\alpha=\frac{1}{3}\log 2\approx 0.2310$.

In the following, we present yet another example of irreducible M in order to (a) demonstrate the upper and lower limit of the conditional sample mean might differ, and (b) compare the conditional sample mean, unconditional sample mean, and their expectations.

Example 3.12. We consider $d=2$ and

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

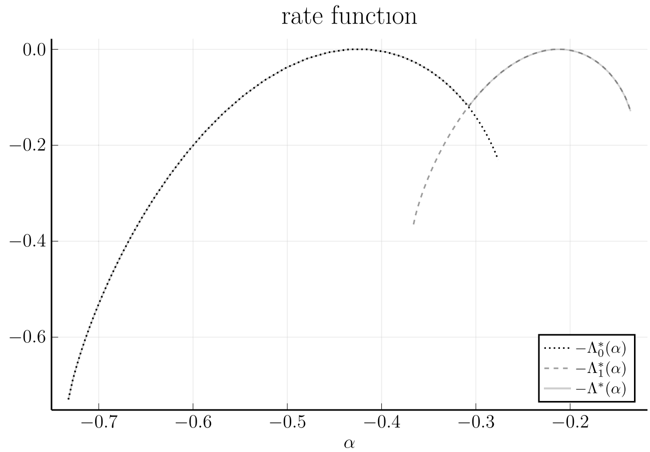

FIGURE 2 Rate function Λ_0^*, Λ_1^*, and Λ^*.

for which the conditional sample mean associated with $\mathcal{A}_0 := \{0\}$ and on $\mathcal{A}_1 := \{1, 2\}$ are plotted in Figure 2, respectively. To address (a), let α_0^* and α_1^* be the maximum points of Λ_0^* and Λ_1^*, respectively. Theorem 1.2 then guarantees that almost surely

$$\alpha^- := \liminf_{n\to\infty} Y_n = \alpha_1^* \quad \text{and} \quad \alpha^+ := \limsup_{n\to\infty} Y_n = \alpha_0^*.$$

Regarding (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 attains its maximum at α_0^* and α_1^*. For the expectation of sample mean, we deduce by the dominated convergence theorem that

$$\beta^- := \liminf_{n\to\infty} \mathbb{E}(Y_n) = \min_{0\leqslant i\leqslant 1} \sum_{j=0}^{1} \sum_{a\in\mathcal{A}_j} \pi_a \alpha_{i+j}^*,$$

$$\beta^+ := \limsup_{n\to\infty} \mathbb{E}(Y_n) = \max_{0\leqslant i\leqslant 1} \sum_{j=0}^{1} \sum_{a\in\mathcal{A}_j} \pi_a \alpha_{i+j}^*.$$

Notably, $\alpha^- < \beta^- < \beta^+ < \alpha^+$, as opposed to the case when M is primitive and the four numbers coincide. This indicates that the conditional sample mean provides more information than the unconditional one or 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 observe that

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

and that for every $t \in \mathcal{T}_{\mathbf{M}}(= \mathrm{supp}\mathbb{P})$, the sample mean can be explicitly calculated as

$$Y_n(t) \in \left\{\alpha^- + o(1), \alpha^+ + o(1)\right\}.$$

Nevertheless, it is interesting yet unsurprising that not a single labeled tree $t \in \mathcal{T}_{\mathbf{M}}$ admits a sample mean approaching $\beta^+ = \beta^-$.

4 | HAUSDORFF DIMENSIONS OF IRREDUCIBLE MARKOV HOM TREE-SHIFTS

The section is dedicated to presenting the Hausdorff dimension formula for Markov hom tree-shifts $\mathcal{T}_{\mathbf{M}} \subset \mathcal{A}^T$ when $\mathbf{M}$ is an irreducible matrix. We begin by considering the following metric on $\mathcal{A}^T$:

$$D(x, y) = e^{-\sup\{|\Delta_n| : x|_{\Delta_n} = y|_{\Delta_n}\}}. \tag{19}$$

Notably, this metric coincides with the canonical metric when $T = \mathbb{Z}_+$ is degenerate and the choice is due to the its intimate relation with topological entropy of tree-shifts, which was considered by Petersen and Salama [24, 25] as a natural generalization of its counterpart for one-sided subshifts. See also [26] for related results for asymptotic pressure of tree-shifts, a generalized form of the topological entropy. Specifically, the *topological entropy* of a Markov hom tree-shift $\mathcal{T}_{\mathbf{M}}$ is defined by

$$h_{top}(\mathcal{T}_{\mathbf{M}}) := \limsup_{n\to\infty} \frac{\log |B_n(\mathcal{T}_{\mathbf{M}})|}{|\Delta_n|},$$

where the sequence actually converges as established in the aforementioned work. Using this definition and the existence of the limit, we observe the following equalities regarding box-counting dimensions:

$$\overline{\dim}_B \mathcal{T}_{\mathbf{M}} := \limsup_{r\to\infty} -\frac{\log \mathcal{N}_r(\mathcal{T}_{\mathbf{M}})}{\log r} = d \cdot h_{top}(\mathcal{T}_{\mathbf{M}}),$$

$$\underline{\dim}_B \mathcal{T}_{\mathbf{M}} := \liminf_{r\to\infty} -\frac{\log \mathcal{N}_r(\mathcal{T}_{\mathbf{M}})}{\log r} = h_{top}(\mathcal{T}_{\mathbf{M}}),$$

where $\mathcal{N}_r(\mathcal{T}_{\mathbf{M}})$ denotes the minimal number of closed r-balls needed to cover the set $\mathcal{T}_{\mathbf{M}}$. It is worth mentioning that the multiplicative difference by d arises from the fact that closed r-balls $B_r(t)$ are aliases of the same set for all $r \in [e^{-|\Delta_n|}, e^{-|\Delta_{n-1}|}) = [e^{-\frac{d^{n+1}-1}{d-1}}, e^{-\frac{d^n-1}{d-1}})$, causing $\mathcal{N}_r(\mathcal{T}_{\mathbf{M}})$ to remain constant over the interval. As a natural follow-up question, it would be interesting to determine the packing and Hausdorff dimensions of the Markov hom tree-shifts, where the former is a known to coincide with $\overline{\dim}_B \mathcal{T}_{\mathbf{M}}$ when $\mathbf{M}$ is irreducible (see Appendix B) and the latter is our primary focus of this section. As mentioned in the introduction, this goal is achieved by finding a nonlinear transfer operator to link the Hausdorff dimension of $\mathcal{T}_{\mathbf{M}}$ to the eigenvalue of the operator.

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) mappings and reproduces several classical results regarding nonnegative primitive/irreducible matrix transformations. In this work, we will consider such an analysis under the framework of [19], of which the setting can be described as follows. Let $f : \mathbb{R}^{\mathcal{A}}_{\geqslant 0} \to \mathbb{R}^{\mathcal{A}}_{\geqslant 0}$ be a continuous function. It is said to be *order-preserving* if $f(x) \leqslant f(y)$ for every $x \leqslant y$, and it is called *homogeneous* if $f(\alpha x) = \alpha f(x)$ for all $\alpha \in \mathbb{R}_{\geqslant 0}$, 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 $\mathbb{R}^{\mathcal{A}}_{>0}$ (see Definition 3.6), where log and exp are applied entrywise as explained in Section 2. Within this framework, we establish that the operator $\mathcal{L}_{\mathbf{M},r}$ (recall (5)) 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. *Suppose* $\mathbf{M}$*,* $\mathcal{L}_{\mathbf{M},r}$*, and* $r \in \mathcal{R}_{p,d}$ *be as in Theorem 1.4. Then, the following properties hold.*

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

(2) *Suppose* $\mathcal{L}_{\mathbf{M},r}$ *maps* $C' = \{x \in \mathbb{R}^{\mathcal{A}}_{\geqslant 0} : x_a = 0 \text{ if } a \notin \mathcal{A}'\}$ *into itself for some* $\mathcal{A}' \subseteq \mathcal{A}$*. Then, there exists an eigenvector* $v \in C'$ *associated with eigenvalue* $\rho(\mathcal{L}_{\mathbf{M},r}|_{C'})$ *such that*

$$\limsup_{n\to\infty} \left\| \mathcal{L}^n_{\mathbf{M},r}(w) \right\|^{1/n} \leqslant \rho(\mathcal{L}_{\mathbf{M},r}|_{C'}) \text{ for all } w \in C'.$$

In particular, if $\mathbf{M}$ *is irreducible, for each* $\mathcal{A}_j \in \mathcal{P}(\mathbf{M})$*, there exists a unique (up to scaling) eigenvector* $v^{(j)} \in C_j$*, whose associated eigenvalue* $\rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{M},r})$ *satisfying the stated property on* C_j*, and* $v^{(j)}_a > 0$ *for all* $\mathcal{A}_j$*.*

(3) *If* $\mathbf{M}$ *is irreducible, then for every* $x \in C_j^+ := \{x \in C_j : x_a > 0 \text{ if } a \in \mathcal{A}_j\}$*, there exists* $0 < \theta < 1$ *such that*

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

where $v^{(j)} \in C_j^+$ *is the unique eigenvector as in (2).*

Proof.

(1) The first four conditions follow from a routine check of definition, while the multiplicative convexity follows from Lemma 3.7 applied to $\Psi_{\mathbf{M},s}$ $(s > 0)$. which extends naturally to $\mathcal{L}_{\mathbf{M},r} = \Psi_{\mathbf{M},r_{p-1}} \circ \cdots \circ \Psi_{\mathbf{M},r_0}$.

(2) According to [19, Corollary 5.4.2], $\mathcal{L}_{\mathbf{M},r}$ admits an eigenvector $v \in C'$ such that $\mathcal{L}_{\mathbf{M},r}(v) = \rho(\mathcal{L}_{\mathbf{M},r}|_{C'}) \cdot v$, while [19, Theorem 5.3.1] asserts that $\limsup_{n\to\infty} \left\|\mathcal{L}_{\mathbf{M},r}(w)\right\|^{1/n} \leqslant \rho(\mathcal{L}_{\mathbf{M},r}|_{C'})$ for all $w \in C'$. We observe that for any $x \in \mathbb{R}^{\mathcal{A}}_{\geqslant 0}$, $\Psi_{\mathbf{M},s}(x)_a > 0$ if and only if $(\mathbf{M}x)_a > 0$. Recursively applying this property, one deduces that $\mathcal{L}_{\mathbf{M},r}(x)_a > 0$ if and only if $(\mathbf{M}^p x)_a > 0$. Now if $\mathbf{M}$ is irreducible, $\mathcal{P}(\mathbf{M})$ is a partition and $\mathbf{M}_{a,b} > 0$ if and only $(a, b) \in \mathcal{A}_j \times \mathcal{A}_{j+1}$ for some $j \in [p]$. Notably, $\mathcal{L}_{\mathbf{M},r}$ maps C_j to C_j, each of which hence admits an eigenvector $v^{(j)} \in C_j$.

Moreover, $v_a^{(j)} > 0$ for all $a \in \mathcal{A}_j$ as every state a plays the role of a_0 in (A1). By denoting the natural inclusion as $\iota_j : \mathcal{C}_j \to \mathbb{R}_{\geqslant 0}^{\mathcal{A}}$, we have that $\iota_j^{-1} \circ \mathcal{L}_{\mathbf{M},r} \circ \iota_j$ maps $\mathbb{R}_{\geqslant 0}^{\mathcal{A}_j}$ into $\mathbb{R}_{\geqslant 0}^{\mathcal{A}_j}$ and the derivative of $\iota_j^{-1} \circ \mathcal{L}_{\mathbf{M},r} \circ \iota_j$ at $\iota_j^{-1}(v^{(j)})$ is primitive. Therefore, by [19, 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 $\mathcal{L}_{\mathbf{M},r}$ in (1) are inherited by $\iota_j^{-1} \circ \mathcal{L}_{\mathbf{M},r} \circ \iota_j$. Hence, the existence of θ follows as a consequence of [20, Theorem 4.7]. □

4.2 | Some lemmas

For conciseness, we suppress, unless mentioned otherwise, $\mathbf{M}$ from $\mathcal{F}_{n:m}(\mathsf{p}, \mathsf{P}; \mathbf{M})$ for the rest of the paper. We first state a number of technical lemmas, whose proofs are postponed until the appendix. These lemmas establish the desired minimax properties, bridging the upper bound of the Hausdorff dimension obtained through efficient covering with the lower bound obtained via the mass distribution principle, as discussed in the coming subsections.

Lemma 4.2. *The following minimax property holds.*

$$\max_{\substack{(\mathsf{p},\mathsf{P}) \in Z_0: \\ p\text{-periodic}}} \min_{s \in \Gamma_{[p]}} \sum_{m=0}^{p-1} s_m \mathcal{F}_{m:\infty}(\mathsf{p}, \mathsf{P}) = \min_{s \in \Gamma_{[p]}} \max_{\substack{(\mathsf{p},\mathsf{P}) \in Z_0: \\ p\text{-periodic}}} \sum_{m=0}^{p-1} s_m \mathcal{F}_{m:\infty}(\mathsf{p}, \mathsf{P}). \tag{20}$$

The proof of Lemma 4.2 is presented in Appendix A.1.

Remark 4.3. We should note that $\mathcal{P}(\mathbf{M})$ is a partition if $\mathbf{M}$ is irreducible. Under the circumstances, if (p, P) is a p-periodic optimizer in Lemma 4.2, then so is $(\mathsf{p}, \mathsf{P}^*)$, where

$$\mathsf{P}^{*(i)}_{a,b} = \mathsf{P}^{*(0)}_{a,b} := \mathsf{P}^{(j)}_{a,b} \quad \text{for all } (a, b) \in \mathcal{A}_{-j-1} \times \mathcal{A}_{-j} \text{ and } i \in [p].$$

Therefore, the value of (A.1) remains unchanged even with every appearance of "$\max_{\substack{\mathsf{P} \in \Upsilon^p \\ \mathsf{p}^{(p)} \in \mathcal{C}_0}}$" 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 sequence before we present our quintessential lemma.

$$t^*_{n,N} = \begin{cases} \left(N + \frac{1}{d^p - 1}\right)^{-1} \cdot \frac{d^p}{d^p - 1} & \text{if } n = 0; \\ \left(N + \frac{1}{d^p - 1}\right)^{-1} & \text{if } p \mid n \text{ and } 0 < n < pN; \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

Lemma 4.4. *Suppose that* $\mathbf{M}$ *is irreducible and* $\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M},r})$ *is as in Proposition 4.1. The following minimax properties hold.*

$$\min_{r \in \mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1} \prod_{i=0}^{\ell} r_i^{-1} \right)^{-1} \cdot \log \rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M},r}) \tag{22}$$

$$= \lim_{N\to\infty} \min_{s\in\Gamma_{[p]}} \max_{(\mathsf{p},\mathsf{P})\in Z_0} \sum_{m=0}^{p-1} s_m \sum_{n=0}^{pN-1} t^*_{n,N} \mathcal{F}_{n+m:\infty}(\mathsf{p},\mathsf{P}) \tag{23}$$

$$= \min_{s\in\Gamma_{[p]}} \max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ \mathsf{P} \text{ is 1-periodic}\\ \mathsf{p} \text{ is } p\text{-periodic}}} \sum_{m=0}^{p-1} s_m \mathcal{F}_{m:\infty}(\mathsf{p},\mathsf{P}) \tag{24}$$

$$= \max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ \mathsf{P} \text{ is 1-periodic}\\ \mathsf{p} \text{ is } p\text{-periodic}}} \min_{s\in\Gamma_{[p]}} \sum_{m=0}^{p-1} s_m \mathcal{F}_{m:\infty}(\mathsf{p},\mathsf{P}). \tag{25}$$

Furthermore, any maximizer $(\mathsf{p}^*,\mathsf{P}^*)$ *of (25) can be chosen such that* $\mathsf{P}^{(0)}$ *is irreducible and* $\mathsf{p}^{(0)^*}$ *is the unique left eigenvector of* $(\mathsf{P}^{(0)^*})^p$ *in* $\mathcal{C}_0$.

The proof of Lemma 4.4 is presented in Appendix A.2.

Remark 4.5. It can be seen, for example, from the cyclic symmetry of s in (24) that

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

is actually independent of j and therefore further coincides with

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

where ρ_* is defined in Theorem 1.4.

As a consequence of Lemma 4.4, we have the following corollary.

Corollary 4.6. *There exists a Markov measure* $\mathbb{P}$ *associated with an irreducible transition matrix* M *and an initial distribution* π *such that it is supported in* $\mathcal{T}_{\mathbf{M}}$ *and that, almost surely,*

$$\liminf_{n\to\infty} -\frac{\log \mathbb{P}([X|_{\Delta_n}])}{|\Delta_n|} = \min_{r\in\mathcal{R}_{p,d}} \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{-1}\right)^{-1} \cdot \log \rho_{*}(\mathcal{L}_{\mathbf{M},r}), \tag{26}$$

where $[u]$ *denotes the cylinder set associated with the block* u*, namely, given* $u \in B_n(\mathcal{T}_{\mathbf{M}})$*,*

$$[u] := \{t \in \mathcal{T}_{\mathbf{M}} : t_g = u_g, \forall g \in \Delta_n\}. \tag{27}$$

Proof. It straightforwardly follows from Lemma 4.4 and Theorem 1.3 by choosing $A = M$. More precisely, we may (uniquely) choose the initial distribution as in Lemma 4.4 so that the initial distribution of a state a is positive if and only if $a \in \mathcal{A}_0$. The proof is then concluded by Theorem 1.3. □

4.3 | Lower bound when M is irreducible

Suppose that $\mathcal{T}_{\mathbf{M}}$ is a tree-shift with an irreducible incidence matrix $\mathbf{M}$ and that $\mathbb{P}$ is the Markov measure in Corollary 4.6. The lower bound is obtained by applying the mass distribution principle: Due to (26), for every $\delta < \min_{r\in\mathcal{R}_{p,d}}(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{-1})^{-1}\cdot\log\rho_*(\mathcal{L}_{\mathbf{M},r})$, we have

$$\limsup_{n\to\infty}\frac{\mathbb{P}([X|_{\Delta_n}])}{e^{-\delta|\Delta_n|}} = 0 \quad \text{almost surely.}$$

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

$$\mathbb{P}([t|_{\Delta_n}]) \leqslant Ce^{-\delta|\Delta_n|} \text{ for all } n\in\mathbb{N}, t\in S.$$

Consequently, supposing that $\mathcal{S}$ is a cover of S consisting of disjoint cylinder sets, we have

$$\sum_{[u]\in\mathcal{S}}(\mathrm{diam}([u]))^{\delta} \geqslant C^{-1}\sum_{[u]\in\mathcal{S}}\mathbb{P}([u]) \geqslant C^{-1}\mathbb{P}(S) > 0.$$

Therefore, the Hausdorff measure $H^{\delta}(S) \geqslant C^{-1}\mathbb{P}(S) > 0$ is bounded from below and $\dim_H \mathcal{T}_{\mathbf{M}} \geqslant \dim_H S \geqslant \delta$, implying $\dim_H \mathcal{T}_{\mathbf{M}} \geqslant \min_{s\in\mathcal{R}_{p,d}}(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{-1})^{-1}\cdot\log\rho_*(\mathcal{L}_{\mathbf{M},r})$.

4.4 | Upper bound when M is irreducible

The upper bound of the Hausdorff dimension is proved by constructing suitable covers. As will be seen shortly, the study of the operator $\mathcal{L}_{\mathbf{M},r}$ is motivated by the following naïve covering strategy for the space $\mathcal{T}_{\mathbf{M}}$. For convenience, we denote the collection of all k-cylinder sets by $\mathcal{C}_k := \{[t|_{\Delta_n}] : t\in\mathcal{T}_{\mathbf{M}}\}$.

Suppose $n + N > n \gg 0$ ($N\in\mathbb{N}$). We define a cover $\mathcal{S}_{n,N}$ agreeing with the following philosophy.

(a) The cover $\mathcal{S}_{n,N}$ consists of cylinder sets in $\cup_{k=n}^{n+N}\mathcal{C}_k$.
(b) A cylinder set $[t|_{\Delta_k}]$ is contained in $\mathcal{S}_{n,N}$ if

$$\frac{\log|B_k(\tau(t),\eta(t))|}{|\Delta_k|} = \min_{n\leqslant m\leqslant n+N}\frac{\log|B_m(\tau(t),\eta(t))|}{|\Delta_m|}. \tag{28}$$

Condition (a) is based on the belief that the Hausdorff dimension is approached by the intermediate dimension (see [4]) and condition (b) is imposed in the hope that the upper bound of the Hausdorff dimension is minimized. Precisely, if we denote

$$\alpha_{n,N} := \max_{(\mathrm{q},\mathrm{Q})\in W_{n+N}}\min_{n\leqslant m\leqslant n+N}\frac{\log|B_m(\mathrm{q},\mathrm{Q})|}{|\Delta_m|}, \tag{29}$$

and assume $\alpha > \limsup_{N\to\infty} \limsup_{n\to\infty} \alpha_{n,N}$, then the cover $\mathcal{S}_{n,N}$ gives the following upper bound for the α-Hausdorff measure of $\mathcal{T}_{\mathbf{M}}$:

$$\begin{aligned} H^s(\mathcal{T}_{\mathbf{M}}) &\leqslant \limsup_{N\to\infty}\limsup_{n\to\infty} \sum_{k=n}^{n+N} \sum_{[u]\in\mathcal{C}_k\cap\mathcal{S}_{n,N}} (\mathrm{diam}([u]))^\alpha \\ &= \limsup_{N\to\infty}\limsup_{n\to\infty} \sum_{k=n}^{n+N} \sum_{(\mathrm{q},\mathrm{Q})\in W_{n+N}} \sum_{\substack{[u]\in\mathcal{C}_k\cap\mathcal{S}_{n,N}:\\ u\in B_k(\mathrm{q}^{(0:k)},\mathrm{Q}^{(0:k-1)})}} e^{-\alpha|\Delta_k|} \\ &\leqslant \limsup_{N\to\infty}\limsup_{n\to\infty} \sum_{k=n}^{n+N} |W_{n+N}| e^{-(\alpha-\alpha_{n,N})|\Delta_k|}. \end{aligned}$$

Now that Lemma 3.3 asserts

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

we deduce that $\mathcal{T}_{\mathbf{M}}$ is s-Hausdorff null:

$$H^s(\mathcal{T}_{\mathbf{M}}) \leqslant \limsup_{N\to\infty}\limsup_{n\to\infty} \sum_{k=n}^{n+N} e^{-\left(\alpha-\alpha_{n,N}-O\left(\frac{\log|\Delta_{n+N}|}{|\Delta_n|}\right)\right)|\Delta_k|} = 0,$$

which implies $\dim_H \mathcal{T}_{\mathbf{M}} \leqslant \alpha$.

In practice, obtaining an asymptotic estimate for $\alpha_{n,N}$ is facilitated by the following simplification. For each $n, N \in \mathbb{N}$, decompose W_{n+N} into subsets

$$W_{n+N}{}^{(j)} := \{(\mathrm{q},\mathrm{Q}) \in W_{n+N} : \mathrm{q}^{(0)} \in \mathcal{C}_{j-n-N}\} \text{ for } j \in [p].$$

Since it is a finite partition, we may assume, by a proper choice of a_0 in assumption (A1), that the maximum of $\alpha_{n,N}$ is attained in $W_{n+N}{}^{(0)}$ for infinitely many $n, N \in \mathbb{N}$, resulting in

$$\limsup_{N\to\infty}\limsup_{n\to\infty} \alpha_{n,N} = \limsup_{N\to\infty}\limsup_{n\to\infty} \max_{(\mathrm{q},\mathrm{Q})\in W_{n+N}{}^{(0)}} \min_{n\leqslant m\leqslant n+N} \frac{\log|B_m(\mathcal{T}_{\mathbf{M}};\mathrm{q},\mathrm{Q})|}{|\Delta_m|}.$$

Utilizing Lemma 3.4, we deduce that for each $n, N \in \mathbb{N}$,

$$\begin{aligned} \max_{(\mathrm{q},\mathrm{Q})\in W_{n+N}{}^{(0)}} \min_{n\leqslant m\leqslant n+N} \frac{|B_m(\mathcal{T}_{\mathbf{M}};\mathrm{q},\mathrm{Q})|}{|\Delta_m|} &\leqslant \min_{n\leqslant m\leqslant n+N} \max_{(\mathrm{q},\mathrm{Q})\in W_{n+N}{}^{(0)}} \frac{|B_m(\mathcal{T}_{\mathbf{M}};\mathrm{q},\mathrm{Q})|}{|\Delta_m|} \\ &\leqslant \min_{s\in\Gamma_{[N]}} \left[\max_{(\mathrm{p},\mathrm{P})\in Z_0} \sum_{m=0}^{N} s_m \mathcal{F}_{m:n+N}(\mathrm{p},\mathrm{P})\right] + o_n(1), \end{aligned}$$

where $o_n(1)$, according to Lemmas 3.3 and 3.5, is a number independent of N that vanishes as $n \to \infty$. Consequently, by Lemma 4.4 and Remark 4.5,

$$\limsup_{N\to\infty}\limsup_{n\to\infty} \alpha_{n,N} \leqslant \limsup_{N\to\infty} \min_{s\in\Gamma_{[N]}} \max_{(\mathrm{p},\mathrm{P})\in Z_0} \sum_{m=0}^{N} s_m \mathcal{F}_{m:\infty}(\mathrm{p},\mathrm{P})$$

$$\leqslant \limsup_{N\to\infty} \min_{s\in\Gamma_{[p]}} \max_{(\mathsf{p},\mathsf{P})\in Z_0} \sum_{m=0}^{p-1} s_m \sum_{n=0}^{pN-1} t^*_{n,N} \mathcal{F}_{n+m:\infty}(\mathsf{p},\mathsf{P})$$

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

implying that the upper bound coincides with the lower bound.

4.5 | Upper bound for general $\mathcal{T}_{\mathbf{M}}$

In this subsection, we derive an upper bound for $\mathcal{T}_{\mathbf{M}}$ without assumption (A1).

Proof of Corollary 1.5. Let $t^*_{n,N}$ be defined as in (21) with $p = 1$. Let $\alpha_{n,N}$ be as in (29) so that, as shown in Section 4.4, $\dim_H \mathcal{T}_{\mathbf{M}} \leqslant \limsup_{N\to\infty} \limsup_{n\to\infty} \alpha_{n,N}$ and

$$\limsup_{N\to\infty} \limsup_{n\to\infty} \alpha_{n,N} \leqslant \limsup_{N\to\infty} \min_{s\in\Gamma_{[N]}} \left[\max_{(\mathsf{p},\mathsf{P})\in Z} \sum_{m=0}^{N} s_m \mathcal{F}_{m:\infty}(\mathsf{p},\mathsf{P}) \right]$$

$$\leqslant \limsup_{N\to\infty} \max_{(\mathsf{p},\mathsf{P})\in Z} \sum_{m=0}^{N} t^*_{m,N} \mathcal{F}_{m:\infty}(\mathsf{p},\mathsf{P}).$$

Similar to Lemma 3.9, the last expression above admits a rearranged form as (A.4), which can be solved similarly to (A.6) (with $q_{i,N} = (N + \frac{1}{d-1})$ for $0 \leqslant i < N$). Explicitly, this yields

$$\limsup_{N\to\infty} \max_{(\mathsf{p},\mathsf{P})\in Z} \sum_{m=0}^{N} t^*_{m,N} \mathcal{F}_{m:\infty}(\mathsf{p},\mathsf{P}) \leqslant \limsup_{N\to\infty} \frac{1}{N + \frac{1}{d-1}} \log \left(\max_{a\in\mathcal{A}} (\mathbf{M}^N \mathbb{1})_a \right) = \log \rho(\mathbf{M}),$$

proving the proposed inequality. When $\mathbf{M}$ is irreducible, it follows from [2, Theorem 2.1] that the entropy satisfies that

$$\underline{\dim}_B \mathcal{T}_{\mathbf{M}} = h_{top}(\mathcal{T}_{\mathbf{M}}) \geqslant \log \rho(\mathbf{M}) \geqslant \dim_H \mathcal{T}_{\mathbf{M}}, \tag{30}$$

in which the first equality holds if and only if $\sum_{b\in\mathcal{A}} \mathbf{M}_{a,b}$ is identical for every $a \in \mathcal{A}$. Under the circumstances, one can simply calculate $h_{top}(\mathcal{T}_{\mathbf{M}}) = \log \rho(\mathbf{M})$ with $\rho(\mathbf{M}) = \sum_{b\in\mathcal{A}} \mathbf{M}_{a,b}$ for every $a \in \mathcal{A}$ and guess the spectral radius of the transfer operator $\mathcal{L}_{\mathbf{M},r}$ to be $\rho(\mathbf{M})^{\sum_{\ell=0}^{p-1} \prod_{i=0}^{\ell} r_i^{-1}}$, the eigenvalue associated with every eigenvectors $\sum_{a\in\mathcal{A}_j} \mathsf{e}_a \in C_j$, $j \in [p]$. Plugging it into the proposed formula (6) yields that the equalities in (30) hold simultaneously. □

Here, we provide an example to illustrate our formula.

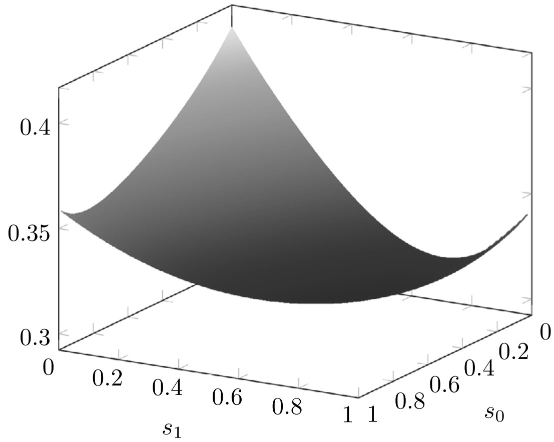

FIGURE 3 Spectral radius $\left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{-1}(s)\right)^{-1}\log\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M},r(s)})$.

Example 4.7. Let $\mathcal{T}_{\mathbf{M}}$ be a Markov hom tree-shift over the 3-tree associated with the incidence matrix $\mathbf{M}$, where

$$\mathbf{M} = \left[\begin{array}{c|c|c} \mathbf{0} & \mathbf{0} & \begin{matrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{matrix} \\ \hline \begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \end{matrix} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \begin{matrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{matrix} & \mathbf{0} \end{array}\right].$$

We compute the Hausdorff dimension numerically by determining the eigenvalues and eigenvectors of the corresponding transfer operator, as prescribed by Proposition 4.1. This is achieved through an iterative process, and the logarithm of the eigenvalue is plotted in Figure 3. Numerical results suggest that the optimal scaling vector s is approximately $(0.312, 0.588, 0.010)$ with $\log\rho_*(\mathcal{L}_{\mathbf{M},r(s^*)}) \approx 0.3027$, which is strictly less than the logarithm of the spectral radius of $\log\rho(\mathbf{M}^T) \approx 0.3208$ and is consistent with Corollary 1.5.

5 | DISCUSSION

Inspired by the classical large deviation principle for Markov chains, this article analogously establishes the large deviation principle for irreducible Markov chains indexed by rooted d-trees. Building on the belief of its potential for providing a better conception of the tree-shifts as well as further applications, the expositions in this paper showcase the capability of the method of types argument, which not only provides a theorem regarding almost sure convergence but also aids the determination of Hausdorff dimension of the set Markov hom tree-shift. Despite these accomplishments, our work is not intended to be an in-depth investigation but rather a starting point for further exploration. Indeed, several questions still remain unanswered in this work, which we list as follows:

- Is the rate function $\Lambda_j^*(\alpha)$ in Theorem 1.2 continuously differentiable if $\mathbf{M}$ is irreducible?
- If $\mathbf{M}$ is irreducible, can the optimal $\mathcal{L}_{\mathbf{M},r}$ in Theorem 1.3 be determined? Can the optimal Markov measure in Lemma 4.4 be determined?
- What is the Hausdorff dimension for general $\mathcal{T}_{\mathbf{M}}$ without the assumption of $\mathbf{M}$ being irreducible?

We hope this paper will inspire further research in this direction, advancing our understanding of tree-indexed Markov chains and tree-shifts.

APPENDIX A: PROOFS OF AUXILIARY LEMMAS

A.1 | Proof of Lemma 4.2

Due to the continuity of the objective function, it suffices to show the equality with $\inf_{s\in\mathring{\Gamma}_{[p]}}$ in place of all appearances of $\min_{s\in\Gamma_{[p]}}$ in (20). Under the circumstances, we first rephrase, by virtue of periodicity, the left-hand side of (20) as

$$\max_{\substack{\mathsf{P}\in\Upsilon^p\\ \mathsf{p}^{(p)}\in\mathcal{C}_0}}\ \inf_{\substack{s\in\mathring{\Gamma}_{[p]}\\ \mu\in\mathbb{R}^{\mathcal{A}}}}\left[-\sum_{i=0}^{p-1}\frac{\sum_{m=1}^{p}s_{m+i}d^m}{\sum_{m=1}^{p}d^m}\mathsf{p}^{(p)}\left(\prod_{\ell=1}^{p-i-1}\mathsf{P}^{(p-\ell)}\right)D_{\mathrm{KL}}(\mathsf{P}^{(i)}||\mathbf{M})+\left(\mathsf{p}^{(p)}\prod_{\ell=1}^{p-1}\mathsf{P}^{(p-\ell)}-\mathsf{p}^{(p)}\right)\mu\right], \tag{A.1}$$

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

As always, the expression on the left of (20) 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 3.7. We let $q_i(s)=(\sum_{m=1}^{p}d^m)^{-1}(\sum_{m=1}^{p}s_{m+i}d^m)$, and define

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

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 (A.1) into

$$\max_{\substack{\mathsf{P}^{(1:p-1)}\in\Upsilon^p\\ \mathsf{p}^{(p)}\in\mathcal{C}_0}}\ \inf_{\substack{s\in\mathring{\Gamma}_{[p]}\\ \mu\in\mathbb{R}^{\mathcal{A}}}}\left[-\sum_{i=1}^{p-1}\frac{\sum_{m=1}^{p}s_{m+i}d^m}{\sum_{m=1}^{p}d^m}\mathsf{p}^{(p)}\left(\prod_{\ell=1}^{p-i-1}\mathsf{P}^{(p-\ell)}\right)D_{\mathrm{KL}}(\mathsf{P}^{(i)}||\mathbf{M})\right.$$

$$\left.+\,\mathsf{p}^{(p)}\left(\prod_{\ell=1}^{p-1}\mathsf{P}^{(p-\ell)}\right)\lambda^{(1)}(s,\mu)-\mathsf{p}^{(p)}\mu\right],$$

Recursively using the convexity of $\lambda^{(i)}$ and applying the minimax theorem, one will move all max in (A.1) to the right of inf while retaining equality. In particular,

$$\max_{\substack{\mathsf{P}\in\Upsilon^p\\ \mathsf{p}^{(p)}\in C_0}} \inf_{\substack{s\in\mathring{\Gamma}_{[p]}\\ \mu\in\mathbb{R}^{\mathcal{A}}}} F(\mathsf{p}^{(p)},\mathsf{P},s,\mu) = \inf_{\substack{s\in\mathring{\Gamma}_{[p]}\\ \mu\in\mathbb{R}^{\mathcal{A}}}} \max_{\substack{\mathsf{P}\in\Upsilon^p\\ \mathsf{p}^{(p)}\in C_0}} F(\mathsf{p}^{(p)},\mathsf{P},s,\mu)$$

$$\geqslant \inf_{\substack{s\in\mathring{\Gamma}_{[p]}\\ \mu\in\mathbb{R}^{\mathcal{A}}}} \max_{\substack{\mathsf{P}\in\Upsilon^p\\ \mathsf{p}^{(p)}\in C_0\\ \mathsf{p}^{(p)}=\mathsf{p}^{(p)}\prod_{\ell=1}^{p-1}\mathsf{P}^{(p-\ell)}}} F(\mathsf{p}^{(p)},\mathsf{P},s,\mu)$$

$$= \inf_{s\in\mathring{\Gamma}_{[p]}} \max_{\substack{\mathsf{P}\in\Upsilon^p\\ \mathsf{p}^{(p)}\in C_0\\ \mathsf{p}^{(p)}=\mathsf{p}^{(p)}\prod_{\ell=1}^{p-1}\mathsf{P}^{(p-\ell)}}} F(\mathsf{p}^{(p)},\mathsf{P},s,0) = \min_{s\in\Gamma_{[p]}} \max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ p\text{-periodic}}} \sum_{m=0}^{p-1} s_m\mathcal{F}_m(\mathsf{p},\mathsf{P}),$$

which agrees with the right-hand side of (20) (with $\inf_{s\in\mathring{\Gamma}_{[p]}}$ in place of $\min_{s\in\Gamma_{[p]}}$.) The lemma is hence proved.

A.2 | Proof of Lemma 4.4

For the sake of simplicity, we denote by $F_N(\mathsf{p},\mathsf{P},s)$ the objective function of (23) throughout the subsection.

A.2.1 | Auxiliary lemmas

We first prove the following two auxiliary lemmas.

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

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

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

Proof. We note that q_i is a bijection between $\Gamma_{[p]}$ and $\{q\in\Gamma_{[p]} : q_0 \leqslant dq_1 \leqslant \cdots \leqslant d^{p-1}q_{p-1} \leqslant d^p q_0\}$. Indeed, $s\mapsto q_i(s)$ is an invertible linear transformation whose inverse can be found by Gaussian elimination, and every element in the latter set admits a preimage in the former. It remains to show that $q(s)\mapsto r(s)$ maps $\{q\in\Gamma_{[p]} : q_0 \leqslant dq_1 \leqslant \cdots \leqslant d^{p-1}q_{p-1} \leqslant d^p q_0\}$ bijectively into $\mathcal{R}_{p,d}$. It is not hard to see that such a 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 calculated as $q_0(s) = (\sum_{i=0}^{p-1}\prod_{\ell=0}^{i} r_\ell^{-1}(s))^{-1}$ and $q_i(s) = \prod_{\ell=0}^{i-1} r_\ell^{-1}(s) q_0(s)$ for $1\leqslant i<p$. □

Lemma A.2. *For every $s,s'\in\Gamma_{[p]}$ and every $N\in\mathbb{N}$*

$$\left|\max_{(\mathsf{p},\mathsf{P})\in Z_0} F_N(\mathsf{p},\mathsf{P},s) - \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_N(\mathsf{p},\mathsf{P},s')\right| \leqslant 2\cdot\log|\mathcal{A}|\cdot \mathrm{d}_v(s,s'). \qquad \text{(A.2)}$$

Proof. It is a consequence of the uniform boundedness $0\leqslant \mathcal{F}_{n:\infty}(\mathsf{p},\mathsf{P}) \leqslant \log|\mathcal{A}|$. □

A.2.2 | Proof setup

Denote $s_{p+i} = s_i$ for $i \in \mathbb{Z}_+$ and set

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

Multiplying by $(N + \frac{1}{d^p-1})$ and taking the limit $N \to \infty$, we obtain a sequence $(\overline{q}_i(s))_{i\in\mathbb{N}}$:

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

from which we define a p-tuple $r(s)$ that is previously studied in Lemma A.1:

$$r(s) = \left(\frac{\overline{q}_p(s)}{\overline{q}_{p+1}(s)}, \frac{\overline{q}_{p+1}(s)}{\overline{q}_{p+2}(s)}, \dots, \frac{\overline{q}_{2p-1}(s)}{\overline{q}_p(s)} \right). \tag{A.3}$$

Our discussions rely heavily on the following alternative expression of the objective function F_N:

$$F_N(\mathsf{p}, \mathsf{P}, s) = -\sum_{i=0}^{\infty} q_{i,N}(s)\mathsf{p}^{(i+1)} D_{\mathrm{KL}}(\mathsf{P}^{(i)}||\mathbf{M}). \tag{A.4}$$

Based on this expression, we consider the following finite approximation:

$$F_{N,m}(\mathsf{p}, \mathsf{P}, s) := -\sum_{i=0}^{m-1} q_{i,N}(s)\mathsf{p}^{(i+1)} D_{\mathrm{KL}}(\mathsf{P}^{(i)}||\mathbf{M}),$$

which, by Lemma A.2, admits the following error bound.

$$\left| \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_N(\mathsf{p}, \mathsf{P}, s) - \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_{N,pN}(\mathsf{p}, \mathsf{P}, s) \right| \leqslant \frac{(d-1)q_{pN-1,N}(s)}{d} \log |\mathcal{A}|. \tag{A.5}$$

Similar to Lemma 3.9 (or, originally, [3, Proposition 3.3]), the maximum of $F_{N,pN}(\mathsf{p}, \mathsf{P}, s)$ can be determined as

$$\max_{(\mathsf{p},\mathsf{P})\in Z_0} F_{N,pN}(\mathsf{p}, \mathsf{P}, s) = \max_{\mathsf{p}^{(pN)}=\mathsf{e}_a : a\in\mathcal{A}_0} \mathsf{p}^{(pN)} \lambda^{(pN),N}, \tag{A.6}$$

where

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

Notably, for $0 \leqslant i < pN$, the above can be expressed, by taking $\nu^{(i)}(s) = e^{\frac{\lambda^{(i),N}(s)}{q_{i,N}(s)}}$, as

$$\nu^{(i)}(s) = \mathbb{1} \text{ and } \nu^{(i+1)}(s) = \begin{cases} \Psi_{\mathbf{M},\frac{\overline{q}_i(s)}{\overline{q}_{i+1}(s)}}\left(\nu^{(i)}(s)\right) & \text{if } q_{i,N}(0) \neq 0, \\ \mathbb{1} & \text{otherwise,} \end{cases} \tag{A.7}$$

which is independent of N. In particular, this implies that the iteration satisfies

$$\nu^{(p+i)}(s) = \Psi_{\mathbf{M},r_i(s)}\left(\nu^{(p+i-1)}(s)\right) \quad \text{for } i = 0, \dots, p(N-1)-1,$$

which, in turn, yields

$$\nu^{(pN)}(s) = \mathcal{L}^{N-1}_{\mathbf{M},r^*}(\nu^{(p)}(s)). \tag{A.8}$$

A.2.3 | Proof of the lemma

Proof of Lemma 4.4. We first prove (23)≤(22). For every $\varepsilon > 0$, choose $r^* \in \mathcal{R}_{p,d}$ such that

$$\left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{*-1}\right)\log\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M},r^*}) < \inf_{r\in\mathcal{R}_{p,d}}\left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{-1}\right)\log\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M},r}) + \varepsilon,$$

so that, by Lemma A.1, we can find some $s^* \in \Gamma_{[p]}$ satisfying $r^* = r(s^*)$. Combining (A.5), (A.7), (A.8), we have

$$\begin{aligned}\limsup_{N\to\infty}\max_{(\mathsf{p},\mathsf{P})\in Z_0} F_N(\mathsf{p},\mathsf{P},s^*) &= \limsup_{N\to\infty}\max_{\mathsf{p}^{(pN)}=\mathsf{e}_a : a\in\mathcal{A}_0} \mathsf{p}^{(pN)}\lambda^{(pN),N}(s^*)_a \\ &= \limsup_{N\to\infty}\frac{\overline{q}_p(s)}{N+\frac{1}{d^p-1}}\log\left(\max_{a\in\mathcal{A}_0}\mathcal{L}^{N-1}_{\mathbf{M},r^*}(\nu^{(p)}(s))\right),\end{aligned} \tag{A.9}$$

which, by Proposition 4.1 (2), is bounded from above by

$$\begin{aligned}\overline{q}_p(s^*)\log\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M},r^*}) &= \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{*-1}\right)^{-1}\log\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M},r^*}) \\ &\leqslant \inf_{r\in\mathcal{R}_{p,d}}\left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{-1}\right)^{-1}\log\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M},r}) + \varepsilon,\end{aligned}$$

where the equality in the first line results from the following:

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

To prove (23)≥(22), we note that one can, by compactness of $Z_0 \times \Gamma_{[p]}$ and continuity of F_N, 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 Lemma A.2, further admits a convergent subsequence $s(N_i) \in \Gamma_{[p]}$ whose limit s^* satisfies

$$\liminf_{N\to\infty} \min_{s\in\Gamma_{[p]}} \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_N(\mathsf{p},\mathsf{P},s) = \lim_{i\to\infty} \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_{N_i}(\mathsf{p},\mathsf{P},s^*). \tag{A.10}$$

We henceforth fix $r^* = r(s^*)$, $\lambda^{(i),j} = \lambda^{(i),j}(s^*)$, and $\nu^{(i)} = \nu^{(i)}(s^*)$ as given in (A.3), (A.6), and (A.7), respectively. Recall that, as argued in Proposition 4.1 (2), $(\mathcal{L}_{\mathbf{M},r^*}(x)_a)_{a\in\mathcal{A}_0}$ depends only on $(x_a)_{a\in\mathcal{A}_0}$. Therefore, by choosing eigenvector $v^{(0)}$ satisfying $(v_a^{(0)})_{a\in\mathcal{A}_0} \leqslant (\nu_a^{(p)})_{a\in\mathcal{A}0}$, we derive the following by the order-preserving property of $\mathcal{L}_{\mathbf{M},r}$:

$$\lim_{i\to\infty} \max_{(\mathsf{p},\mathsf{P})\in Z_0} F_{N_i,pN_i}(\mathsf{p},\mathsf{P},s^*) = \lim_{i\to\infty} \frac{\overline{q}_p(s^*)}{N+\frac{1}{d^p-1}} \log\left(\max_{a\in\mathcal{A}_0} \mathcal{L}_{\mathbf{M},r^*}^{N-1}(\nu^{(p)})\right)$$
$$\geqslant \lim_{i\to\infty} \frac{\overline{q}_p(s)}{N+\frac{1}{d^p-1}} \log\left(\max_{a\in\mathcal{A}_0} \mathcal{L}_{\mathbf{M},r^*}^{N-1}(v^{(0)})\right) = \left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell} r_i^{*-1}\right)^{-1} \log\rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{M},r^*}).$$

Particularly, the argument here also justifies that "min" in (22) is attained: For the very same r^*, we see from the part (23)≤(22) that for any $s \in \Gamma_{[p]}$,

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

as desired.

We now prove (23)≤(24). To begin with, we fix onward s^* to be a minimizer of (24) and hence also $r^* = r(s^*)$, $\lambda^{(i),j} = \lambda^{(i),j}(s^*)$, and $\nu^{(i)} = \nu^{(i)}(s^*)$ as given in (A.3), (A.6), and (A.7), respectively. In view of (A.5), it suffices to prove the inequality holds with $F_N(\mathsf{p},\mathsf{P},s^*)$ in (23) replaced by $F_{N,pN}(\mathsf{p},\mathsf{P},s^*)$. Under the circumstances, for all $N \in \mathbb{N}$, we can choose, as in [3, Proposition 3.3], a maximizer $(\mathsf{p}^*,\mathsf{P}^*)$ of $F_{N,pN}(\mathsf{p},\mathsf{P},s^*)$ satisfying the following: For all $0 \leqslant i < pN$,

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

We should note that this expression is independent of N. Furthermore, without loss of generality, we may assume $\mathsf{p}^{(pN)} = \mathsf{e}_{a_0}$ for all N by passing to subsequence and changing our choice of $a_0 \in \mathcal{A}_0$ if necessary. Next, we construct a periodic pair $(\mathsf{q}^*,\mathsf{Q}^*)$ in the feasible domain of (24). Suppose $v^{(j)} \in C_j^+$, $j \in [p]$, are normalized eigenvectors given in Proposition 4.1 (b), which we again extend periodically by setting $v^{(j)} = v^{(j+p)}$ for all $j \in \mathbb{Z}$. Notably, due to the uniqueness of the eigenvectors, the following holds for all $j \in \mathbb{Z}$:

$$v^{(j-1)} = \Psi_{\mathbf{M},r_j}(v^{(j)})/\|\Psi_{\mathbf{M},r_j}(v^{(j)})\|.$$

Our 1-periodic $\mathsf{Q}^* \in \mathsf{Y}^{\mathbb{Z}_+}$ is then defined by

$$\mathsf{Q}^{*(i)}_{a,b} = \mathsf{Q}^{*(0)}_{a,b} := \begin{cases} \dfrac{\mathbf{M}_{a,b} v_b^{(j)}}{\sum_{c:c\in\mathcal{A}_{j-1}} \mathbf{M}_{a,c} v_c^{(j)}} & \text{if } a \in \mathcal{A}_{j-1}; \\ 0 & \text{otherwise.} \end{cases} \tag{A.11}$$

As $\mathsf{Q}^{*(i)} = \mathsf{Q}^{*(0)}$ are irreducible, each of the matrix is, by the Perron-Frobenius theorem, associated with a unique probability vector $\mathsf{q}^{*(i)} \in \Gamma \cap C_j^+$ satisfying $\mathsf{q}^{*(i)} = \mathsf{q}^{*(i+1)}\mathsf{Q}^{(0)}$ for all $i \in \mathbb{Z}_+$. To establish the aforementioned inequality, we need the following uniform estimate for the distance between $(\mathsf{p}^*, \mathsf{P}^*)$ and $(\mathsf{q}^*, \mathsf{Q}^*)$: There exists $C > 0$ and $0 < \theta < 1$, independent of N, such that the following holds for all $n \in \mathbb{Z}_+$:

$$\begin{cases} \sup_{a\in\mathcal{A}_{-i-1}} \mathsf{d}_v((\mathsf{P}^{*(i)}_{a,b})_{b\in\mathcal{A}}, (\mathsf{Q}^{*(i)}_{a,b})_{b\in\mathcal{A}}) \leqslant C\theta^i, \\ \mathsf{d}_v(\mathsf{p}^{*(i)}, \mathsf{q}^{*(i)}) \leqslant C(\theta^i + \theta^{pN-i}), \\ |\mathsf{p}^{*(i+1)} D_{\mathrm{KL}}(\mathsf{P}^{*(i)}||\mathbf{M}) - \mathsf{q}^{*(i+1)} D_{\mathrm{KL}}(\mathsf{Q}^{*(i)}||\mathbf{M})| \leqslant C(\theta^i + \theta^{pN-i}). \end{cases}$$

The first estimate results from Proposition 4.1 (3). For the second, note that

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

where the first term in the last line follows from the first estimate and the second from the Perron–Frobenius theorem. Finally, the last estimate is a consequence of the former together with the observation $x \mapsto x \log x$ is Lipschitz on any compact subinterval of (0,1). To complete the proof, note that $q_{i,N}(s^*) = O(1/N)$, and thus,

$$|F_{N,pN}(\mathsf{q}^*, \mathsf{Q}^*, s^*) - F_{N,pN}(\mathsf{p}^*, \mathsf{P}^*, s^*)| \leqslant \sum_{i=0}^{pN-1} q_{i,N}(s^*) C(\theta^i + \theta^{pN-i}) \to 0 \text{ as } N \to \infty.$$

The proof of the remaining inequalities is rather straightforward: (23)≥(24) follows from definition, while (24)≥(25) is the weak duality.

To prove the irreducibility of optimal transition matrix, we first note that if $\mathbf{M}' \prec \mathbf{M}$, then $\mathcal{L}_{\mathbf{M}',r}$ still maps C_j into C_j, and thus, $\rho_{\mathcal{A}_j}(\mathcal{L}_{\mathbf{M}',r})$ in Proposition 4.1 (2) are still well defined. We then suppose that $(\mathsf{p}^*, \mathsf{P}^*, s^*)$ is an optimizer of (25), r^* is a minimizer of (22), and $v^{(0)} \in C_0$ is an eigenvector of $\mathcal{L}_{\mathbf{M},r^*}$. Now, if $\mathsf{P}^{(0)}_{a',b'} = 0$ yet $\mathbf{M}_{a',b'} = 1$ for some $a', b' \in \mathcal{A}$, we denote the incidence matrix of $\mathsf{P}^{(0)}_{a',b'}$ as $\mathbf{M}'$ and write

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

We can then reproduce (22)≥(23) as before:

$$\inf_{r\in\mathcal{R}_{p,d}}\left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell}r_i^{-1}\right)^{-1}\cdot\log\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M}',r})\geqslant\limsup_{N\to\infty}\min_{s\in\Gamma_{[p]}}\max_{(\mathsf{p},\mathsf{P})\in Z_0'}\sum_{m=0}^{p-1}s_m\sum_{n=0}^{pN-1}t^*_{n,N}f_{n+m}(\mathsf{p},\mathsf{P}).$$

Additionally, we note that (23)≥(24)≥(25) follows naturally by definition. These altogether imply that

$$\begin{aligned}&\inf_{r\in\mathcal{R}_{p,d}}\left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell}r_i^{-1}\right)^{-1}\cdot\log\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M}',r})\\&\geqslant\max_{\substack{(\mathsf{p},\mathsf{P})\in Z_0:\\ \mathsf{P}\text{ is 1-periodic}\\ \mathsf{p}\text{ is }p\text{-periodic}}}\min_{s\in\Gamma_{[p]}}\sum_{m=0}^{p-1}s_m\mathcal{F}_{m:\infty}(\mathsf{p},\mathsf{P})=\sum_{m=0}^{p-1}s^*_m\mathcal{F}_{m:\infty}(\mathsf{p}^*,\mathsf{P}^*)\\&=\min_{r\in\mathcal{R}_{p,d}}\left(\sum_{\ell=0}^{p-1}\prod_{i=0}^{\ell}r_i^{-1}\right)^{-1}\cdot\log\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M},r}).\end{aligned}\tag{A.12}$$

However, since $\mathbf{M}$ is irreducible, there exists n such that $\mathcal{L}^n_{\mathbf{M}',r^*}(v^{(0)})_a<\rho_{\mathcal{A}_0}(\mathcal{L}_{\mathbf{M}',r})^n v_a^{(0)}$ for all $a\in\mathcal{A}_0$. As a consequence of the Collatz–Wielandt formula [19, Theorem 5.6.1], we deduce that

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

This implies

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

contradicting (A.12). Hence, $\mathsf{P}^*\sim\mathbf{M}$ is irreducible. In this case, the only feasible $\mathsf{p}^{(0)}$ is the left eigenvector of $(\mathsf{P}^{(0)})^p$ in $\mathcal{C}_0$. □

APPENDIX B: PACKING DIMENSION OF $\mathcal{T}_\mathbf{M}$ WHEN M IS IRREDUCIBLE

When the incidence matrix $\mathbf{M}$ is irreducible, the packing dimension satisfies $\dim_P\mathcal{T}_\mathbf{M}=\overline{\dim}_B\mathcal{T}_\mathbf{M}=d\cdot h_{top}(\mathcal{T}_\mathbf{M})$, a result that, to the best of our knowledge, has not been explicitly addressed in the literature. For completeness, we provide a proof below.

Proposition B.1. *Suppose that* $\mathbf{M}$ *is an irreducible incidence matrix. Then,* $\dim_P\mathcal{T}_\mathbf{M}=\overline{\dim}_B\mathcal{T}_\mathbf{M}=d\cdot h_{top}(\mathcal{T}_\mathbf{M})$.

Proof. Since $\dim_P\mathcal{T}_\mathbf{M}\leqslant\overline{\dim}_B\mathcal{T}_\mathbf{M}$ by definition and $\overline{\dim}_B\mathcal{T}_\mathbf{M}=d\cdot h_{top}(\mathcal{T}_\mathbf{M})$ as discussed in Section 4, it remains to show $\dim_P\mathcal{T}_\mathbf{M}\geqslant d\cdot h_{top}(\mathcal{T}_\mathbf{M})$.

We recall that when $\mathbf{M}$ is irreducible, there exists $a \in \mathcal{A}$ such that

$$\lim_{n\to\infty} \frac{\log\left|\left\{u \in B_{pn}(\mathcal{T}_{\mathbf{M}}) : u_\epsilon = a\right\}\right|}{|\Delta_{pn}|} = h_{top}(\mathcal{T}_{\mathbf{M}}), \tag{B.1}$$

where p is the common period of the states. This result appears as early as in [24, Proposition 3.1] and follows also as a consequence of [3, Theorem 3.1]. We henceforth fix a_0 in assumption (A1) to be one satisfying (B.1). The remainder of the argument is based on the mass distribution principle and is somewhat similar to Section 4.3.

Due to irreducibility, there exists $n_0 \in \mathbb{N}$ and a collection $\mathcal{B} = \{(b^{a,i})_{0\leqslant i\leqslant pn_0} : a \in \mathcal{A}\}$ of admissible paths from a to a_0 by incidence matrix $\mathbf{M}$, namely, $b^{a,0} = a$, $b^{a,pn_0} = a_0$, and $\mathbf{M}_{b^{a,i},b^{a,i+1}} = 1$ for all i. Choose increasing sequences $(\underline{N}_k)_{k\in\mathbb{N}}, (\overline{N}_k)_{k\in\mathbb{N}} \subset \mathbb{N}$ with each term divisible by p such that $\lim_{k\to\infty} \underline{N}_{k+1}/\overline{N}_k = \infty$ and that

$$0 = \overline{N}_0 < \underline{N}_1 < \overline{N}_1 := \underline{N}_1 + pn_0 < \underline{N}_2 < \overline{N}_2 := \underline{N}_2 + pn_0 < \cdots$$

We iteratively define a measure $\mathbb{P}$ over cylinder sets $C_{\overline{N}_k}$ $(k \in \mathbb{Z}_+)$ that concentrates on configurations with states a_0 on the boundary:

$$\mathbb{P}\Big(\bigcup \Big\{[u] \in C_{\overline{N}_k} : u_g = a_0, \forall g \in \Xi_{\overline{N}_k}\Big\}\Big) = 1. \tag{B.2}$$

To begin, define $\mathbb{P}$ by $\mathbb{P}([a_0]) = 1$. Suppose that $\mathbb{P}$ is defined for $C_{\overline{N}_k}$ and satisfies (B.2), extend it to a measure over $C_{\overline{N}_{k+1}}$ by first distributing uniform mass to each admissible configuration in $[u] \in C_{\underline{N}_{k+1}}$:

$$\begin{aligned}\mathbb{P}([u]) &= \mathbb{P}([u|_{\Delta_{\overline{N}_k}}])\left|\left\{v \in B_{\overline{N}_k:\underline{N}_{k+1}} : v_g = a_0, \forall g \in \Xi_{\overline{N}_k}\right\}\right|^{-1} \\ &= \mathbb{P}([u|_{\Delta_{\overline{N}_k}}])\left(\left|\left\{v \in B_{\underline{N}_{k+1}-\overline{N}_k} : v_\epsilon = a_0\right\}\right|\right)^{-d^{\overline{N}_k}},\end{aligned} \tag{B.3}$$

and then giving all its mass to the (unique) configuration $w \in C_{\overline{N}_{k+1}}$ whose paths from level $\underline{N}_{k+1}$ to $\overline{N}_{k+1}$ are of the form of $b^{a,i}$:

$$\mathbb{P}([w]) = \begin{cases} \mathbb{P}([w|_{\Delta_{\underline{N}_{k+1}}}]) & \text{if } (w_g)_{g'\leqslant g\leqslant g''} \in \mathcal{B}, \forall g' \in \Xi_{\underline{N}_{k+1}}, g'' \in \Xi_{\overline{N}_{k+1}} \\ 0 & \text{otherwise.}\end{cases}$$

In this manner, the resultant $\mathbb{P}$ satisfies (B.2) for configurations in $C_{\overline{N}_{k+1}}$. This definition uniquely extends $\mathbb{P}$ to a probability measure supported in $\mathcal{T}_{\mathbf{M}}$. In addition, combining (B.3) with $\lim_{k\to\infty} \underline{N}_{k+1}/\overline{N}_k = \infty$ and (B.1), we have almost surely,

$$\limsup_{n\to\infty} \frac{\log \mathbb{P}(B_r(X))}{\log r} = \limsup_{n\to\infty} -\frac{\log \mathbb{P}([X|_{\Delta_n}])}{|\Delta_{n-1}|} = d \cdot h_{top}(\mathcal{T}_{\mathbf{M}}).$$

To conclude the proof, we apply the mass distribution principle to $\mathbb{P}$. Note that $\cup_{i=0}^{\infty} C_i$ is a countable collection of sets and that for two cylinder sets, either they are disjoint or one is contained in the other. Hence, given any $\delta < d \cdot h_{top}(\mathcal{T}_{\mathbf{M}})$, any countable partition $(S_i)_{i\in\mathbb{N}}$ of $\mathrm{supp}(\mathbb{P})$, and every

$\varepsilon > 0$, one can find for each S_i a disjoint countable ε-cover $\mathcal{S}_i$ (and hence an ε-packing) satisfying

$$r^\delta \leqslant \mathbb{P}(\overline{B}_r(t)) \quad \text{for all } \overline{B}_r(t) \in \mathcal{S}_i,$$

where $\overline{B}_r(t)$ denotes the closed r-ball centered at t. This implies that the packing premeasure $P_0^\delta(S_i)$ is bounded from below by $\mathbb{P}(S_i)$:

$$P_0^\delta(S_i) \geqslant \sum_{\overline{B}_r(t)\in\mathcal{S}_i} r^\delta \geqslant \sum_{\overline{B}_r(t)\in\mathcal{S}_i} \mathbb{P}(\overline{B}_r(t)) \geqslant \mathbb{P}(S_i),$$

and therefore,

$$\sum_{i=1}^{\infty} P_0^\delta(S_i) \geqslant 1.$$

As a result, the packing measure $P^\delta(\mathrm{supp}(\mathbb{P}))$ is at least 1. This proves $\dim_P \mathcal{T}_{\mathbf{M}} \geqslant \dim_P \mathrm{supp}(\mathbb{P}) \geqslant d \cdot h_{top}(\mathcal{T}_{\mathbf{M}})$. □

ACKNOWLEDGMENTS

Ban is partially supported by the National Science and Technology Council, R.O.C. (Contract No. NSTC 111-2115-M-004-005-MY3) and National Center for Theoretical Science. Lai is partially supported by the National Science and Technology Council, R.O.C. (Contract No. NSTC 111-2811-M-004-002-MY2). The authors extend their gratitude to the anonymous referee for helpful comments that greatly improve the readability of the manuscript.

Open access publishing facilitated by Oulun yliopisto, as part of the Wiley - FinELib agreement.

JOURNAL INFORMATION

The *Journal of the London Mathematical Society* is wholly owned and managed by the London Mathematical Society, a not-for-profit Charity registered with the UK Charity Commission. All surplus income from its publishing programme is used to support mathematicians and mathematics research in the form of research grants, conference grants, prizes, initiatives for early career researchers and the promotion of mathematics.

ORCID

Yu-Liang Wu https://orcid.org/0000-0002-5181-056X

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, C.-H. Chang, W.-G. Hu, and Y.-L. Wu, *Topological entropy for shifts of finite type over $\mathbb{Z}$ and trees*, Theor. Comput. Sci. **930** (2022), 24–32, 2006.13415.
3. J.-C. Ban and Y.-L. Wu, *On the topological pressure of axial product on trees*, Stoch. Dyn. **25** (2025), no. 01, 2550007.
4. A. Banaji and I. Kolossváry, *Intermediate dimensions of Bedford–Mcmullen carpets with applications to Lipschitz equivalence*, Adv. Math. **449** (2024), 109735.
5. I. Benjamini and Y. Peres, *Markov chains indexed by trees*, Ann. Probab. **22** (1994), no. 1.
6. I. Benjamini and Y. Peres, *Tree-indexed random walks on groups and first passage percolation*, Probab. Theory Related Fields **98** (1994), no. 1, 91–112.

7. I. Benjamini and O. Schramm, *Percolation beyond $\mathbb{Z}^d$, nany questions and a few answers*, Electron. Commun. Probab. **1** (1996), 71–82.
8. T. Berger and Z. Ye, *Entropic aspects of random fields on trees*, IEEE Trans. Inform. Theory **36** (1990), no. 5, 1006–1018.
9. G. Brightwell and P. Winkler, *Hard constraints and the Bethe lattice: adventures at the interface of combinatorics and statistical physics*, Proceedings of the International Congress of Mathematicians, vol. III, Higher Education Press, Beijing, China, 2002, pp. 605–624.
10. R. M. Burton, C.-E. Pfister, and J. E. Steif, *The variational principle for Gibbs states fails on trees*, Markov Process. Related Fields **1** (1995), no. 3, 387–406.
11. A. Dembo, P. Morters, and S. Sheffield, *Large deviations of Markov chains indexed by random trees*, Ann. Inst. Henri Poincaré Probab. Stat. **41** (2005), no. 6, 971–996.
12. A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, 2nd ed., Stochastic modelling and applied probability, Springer, Berlin, Germany, 2009 (en).
13. T. P. Eggarter, *Cayley trees, the Ising problem, and the thermodynamic limit*, Phys. Rev. B **9** (1974), no. 7, 2989–2992.
14. A. Fan and J. P. Kahane, *How many intervals cover a point in random dyadic covering?*, Port. Math. (N.S.) **58** (2001), no. 1, 59–75 (eng).
15. H.-O. Georgii, *Gibbs measures and phase transitions*, De Gruyter, Berlin, New York, 2011.
16. R. A. Horn and C. R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 2012.
17. H. Huang, *The generalized entropy ergodic theorem for nonhomogeneous Markov chains indexed by a homogeneous tree*, Probab. Engrg. Inform. Sci. **34** (2019), no. 2, 221–234.
18. H. Komiya, *Elementary proof for Sion's minimax theorem*, Kodai Math. J. **11** (1988), no. 1.
19. B. Lemmens and R. Nussbaum, *Nonlinear Perron-Frobenius theory*, Cambridge University Press, Cambridge, 2012.
20. B. Lins, *Convergence of iterates in nonlinear Perron-Frobenius theory*, Discrete Contin. Dyn. Syst. Ser. B **28** (2023), no. 7, 3868–3886.
21. R. Lyons and R. Pemantle, *Random walk in a random environment and first-passage percolation on trees*, Ann. Probab. **20** (1992), no. 1.
22. R. Pemantle, *Automorphism invariant measures on trees*, Ann. Probab. **20** (1992), no. 3.
23. W. Peng, W. Yang, and B. Wang, *A class of small deviation theorems for functionals of random fields on a homogeneous tree*, J. Math. Anal. Appl. **361** (2010), no. 2, 293–301.
24. K. Petersen and I. Salama, *Tree shift topological entropy*, Theoret. Comput. Sci **743** (2018), 64–71.
25. K. Petersen and I. Salama, *Entropy on regular trees*, Discrete Contin. Dyn. Syst. **40** (2020), no. 7, 4453–4477.
26. K. Petersen and I. Salama, *Asymptotic pressure on some self-similar trees*, Stoch. Dyn. **23** (2022), no. 02.
27. C. J. Preston, *Gibbs states on countable sets*, Cambridge University Press, Cambridge, 1974.
28. Z. Shi and C. Ding, *A class of small deviation theorems for functionals of random fields on a tree with uniformly bounded degree in random environment*, Probab. Engrg. Inform. Sci. **36** (2022), no. 1, 169–183.
29. Z. Shi, Z. Wang, P. Zhong, and Y. Fan, *The generalized entropy ergodic theorem for nonhomogeneous bifurcating Markov chains indexed by a binary tree*, J. Theoret. Probab. **35** (2022), no. 3, 1367–1390.
30. F. Spitzer, *Markov random fields on an infinite tree*, Ann. Probab. **3** (1975), no. 3, 387.
31. W. Yang, *Some limit properties for Markov chains indexed by a homogeneous tree*, Statist. Probab. Lett. **65** (2003), no. 3, 241–250.
32. W. Yang, *A class of strong deviation theorems for the random fields associated with nonhomogeneous Markov chains indexed by a Bethe tree*, Stoch. Anal. Appl **30** (2012), no. 2, 220–237.
33. W. Yang and W. Liu, *Strong law of large numbers for Markov chains field on a Bethe tree*, Statist. Probab. Lett. **49** (2000), no. 3, 245–250.
34. W. Yang and Z. Ye, *The asymptotic equipartition property for nonhomogeneous Markov chains indexed by a homogeneous tree*, IEEE Trans. Inf. Theory **53** (2007), no. 9, 3275–3280.
35. Z. Ye and T. Berger, *Ergodic, regularity and asymptotic equipartition property of random fields on trees*, Comb. Inform. Syst. Sci. **21** (1996), 157–184.
36. Z. Ye and T. Berger, *Information measures for discrete random fields*, Science Press, Beijing, 1998.

IV

Mathematische Annalen
https://doi.org/10.1007/s00208-025-03233-3

Van der Corput and metric theorems for geometric progressions for self-similar measures

Amir Algom[1] · Yuanyang Chang[2] · Meng Wu[3] · Yu-Liang Wu[4]

Received: 4 July 2024 / Revised: 22 June 2025 / Accepted: 25 June 2025

Abstract

We prove a van der Corput lemma for non-atomic self-similar measures μ. As an application, we show that the correlations of all finite orders of $(x^n \mod 1)_{n\geq 1}$ converge to the Poissonian model for μ-a.e. x, assuming $x > 1$. We also complete a recent result of Algom, Rodriguez Hertz, and Wang (obtained simultaneously by Baker and Banaji), showing that any self-conformal measure with respect to a non-affine real analytic IFS has polynomial Fourier decay.

1 Introduction

1.1 A van der Corput lemma for fractal measures

A fundamental result in harmonic analysis is the van der Corput Lemma:

✉ Amir Algom
amir.algom@math.haifa.ac.il

Yuanyang Chang
chrischang2016@gmail.com

Meng Wu
meng.wu@oulu.fi

Yu-Liang Wu
yuliang.wu@oulu.fi

[1] Department of Mathematics, The University of Haifa at Oranim, Tivon 36006, Israel

[2] Department of Mathematics, School of Mathematics and Statistics, Wuhan University of Technology, Wuhan, China

[3] School of Mathematics, Hunan University, Changsha 410082, China

[4] Department of Mathematical Sciences, University of Oulu, Oulu 90014, Finland

Published online: 06 August 2025

Proposition *[37, Proposition 2 in Chapter 8] Let g be a real-valued smooth function on an interval $J \subset \mathbb{R}$. Suppose $|g^{(k)}(x)| \geq 1$ for all $x \in J$, where $k \geq 1$ is an integer. If $k = 1$ and g' is monotonic, or if $k > 1$, then*

$$\left| \int_J e^{2\pi i \lambda g(x)} dx \right| \leq c_k \lambda^{-\frac{1}{k}} \ for \ \lambda > 0, where c_k > 0 is independent of \lambda. \quad (1)$$

The purpose of this paper is to prove variants of this result when the underlying measure in (1) is a self-similar measure, while assuming as little as possible about the smoothness of the phase function *g*. We will then apply this to study the fine statistics of sequences of the form $(x^n \mod 1)_{n \geq 1}$ where *x* is drawn according to such a measure, and to study the Fourier decay problem for self-conformal measures.

Self-similar measures are defined as follows: Let $\Phi = \{f_1, ..., f_n\}$ be a finite set of real non-singular contracting similarity maps of a compact interval $J \subset \mathbb{R}$. That is, for every *i* we can write $f_i(x) = r_i \cdot x + t_i$ where $r_i \in (-1, 1) \setminus \{0\}$ and $t_i \in \mathbb{R}$, and $f_i(J) \subseteq J$. We will refer to Φ as a *self-similar IFS* (Iterated Function System). It is well known that there exists a unique compact set $\emptyset \neq K = K_\Phi \subseteq J$ such that

$$K = \bigcup_{i=1}^{n} f_i(K).$$

The set *K* is called a *self-similar set*, and the *attractor* of the IFS Φ. We always assume that there exist $i \neq j$ such that the fixed point of f_i is not equal to the fixed point of f_j. This ensures that *K* is infinite.

Next, let $\mathrm{p} = (p_1, ..., p_n)$ be a strictly positive probability vector, that is, $p_i > 0$ for all *i* and $\sum_i p_i = 1$. It is well known that there exists a unique Borel probability measure μ such that

$$\mu = \sum_{i=1}^{n} p_i \cdot f_i \mu, \qquad \text{where } f_i\mu \text{ is the push-forward of } \mu \text{ via } f_i. \quad (2)$$

The measure μ is called a *self-similar measure*, and is supported on *K*. Our assumptions that *K* is infinite and that $p_i > 0$ for every *i* are known to imply that μ is non-atomic. In particular, all self-similar measures in this paper are non-atomic. It is a standard fact that the Lebesgue measure on any bounded interval is a self-similar measure, but there are many self-similar measures that are not absolutely continuous.

We can now state one of the main results of this paper:

Theorem 1.1 *Let μ be a non-atomic self-similar measure supported on an interval $J \subseteq \mathbb{R}$. Let $g \in C^k(J)$ for some integer $k \geq 2$ be such that for some $c_0 > 0$, $\max_{2 \leq j \leq k} |g^{(j)}(x)| \geq c_0$ for all $x \in J$. Then there exist constants $\tau = \tau(\mu, k)$ and $C = C(g) > 0$ such that*

$$\left|\int_J e^{2\pi i g(x)\lambda} d\mu(x)\right| \leq C\lambda^{-\tau} \quad for\ \lambda > 0. \tag{3}$$

In fact, we will prove a more general result, where the phase g is only required to be $C^{1+\gamma}(\mathbb{R})$ and non-flat in an appropriate sense; see Theorem 1.6 in Section 1.4 for more details.

The exponent τ as in Theorem 1.1 depends on the geometry of μ, and is uniform across all g satisfying its assumptions (see Section 1.5 for comments on quantitative aspects of the proof). Some special cases of Theorem 1.1 had been established in the past: Kaufman [22] proved, among other things, that if g is any C^2 diffeomorphism on [0, 1] such that $g'' > 0$, then (3) holds for the Cantor-Lebesgue measure on the middle-thirds Cantor set. This was later extended to all self-similar measures with a uniform contraction ratio by Mosquera-Shmerkin [31]. Since the work of Mosquera-Shmerkin it remained an open problem as to whether the result holds for *all* non-atomic self-similar measures (with no further assumptions). Theorem 1.1 settles this problem, while also relaxing the assumptions on the phase function. Finally, Algom, Rodriguez Hertz, and Wang [5, 7] recently proved logarithmic decay for $C^{1+\gamma}(\mathbb{R})$ phases under certain Diophantine assumptions on the IFS. No such assumptions are required for our corresponding result, Theorem 1.6 below. See Section 1.4 for more details.

We also remark that, simultaneously and independently of our work, Baker and Banaji [10] obtained similar results but with a different method, and a somewhat different range of applications. Baker and Banaji rely on a disintegration technique for self-similar measures as in [4], and apply this to study Fourier decay for certain smooth images of fractal measures on the so-called fibered IFSs. Their technique applies in higher dimensions and to infinite IFSs. In contrast, we rely on an application of Tsujii's large deviations estimate [40] (see Section 1.5), and apply this to study the fine statistics of sequences of the form $(x^n \mod 1)_{n\geq 1}$ (Section 1.2). However, both results can be applied to show that any non-atomic self-conformal measure with respect to a non-affine real analytic IFS has polynomial Fourier decay, a result first stated in the recent work of Algom, Rodriguez Hertz, and Wang [6, Corollary 1.2]. See Section 1.3 for more details.

1.2 Applications to higher order correlations of sequences modulo 1

A sequence $(x_n)_{n\geq 1}$ taking values in [0, 1) is called *uniformly distributed* or *equidistributed* if for any sub-interval $J \subseteq [0, 1)$,

$$\lim_{N\to\infty} \frac{1}{N} \sharp\{1 \leq n \leq N : x_n \in J\} = |J|, \text{ where } |J| := \text{ the Lebesgue measure of } J.$$

A real valued sequence $(x_n)_{n\geq 1}$ is called *uniformly distributed modulo one* if the sequence of fractional parts $(\{x_n\})_{n\geq 1} \subset [0, 1)$ is uniformly distributed in [0, 1). This notion concerns the proportion of the fractional parts of the sequence in a fixed interval. The *fine-scale* statistics of sequences modulo one, which describe the behaviour of a sequence on the scale of mean gap 1/*N*, has attracted growing attention in recent years. Two of the most important fine-scale statistics are the *k*-level correlations and the distribution of level spacings (also called nearest-neighbour gaps), which are defined as follows.

For every integer $N \geq 1$ and $k \geq 2$ let $\mathcal{U}_k = \mathcal{U}_k(N)$ denote the set of distinct integral *k*-tuples taking values in $\{1, \ldots, N\}$. That is,

$$\mathcal{U}_k = \{\mathrm{u} = (u_1, \ldots, u_k) : \, u_i \in \{1, \ldots, N\}, \, u_i \neq u_j \text{ for all } i \neq j\} \, .$$

For each $\mathrm{u} \in \mathcal{U}_k$ and real valued sequence $(x_n)_{n\geq 1}$ consider the difference vector

$$\Delta(\mathrm{u}, (x_n)_{n\geq 1}) = (x_{u_1} - x_{u_2}, \ldots, x_{u_{k-1}} - x_{u_k}) \in \mathbb{R}^{k-1}.$$

Given $f : \mathbb{R}^{k-1} \to \mathbb{R}$ with compact support, the *k-level correlation function* is defined to be

$$R_k(f, (x_n)_{n\geq 1}, N) := \frac{1}{N} \sum_{\mathrm{u}\in\mathcal{U}_k} \sum_{\mathrm{l}\in\mathbb{Z}^{k-1}} f(N(\Delta(\mathrm{u}, \{x_n\}_{n\geq 1}) + \mathrm{l})). \tag{4}$$

If

$$\lim_{N\to\infty} R_k(f, (x_n)_{n\geq 1}, N) = \int_{\mathbb{R}^{k-1}} f(\mathrm{x}) d\mathrm{x}, \ \forall f \in C_c^{\infty}(\mathbb{R}^{k-1}) \tag{5}$$

then we say that *the k-level correlation of* $(\{x_n\})_{n\geq 1}$ *is Poissonian*. This notion alludes to the fact that such behaviour is consistent with the almost sure behaviour of a Poisson process with intensity one.

We now define the distribution of level spacings or nearest-neighbour gaps of the sequence $(\{x_n\})_{n\geq 1}$, that is, the gaps between consecutive elements of $(\{x_n\})_{n\geq 1}$. For each $N \geq 1$, we reorder the sequence $\{x_n\}_{n=1}^{N}$ and label them as

$$0 \leq \theta_{1,N} \leq \theta_{2,N} \leq \cdots \leq \theta_{N,N} \leq 1, \quad \text{ and set } \theta_{0,N} := \theta_{N,N} - 1 \pmod 1.$$

Suppose that for every $s \geq 0$ the limit as $N \to \infty$ of the function

$$G(s, \{x_n\}_{n\geq 1}, N) := \frac{1}{N} \sharp \{1 \leq n \leq N : \, N\,(\theta_{n,N} - \theta_{n-1,N}) \leq s\}$$

exists. Then the limit function *G*(*s*) is called the asymptotic distribution function of the level spacings of $(\{x_n\})_{n\geq 1}$. We say that the level spacings are *Poissonian* if $G(s) = 1 - e^{-s}$, which agrees with the distribution of the waiting time of a Poisson process with intensity one. We refer the reader to [2] for more exposition on why and how these notions are used to capture pseudo-randomness properties of sequences.

Recently, Aistleitner and Baker [1] proved that $(\{\alpha^n\})_{n\geq 1}$ has Poissonian pair correlations for almost every $\alpha > 1$. This refined a classical Theorem of Koksma that such sequences are equidistributed (almost surely), since it is well known that a sequence with Poissonian pair correlations is always equidistributed [3, 20, 29]. Aistleitner and Baker [1] further conjectured that typically such sequences should have Poissonian *k*-level correlations. Very recently, Aistleitner, Baker, Technau, and Yesha [2] proved this conjecture, and thus obtained that the level spacings of $\{\alpha^n\}_{n\geq 1}$ are Poissonian for a.e. $\alpha > 1$; here we use the well-known fact that if the *k*-level correlation of a sequence is Poissonian for all $k \geq 2$, then the level spacings are also Poissonian [24, Appendix A].

The results cited in the previous paragraph hold for α distributed according to the Lebesgue measure. The next theorem generalizes them to the setting where α is distributed according to an arbitrary non-atomic self-similar measure.

Theorem 1.2 *Let μ be a non-atomic self-similar measure on $\mathbb{R}$ such that μ-a.e. x is larger than 1. Let $\xi \in \mathbb{R} \setminus \{0\}$. Then for μ-almost every x the k-level correlation of $(\{\xi x^n\})_{n\geq 1}$ is Poissonian for all $k \geq 2$. In particular, the level spacings are also Poissonian, and the sequence $(\{x^n\})_{n\geq 1}$ is uniformly distributed.*

Baker [9] recently proved that for some self-similar measures, $(\{x^n\})_{n\geq 1}$ is uniformly distributed almost surely. For example, this was shown for appropriate translates of the Cantor-Lebesgue measure on the middle-thirds Cantor set. Baker further conjectured [9, Conjecture 1.1] that this should hold for all self-similar measures supported on $[1, \infty)$. Theorem 1.2 proves this conjecture. Also, as the Lebesgue measure restricted on any interval is a self-similar measure, Theorem 1.2 generalizes the results of [2] discussed above; moreover, our methods also yield an alternative proof of the Lebesgue case treated in that paper.

Let us now say a brief word about the relation between Theorem 1.1, which concerns Fourier decay of smooth images of a non-atomic self-similar measure μ, and Theorem 1.2, which studies the fine-scale statistics of sequences of the form $(\{x^n\})_{n\geq 1}$ where $x \sim \mu$. The idea is that, by virtue of Poisson summation and an observation of Technau and Yesha [39, Proposition 7.1] (see also Proposition 4.8 below), the limit of the *k*-level correlations in (5) exists if: for some $\tau > 0$ that is independent of the $f \in C_C^\infty(\mathbb{R}^{k-1})$ we have

$$\left\| \frac{1}{N^k} \sum_{\mathbf{l}\in\mathbb{Z}^{k-1}\setminus\{0\}} \widehat{f}\left(\frac{\mathbf{l}}{N}\right) \sum_{\mathbf{u}\in\mathcal{U}_k} e^{-2\pi i \sum_{j=1}^{k-1} l_j (x^{u_j} - x^{u_{j+1}})} \right\|_{L^2(\mu)} = O\left(N^{-\tau}\right), \text{ as } N \to \infty.$$

In Theorem 4.1 we will prove super-polynomial decay of the μ-means of individual oscillatory integrals that arise from the sum above. This can be recast as providing a uniform van der Corput estimate for phases arising from a certain family of polynomials (see Definition 4.2); whence the relation with Theorem 1.1. These estimates rely on a non-trivial adaptation of the proof of our main technical result, Theorem 1.6, that is discussed in Section 1.4.

1.3 Applications to fourier decay of self-conformal measures with respect to real analytic IFSs

An important special case of Theorem 1.1 is when the phase g is real analytic and not affine. By virtue of Lemma 2.7 proved below, Theorem 1.1 applies to such phase functions, and we thus obtain the following notable corollary:

Corollary 1.3 *Let μ be a non-atomic self-similar measure supported on an interval $J \subseteq \mathbb{R}$, and let $g \in C^\omega(J)$ be a non-affine real analytic function. Then there exist $\tau > 0$ and $C > 0$ such that*

$$\left| \int_J e^{2\pi i g(x)\lambda} d\mu(x) \right| \leq C\, |\lambda|^{-\tau}, as\ \ |\lambda| \to \infty. \tag{6}$$

Corollary 1.3 has important applications regarding the Fourier decay problem for self-conformal measures: Let Φ be a $C^\omega(J)$ IFS; This simply means that all the maps in Φ are real analytic strict contractions on the interval J. A great deal of attention has been given to the study of the rate at which the Fourier transform of self-conformal measures with respect to such IFSs (i.e., measures that satisfy (2) with respect to the IFS) decays. They arise naturally, e.g., as (finitely many) inverse branches of the Gauss map in number theory [21] and as Furstenberg measures for some $\mathrm{SL}(2, \mathbb{R})$ cocycles [8, 42], and are closely related to Patterson-Sullivan measures on limit sets of some Schottky groups [12, 27, 32], among others.

Combining Corollary 1.3 with very recent results of Algom, Rodriguez Hertz, and Wang [6], we can show that all such measures have polynomial Fourier decay, as long as Φ is not entirely made up of affine maps. Denoting by $\widehat{\nu}(\lambda)$ the Fourier transform of a Borel probability measure ν, we have:

Theorem 1.4 *Let Φ be a $C^\omega(\mathbb{R})$ IFS. If Φ contains a non-affine map then every non-atomic self-conformal measure ν admits some $\tau > 0$ such that*

$$\widehat{\nu}(\lambda) = O\left(\frac{1}{|\lambda|^\tau} \right)\ as\ \ |\lambda| \to \infty.$$

This result was first announced in the work of Algom, Rodriguez Hertz, and Wang [6, Corollary 1.2]. The last part of its proof in [6, Section 6] required the estimate (6) that was announced in [6, Theorem 6.5]. Thus, with (6) we complete the proof of Theorem 1.4. Also, recall from our discussion at the end of Section 1.1 that Baker and Banaji [10] obtained Theorem 1.4 simultaneously and independently of our work. See also the related work of Baker-Sahlsten [11] regarding the non-linearizable case.

We emphasize that except for the existence of a non-affine map, no further conditions are imposed on the IFS. Thus, Theorem 1.4 extends the recent work of Sahlsten-Stevens [35, Theorem 1.1] by removing their total non-linearity and separation assumptions. Other related significant works include those of Li [25, 26] about Furstenberg measures, Bourgain-Dyatlov [12] about Patterson-Sullivan measures, Baker and Sahlsten [11] about totally non-linear C^2 IFSs, and our previous works [5, 7].

See the recent survey [34] for more discussion about the recent breakthroughs on the Fourier decay problem.

Finally, we remark that the only concrete examples of polynomial Fourier decay in the fully self-similar setting were given by Dai-Feng-Wang [17] and Streck [38] for some homogeneous IFSs. It is known, though, that most self-similar IFSs should satisfy this property; see Solomyak [36]. For more recent results on Fourier decay for self-similar measures we refer to [7, 13, 14, 16, 17, 28, 31, 33, 36, 38, 41] and references therein.

1.4 Main technical theorem: van der Corput for phases with non-flat derivatives

In this Section we will state and discuss our main technical result, Theorem 1.6. It implies Theorem 1.1, and thus Corollary 1.3, as special cases. The results regarding fine statistics of sequences, Theorem 1.2, follow from Theorem 4.1, which requires delicate fine-tuning of the proof of Theorem 1.6 for certain exponential functions.

We first formulate the following definition. For a given bounded set $A \subseteq \mathbb{R}$ let $\mathcal{N}_r(A)$ denote the minimal number of balls of diameter *r* needed to cover *A*.

Definition 1.5 Let $0 < \delta \leq 1$. A function $h : \mathbb{R} \to \mathbb{R}$ is called δ-non-flat on an interval $J \subset \mathbb{R}$ if:

For all $\epsilon > 0$ there exists $n_0 = n_0(\epsilon) \geq 1$ such that for all $n > n_0$ and every bounded subset $A \subset J$ we have

$$\mathcal{N}_{2^{-\delta n}}(A) \geq 2^{\epsilon n} \Rightarrow \operatorname{diam}(h(A)) \geq 2^{-n}.$$

The idea is that the range of a δ-non-flat function on a set of positive dimension can not be too small. Note that by definition, non-singular affine functions on $\mathbb{R}$ are 1-non-flat, and for any $0 < \delta_1 < \delta_2 \leq 1$, δ_2-non-flatness implies δ_1-non-flatness. In addition, in [18, prior to Theorem 6.2], de Melo and van Strien discuss a related notion of non-flatness of a C^1 function: By their definition, a function $h \in C^1(J)$ is non-flat at a point $c \in J$ if there is a $C^2(\mathbb{R}, J)$ diffeomorphism φ such that $\varphi(0) = c$ and $h \circ \varphi$ is a polynomial (that we assume is non-trivial). It follows from the results in Sect. 2.3 that, locally, if *h* is non-flat in the sense of de Melo and van Strien then it is non-flat in the sense of Definition 1.5.

We can now state our main technical result.

Theorem 1.6 *Let μ be a non-atomic self-similar measure supported on a closed interval $J \subseteq \mathbb{R}$. Let $g \in C^{1+\alpha}(J)$ for some $0 < \alpha \leq 1$. Suppose that g' is δ-non-flat on J, where $0 < \delta \leq 1$. Then there exist constants $\tau = \tau(\mu, \delta, \alpha)$ and $C = C(g) > 0$ such that*

$$\left| \int e^{2\pi i g(x)\lambda} d\mu(x) \right| \leq C\lambda^{-\tau} \quad for\ \lambda > 0.$$

By Lemma 2.6 if $g \in C^k(J)$ with $k \geq 1$ and there is some $c_0 > 0$ with $\max_{1 \leq j \leq k} |g^{(j)}(x)| \geq c_0$ for all $x \in J$, then it is $\frac{1}{2k}$-non-flat. Thus, Theorem 1.6 implies Theorem 1.1 as a special case. Finally, consider a non-flat $h \in C^{1+\alpha}(J)$. Definition 1.5 suggests that the optimal non-flatness exponent should roughly correspond to the Hölder exponent of h^{-1} (which might only be defined locally on $h(J)$). Based on this heuristic, we believe a function $g \in C^k(J)$ as above should, in fact, be $\frac{1}{k}$-non-flat. While this potential improvement is not essential for the purposes of the present paper, it may play a role in obtaining effective estimates on τ, as discussed at the end of the next section.

1.5 Outline of method

In this section we sketch the proof of Theorem 1.1. We work under the stronger assumption that the phase g satisfies $g'' > 0$ on the interval; the general case, as well as our other main results, is proved via subtle optimizations of this method. So, let μ be a non-atomic self-similar measure supported on, say, [0, 1], and let $g \in C^2([0, 1])$ be a diffeomorphism such that $g'' > 0$. We show that there exists some $\alpha > 0$ such that for $\lambda > 0$

$$\left| \int_0^1 e^{2\pi i g(x)\lambda} d\mu(x) \right| = O\left(\lambda^{-\alpha}\right).$$

First, we recall the main ingredients in the approach of Kaufman [22] and Mosquera-Shmerkin [31], in the special case where the IFS is homogeneous; that is, all the maps in $\Phi = \{f_1, ..., f_n\}$ share the same contraction ratio. Here, the measure μ is known to have a convolution structure, so we may write explicitly

$$\widehat{\mu}(\lambda) = \prod_{k=1}^{\infty} \left(\sum_{i=1}^{n} p_i \exp\left(2\pi t_i r^k \lambda\right) \right), \text{ where } f_i(x) = r \cdot x + t_i \text{ for each } 1 \leq i \leq n.$$

From this decomposition, Kaufman, and later Mosquera-Shmerkin, first showed polynomial decay outside of a sparse set of frequencies:

$$\forall \epsilon > 0\, \exists c_0 > 0 \text{ s.t. for } T \gg 1,\ \left\{\lambda \in [-T, T] : |\widehat{\mu}(\lambda)| \geq T^{-c_0}\right\} \text{ can be covered by } T^{\epsilon} \text{ intervals of length 1.} \quad (7)$$

In fact, Mosquera-Shmerkin [31, Proposition 2.3] proved a quantitative version of this result, which enabled them to give explicit lower bounds on the Fourier dimension of $g\mu$ in some cases. The proof then proceeded by combining the previous two displayed equations; see, e.g., [31, Theorem 3.1]. Tsujii [40] extended (7) to all self-similar measures (without relying on convolution structure). Recently, Khalil [23, Corollary 1.8] proved a very general form of this result.

The main innovation of this paper is the introduction of a new method to obtain polynomial Fourier decay for $g\mu$ from a large deviations estimate as in (7), for μ

without a convolution structure, and for more general phase functions. Our initial observation is that we can rewrite (7) as follows:

$$\forall\epsilon>0\,\exists c>0 \text{ s.t. for } t\gg 1,\quad \sharp\left\{n\in\mathbb{Z}:\exists\lambda\in[n,n+1]\cap[-e^t,e^t] \text{ such that } |\widehat{\mu}(\lambda)|\geq e^{-c\cdot t}\right\}\leq e^{\epsilon\cdot t}. \quad (8)$$

At this point we need some notation: Let us fix our self-similar IFS $\Phi=\{f_1,...,f_n\}$ and denote $\mathcal{A}:=\{1,..,n\}$. For every $i\in\mathcal{A}$ we write

$$f_i(x)=r_i\cdot x+t_i,\ \text{where } 0<|r_i|<1 \text{ and } t_i\in\mathbb{R}.$$

For every $\omega\in\mathcal{A}^*:=\cup_{i\geq 1}\mathcal{A}^i$ of length $|\omega|=\ell$ write, recalling that $\mu=\mu_{\mathrm{p}}$ for the probability vector $\mathrm{p}=(p_1,...,p_n)$,

$$r_\omega:=r_{\omega_1}\cdots r_{\omega_\ell},\ \text{and } f_\omega(x)=r_\omega\cdot x+t_\omega,\ \text{where } t_\omega:=f_\omega(0),\ \text{and also } p_\omega=p_{\omega_1}\cdots p_{\omega_\ell}. \quad (9)$$

For a word $\omega\in\mathcal{A}^*$ let ω^- denote the word of length $|\omega|-1$ obtained from ω by forgetting its last letter. For sufficiently small (not necessarily integer) $b>0$ we consider the following cut-set of $\mathcal{A}^{\mathbb{N}}$:

$$\mathcal{W}_b=\{\omega\in\mathcal{A}^*:|r_\omega|\leq b<|r_{\omega^-}|\}\,. \quad (10)$$

Now, the first step in our proof is, using self-similarity and the C^2 assumption, to linearize:

$$\left|\int_0^1 e^{2\pi i\lambda g(x)}dx\right|=|\widehat{g\mu}(\lambda)|\leq\sum_{\omega\in\mathcal{W}_b}p_\omega\cdot|\widehat{\mu}\left(g'(t_\omega)r_\omega\lambda\right)|+O\left(\lambda\cdot b^2\right),$$

where $b=b(\lambda)>0$ is an auxiliary parameter to be chosen later.

Next, fixing yet another auxiliary parameter $\epsilon>0$, there is a $c_0>0$ such that (8) holds when taking $t\approx\log(\lambda\cdot b)$. Thus, we can partition $\mathcal{W}_b$ into

$$X:=\left\{\omega:\ |\widehat{\mu}\left(g'(t_\omega)r_\omega\lambda\right)|\geq|\lambda\cdot r_\omega|^{-c_0}\right\},\ \text{and } \mathcal{W}_b\setminus X.$$

It is clear that

$$\sum_{\omega\in\mathcal{W}_b\setminus X}p_\omega\cdot|\widehat{\mu}\left(g'(t_\omega)r_\omega\lambda\right)|<|\lambda\cdot b|^{-c_0}\,.$$

So, it remains to control

$$\sum_{\omega\in X}p_\omega\cdot|\widehat{\mu}\left(g'(t_\omega)r_\omega\lambda\right)|\,.$$

This is where (8) and the non-linearity condition $g'' > 0$ enter the picture. Fix $n \in \mathbb{Z}$ such that:

$$[n, n+1] \text{ contains } \text{ some } g'(t_\omega) r_\omega \lambda \text{ with } |\widehat{\mu}\left(g'(t_\omega) r_\omega \lambda\right)| \geq |\lambda \cdot r_\omega|^{-c_0}.$$

Roughly speaking, using the non-linearity of *g* (or, more generally, the non-flatness of g'), and the Hölder regularity of μ ([19] and Proposition 2.1), one can show that for some uniform $s_0 = s_0(\mu) > 0$,

$$\sum_{\omega \in \mathcal{W}_b} p_\omega \cdot 1_{\lambda \cdot g'(t_\omega) \cdot r_\omega \in [n, n+1]}(\omega) \leq \mu\left(\text{ some ball of radius } \lambda \cdot b\right) = O\left((\lambda \cdot b)^{-s_0} \right). \quad (11)$$

The last displayed equation means that in each such interval $[n, n+1]$ the probability of having a "bad" frequency is very small; in addition, (8) means that the number of such “bad” intervals is sub-polynomial - here it can be shown to be less than $(\lambda \cdot b)^{\epsilon}$. Thus, taking tally of our calculation and disregarding global multiplicative constants,

$$\left| \int_0^1 e^{2\pi i \lambda g(x)} dx \right| \leq |\lambda \cdot b|^{-c_0} + |\lambda \cdot b|^{\epsilon} \cdot |\lambda \cdot b|^{-s_0} + |\lambda \cdot b|^2.$$

This can be made polynomially small in λ by a careful choice of the parameters *b* and ϵ (note that $s_0 = s_0(\mu)$ and $c_0 = c_0(\epsilon)$ are not free for us to choose).

In addition, we remark that in practice we will derive the inequality (11) not on the full space $\mathcal{W}_b$, but rather only upon a further partition of this space into those ω, η such that $r_\omega = r_\eta$. This introduces another factor, the number of such distinct r_ω, that is large in λ into the previously displayed equation; however, standard properties of self-similar IFSs show that it can be made logarithmic in $b = b(\lambda)$, and so the proof goes through.

Finally, although providing effective estimates is not the primary goal of this paper, let us briefly address the issue. Taking the δ-non-flatness of g' into consideration in (11), the proof of Theorem 1.6 as outlined above gives a decay exponent of the form

$$\tau \approx \min\{\frac{1}{2}, c_0, \frac{s_0 \cdot \delta}{2} - \epsilon\}.$$

See (20) for more details (note that the $\frac{1}{2}$ term arises from the final error term on the right-hand side). In particular, the bound on τ deteriorates as δ decreases. In the context of Theorem 1.1, this translates to slower decay rates as $k \geq 2$, via Lemma 2.6. Also, in Tsujii’s estimate (8) the dependence of c_0 on ϵ is not made explicit, and so we don’t obtain a quantitative decay rate in general. While Shmerkin and Mosquera do provide an effective version of (8) for homogeneous IFSs in [31], our method does not appear to significantly improve on the effective bounds they achieve in e.g. [31, Corollary 3.3]. Thus, the question of optimal decay rates remains open.

2 Preliminaries

2.1 Some properties of self-similar measures

We retain the notations setup in Section 1, and in particular in Section 1.5: Specifically, our self-similar IFS $\Phi = \{f_1, ..., f_n\}$, $\mathcal{A} := \{1, .., n\}$, and the use of (9) for symbolic notation. Recall that there is no loss of generality in assuming that

$$K := K_\Phi \subseteq [0, 1].$$

This assumption will be retained throughout the paper.

Recall that we are working with a self-similar measure $\mu = \mu_{\mathrm{p}}$ with respect to a strictly positive probability vector $\mathrm{p} \in \mathcal{P}(\mathcal{A})$. Since K_Φ is assumed to be infinite, this is well known to imply that μ is not atomic. In fact, as shown by Feng and Lau [19], μ satisfies a much stronger Hölder regularity property:

Proposition 2.1 *[19, Proposition 2.2] Let μ be a non-atomic self-similar measure. Then there exists some $s_0 = s_0(\mu) > 0$ such that for all $r > 0$*

$$\sup_{x \in \mathbb{R}} \mu \left(B(x, r) \right) = O\left(r^{s_0} \right).$$

Proposition 2.1 means that the Frostman exponent of μ is non-zero. We remark that it is quite common for this attribute of the measure to appear when estimating the decay of its Fourier transform, see e.g. [5, 7, 28, 31]; our estimates will involve it as well.

Next, recall the definition of the cut-set $\mathcal{W}_b$ as in (10). We denote the set of the corresponding contraction ratios by

$$\mathcal{C}_b := \{ r_\omega : \omega \in \mathcal{W}_b \} \subseteq (-1, 1). \tag{12}$$

We proceed to record a number of standard facts. Recall that $|\mathcal{A}| = n$. For $\omega \in \mathcal{A}^p$ we write $[\omega]$ to denote the set of all $\eta \in \mathcal{A}^{\mathbb{N}}$ such that $\eta|_p = \omega|_p$.

Lemma 2.2 *The following statements hold true for every $b > 0$:*

1. The collection of sets $\{[\eta] : \eta \in \mathcal{W}_b\}$ forms a partition of $\mathcal{A}^{\mathbb{N}}$. Furthermore,

$$|\mathcal{C}_b| = O\left(\log b\right), \text{ where the implicit constant is uniform in } b.$$

2. For any measurable map $g : \mathbb{R} \to \mathbb{R}$,

$$\int_0^1 e^{2\pi i \lambda g(x)} \, d\mu(x) = \widehat{g\mu}\left(\lambda\right) = \sum_{\omega \in \mathcal{W}_b} p_\omega \cdot \widehat{\mu}\left(g\left(f_\omega(x) \right) \lambda \right).$$

Indeed, Part (1) follows from a standard counting argument, whereas Part (2) follows from the stationarity of μ via Part (1).

Next, we discuss linearization. For any $\eta \in \mathcal{A}^*$ recall that we write

$$f_\eta(x) = r_\eta \cdot x + t_\eta, \text{ where } t_\eta := f_\eta(0).$$

We will require the following standard linear approximation lemma. Recall that we assuming μ is supported on [0, 1].

Lemma 2.3 *Let $g \in C^{1+\alpha}([0,1])$ where $\alpha \in (0,1]$. Then for all $x \in [0,1]$ and $\eta \in \mathcal{A}^*$,*

$$g \circ f_\eta(x) - g(t_\eta) - g'(t_\eta) \cdot r_\eta \cdot x = O_g\left(r_\eta^{1+\alpha}\right). \tag{13}$$

Proof There exists a point $y = y(x,\eta) \in [0,1]$ such that

$$g \circ f_\eta(x) = g(t_\eta) + g'(f_\eta(y)) \cdot r_\eta \cdot x.$$

As $g' \in C^\alpha([0,1])$ and $|x|, |y| \le 1$, we thus have

$$|g \circ f_\eta(x) - g(t_\eta) - g'(t_\eta) \cdot r_\eta \cdot x| = |g'(f_\eta(y)) - g'(t_\eta)| \cdot |r_\eta| \cdot |x| \le O_g\left(|f_\eta(y) - f_\eta(0)|^\alpha\right) \cdot |r_\eta| = O_g\left(r_\eta^{1+\alpha}\right).$$

□

2.2 On Tsujii's large deviations estimate

Recalling our discussion from Section 1.5, let us first formulate Tsujii's original result from [40]. Writing $\mathcal{P}(X)$ for the family of Borel probability measures on the space *X*, and recalling that $|\cdot|$ stands for the Lebesgue measure on $\mathbb{R}$, we have:

Theorem 2.4 *[40] Let $\mu \in \mathcal{P}(\mathbb{R})$ be a non-atomic self-similar measure. Then*

$$\lim_{c\to 0^+} \limsup_{t\to\infty} \frac{1}{t} \log \left|\left\{\lambda \in [-e^t, e^t] : |\widehat{\mu}(\lambda)| \ge e^{-c\cdot t}\right\}\right| = 0.$$

In our argument we will require the following corollary of this result:

Corollary 2.5 *Let $\mu \in \mathcal{P}(\mathbb{R})$ be a non atomic self-similar measure. Then*

$$\lim_{c\to 0^+} \limsup_{t\to\infty} \frac{1}{t} \log \sharp \left\{n \in \mathbb{Z} : \exists \lambda \in [n, n+1] \cap [-e^t, e^t] such that\, |\widehat{\mu}(\lambda)| \ge e^{-c\cdot t}\right\} = 0.$$

In particular, for every $\epsilon > 0$ there exists $c_0 > 0$ such that for all *t* large enough,

$$\sharp\left\{n\in\mathbb{Z}:\exists\lambda\in[n,n+1]\cap[-e^t,e^t] such that\, |\widehat{\mu}(\lambda)|\geq e^{-c\cdot t}\right\}\leq e^{\epsilon\cdot t}. \qquad (14)$$

Proof Since μ is a self-similar measure, it is Borel and compactly supported. It is well known that the Fourier transform of such a measure is uniformly Lipschitz continuous [30, Section 3.2].

Thus, if for some $c>0, t>0$ and $\lambda\in\mathbb{R}$ we have

$$|\widehat{\mu}(\lambda)|\geq e^{-c\cdot t},$$

then, assuming *t* is very large (depending on *c*), for every $s\in[\lambda-e^{-2ct},\ \lambda+e^{-2ct}]$ we have

$$|\widehat{\mu}(s)|\geq\frac{1}{2}\cdot e^{-c\cdot t}.$$

This in turn implies that for every $c>0$ and $t=t(c)>0$ large enough,

$$e^{-2ct}\sharp\left\{n\in\mathbb{Z}:\exists\lambda\in[n,n+1]\cap[-e^t,e^t]\text{ such that }|\widehat{\mu}(\lambda)|\geq e^{-c\cdot t}\right\}$$
$$\leq\left|\left\{\lambda\in[-e^t,e^t]:\ |\widehat{\mu}(\lambda)|\geq\frac{1}{2}\cdot e^{-c\cdot t}\right\}\right|.$$

The first assertion of the corollary is a direct consequence of Theorem 2.4 and the last displayed equation. As (14) is a formal consequence of this result, the corollary is proved. □

2.3 Non-affine continuously differentiable functions are non-flat

The following two lemmas demonstrate the non-flatness of non-linear real continuously differentiable functions.

Lemma 2.6 *Let $J\subset\mathbb{R}$ be a closed interval and $h\in C^k(J)$ for some integer $k\geq 1$. If*

$$\max_{1\leq j\leq k}|h^{(j)}(x)|\geq 1 for all x\in J, \qquad (15)$$

then there exists $\rho>0$ such that h is $\frac{1}{2k}$-non-flat on all sub-intervals $I\subseteq J$ with $|I|\leq\rho$.

We remark that in (15) we can replace 1 with any $c_0>0$.

Proof Since $h\in C^k(J)$ there exists $\rho>0$ such that

$$\max_{0\le j\le k} \sup \left\{ |h^{(j)}(x) - h^{(j)}(y)| : x, y \in J,\ |x-y| \le \rho \right\} \le \frac{1}{4} \tag{16}$$

Let $I \subseteq J$ be any sub-interval with $|I| \le \rho$. We claim that h is $\frac{1}{2k}$-non-flat on I.

Suppose that h is not $\frac{1}{2k}$-non-flat on I. We will show that this contradicts our assumption (15). Now, if h is not $\frac{1}{2k}$-non-flat on I, then there exist $\epsilon > 0$, arbitrarily large $n \ge 1$, and $A \subseteq I$ satisfying

$$\mathcal{N}_{2^{-\frac{n}{2k}}}(A) \ge 2^{n\epsilon}, \qquad \text{and } \operatorname{diam}(h(A)) < 2^{-n}.$$

In particular, for n large enough (depending on ϵ), there exist 2^{k+1} points $\{x_i\}_{i=1}^{2^{k+1}} \subseteq A$ such that

$$x_1 < x_2 < ... < x_{2^{k+1}}, \qquad \text{and } \operatorname{dist}(x_i, x_j) \ge 2^{-\frac{n}{2k}} \text{ for all } i \ne j.$$

We proceed to construct $2^k = 2^{k+1}/2$ points $\{y_i^{(1)}\}_{i=0}^{2^k-1}$ as follows. By the previous two displayed equations, for each $0 \le i \le 2^k - 1$ there exists $y_{2i}^{(1)} \in [x_{2i+1}, x_{2i+2}]$ such that

$$|h'(y_{2i}^{(1)})| = \left| \frac{h(x_{2i+1}) - h(x_{2i+2})}{x_{2i+1} - x_{2i+2}} \right| \le 2^{-n} \cdot 2^{\frac{n}{2k}}.$$

Note that the gaps between the (ordered) 2^k points $\{y_{2i}^{(1)}\}_{i=0}^{2^k-1}$ are at least $2^{-\frac{n}{2k}}$. We next construct $2^{k-1} = 2^k/2$ points $\{y_i^{(2)}\}_{i=0}^{2^{k-1}-1}$ as follows: For each $0 \le i \le 2^{k-1} - 1$, there exists $y_{4i}^{(2)} \in [y_{4i}^{(1)}, y_{4i+2}^{(1)}]$ such that

$$|h^{(2)}(y_{4i}^{(2)})| = \left| \frac{h'(y_{4i+2}^{(1)}) - h'(y_{4i+4}^{(1)})}{y_{4i+2}^{(1)} - y_{4i+4}^{(1)}} \right| \le 2^{-n} \cdot 2 \cdot (2^{\frac{n}{2k}})^2.$$

Note the use of the triangle inequality and the previous construction. Continuing this process $k-2$ more times, we can find k points $z_1, z_2, \cdots, z_k$ such that for every $1 \le j \le k$

$$|h^{(j)}(z_j)| \le 2^{-n} \cdot 2^{k-1} \cdot \left(2^{\frac{n}{2k}}\right)^k \le \frac{1}{2}.$$

In view of (16), it follows that

$$\max_{1\le j\le k} \sup_{x\in I} |h^{(j)}(x)| \le \max_{1\le j\le k} \left[|h^{(j)}(z_j)| + \sup_{x\in I} |h^{(j)}(x) - h^{(j)}(z_j)| \right] \le \frac{3}{4} < 1.$$

This is a contradiction to our assumption (15). We thus conclude that h is not $\frac{1}{2k}$-non-flat on I. $\square$

Let us briefly revisit the discussion of the optimality of the exponent $\frac{1}{2k}$ in Lemma 2.6, as presented in Section 1.4. We believe that, with only minor modifications, the factor $\frac{1}{2}$ in the argument above can be replaced by an arbitrary constant $c < 1$. If so, the exponent $\frac{1}{2k}$ could be improved to $\frac{c}{k}$, where $c < 1$ can be taken arbitrarily close to 1. However, it appears that the same argument does not yield the full improvement to $\frac{1}{k}$.

Now, the following lemma is a direct consequence of the previous one:

Lemma 2.7 *Let $g \in C^{\omega}(J)$ be a non-affine real analytic function on a compact interval $J \subseteq \mathbb{R}$. Then there exist $c > 0$ and an integer $k \geq 2$ such that for all $x \in J$,*

$$\max_{2\leq j\leq k} |g^{(k)}(x)| \geq c.$$

In particular, there exists $\rho > 0$ such that $g \cdot \frac{1}{c}$ is $\frac{1}{2k}$-non-flat on all sub-intervals $I \subseteq J$ with $|I| \leq \rho$.

Proof Suppose, towards a contradiction, that there exists a sequence of points $(x_n)_{n\geq 1}$ in J such that $\max_{2\leq j\leq n} |g^{(j)}(x_n)| \leq \frac{1}{n}$ as $n \to \infty$. By compactness, it has an accumulation point $x \in J$. Now the smoothness of g implies that $g^{(j)}(x) = 0$ for all $j \geq 2$. It thus follows that $g(y) = g(x) + g'(x)(y - x)$ on J, which is an affine function. ☐

3 Proof of Theorem 1.6

In this section we prove Theorem 1.6. Let $\mu = \mu_{\mathrm{p}}$ be our self-similar measure. We retain the assumption, without the loss of generality, that μ is supported on [0, 1], and the other notations and definitions as in the preceding sections, and specifically in Sections 2 and 1.5. Let $g \in C^{1+\alpha}([0, 1])$ be such that g' is a δ-non-flat function on [0, 1], and recall that $0 < \delta \leq 1$.

Let $\lambda > 0$. We aim to prove the existence of constants $\tau = \tau(\mu, \delta) > 0$ and $C = C(g) > 0$ such that

$$\left| \int e^{2\pi i g(x)\lambda} d\mu(x) \right| \leq C\lambda^{-\tau}.$$

We will often work with big O-notation, and moreover sometimes even omit that for notational clarity, under the assumption that $\lambda \gg 1$. In particular, suitable constants C can be worked out from our argument.

We fix a parameter $b = b(\lambda) > 0$ to be chosen later. The proof consists of 4 steps:

Step 1: Linearization. Recall the definition of the cut-set (10),

$$\mathcal{W}_b = \{\omega \in \mathcal{A}^* : |r_\omega| \leq b < |r_{\omega^-}|\}.$$

Then we have:

$$
\begin{aligned}
\left|\widehat{g\mu}(\lambda)\right| &= \left|\sum_{\omega\in\mathcal{W}_b} p_\omega \cdot \widehat{\mu}\left(g\left(f_\omega(x)\right)\lambda\right)\right| \\
&= \left|\sum_{\omega\in\mathcal{W}_b} p_\omega \cdot \int e^{2\pi i\lambda\left(g(t_\omega)+g'(t_\omega)\cdot r_\omega\cdot x+O\left(r_\omega^{1+\alpha}\right)\right)}\, d\mu(x)\right| \qquad (17) \\
&\le \sum_{\omega\in\mathcal{W}_b} p_\omega \cdot \left|\widehat{\mu}\left(g'(t_\omega)\cdot r_\omega\cdot\lambda\right)\right| + O\left(b^{1+\alpha}\cdot\lambda\right).
\end{aligned}
$$

Indeed, the first equality is Lemma 2.2 Part (2), and the second one follows from Lemma 2.3. The last one follows from the triangle inequality and since $e^{2\pi i\lambda x}$ is a $2\pi\lambda$-Lipschitz function, and using that by definition $r_\omega = \Theta(b)$ for all $\omega \in \mathcal{W}_b$, where $\Theta(\cdot)$ is the standard big Theta notation.

Step 2: Applying Tsujii's theorem and the non-flatness assumption. Recall definition (12). Fix $r \in \mathcal{C}_b$ and define

$$
\mathcal{W}_{b,r} = \{\omega \in \mathcal{W}_b : r_\omega = r\}. \qquad (18)
$$

Recalling that we are free to choose $b = b(\lambda)$, we add the assumptions that as $b \to 0$, for all $\omega \in \mathcal{W}_b$,

$$
|\lambda\cdot r_\omega| \to \infty \text{ and } |\lambda\cdot r_\omega^{1+\alpha}| \to 0.
$$

Note that this must hold uniformly in $\omega \in \mathcal{W}_b$ by definition. In particular, for every $r \in \mathcal{C}_b$, we may assume $|\lambda\cdot r|$ is arbitrarily large.

Next, fix some $\epsilon > 0$ to be determined later. Since $g \in C^{1+\alpha}([0,1])$, there exists some $C > 1$ such that

$$
|g'(x)| < C \text{ for all } x \in [0,1].
$$

So, applying the Corollary (14) of Tsujii's theorem (Theorem 2.4) with this ϵ, there exists some $c_0 = c_0(\epsilon) > 0$ such that for all λ large enough, putting $t = \log|\lambda\cdot r\cdot C|$ such that it grows to ∞ with λ, we have

$$
\#\left\{\begin{matrix} n\in\mathbb{Z} : \exists\omega\in\mathcal{W}_{b,r} \text{ with } \lambda\cdot g'(t_\omega)\cdot r \in [n,n+1] \\ \text{such that } |\hat{\mu}(\lambda\cdot g'(t_\omega)\cdot r)| \ge |\lambda\cdot r\cdot C|^{-c_0} \end{matrix}\right\}| \le |\lambda\cdot C\cdot r|^{\epsilon} \qquad (19)
$$

The next Lemma is the only place in the proof where the δ-non-flatness assumption is used. Recall that $s_0 > 0$ is the Frostman exponent of the measure as in Proposition 2.1. Also, to streamline notation, let $\mathbb{P} = \mathrm{p}^{\mathbb{N}}$ be the Bernoulli on $\mathcal{A}^{\mathbb{N}}$. Let $\pi : \mathcal{A}^{\mathbb{N}} \to K$ be the coding map $\pi(\eta) = \lim_{n\to\infty} f_{\eta|_n}(0)$, where $\eta|_n \in \mathcal{A}^n$ is the n-prefix of η. It is standard that $\pi\left(\mathcal{A}^{\mathbb{N}}\right) = K$ and that $\pi\mathbb{P} = \mu$. Finally, for $\omega \in \mathcal{A}^*$ recall that $[\omega] = \{\eta \in \mathcal{A}^{\mathbb{N}} : \eta|_{|\omega|} = \omega\}$.

Lemma 3.1 *For every fixed* $n \in \mathbb{Z}$,

$$\mathbb{P}\left([\omega] : \omega \in \mathcal{W}_{k,r}, \lambda \cdot g'(t_\omega) \cdot r \in [n, n+1]\right) = O\left((r \cdot \lambda)^{\frac{\delta \cdot s_0}{2}} \cdot \left((r \cdot \lambda)^{-\delta} + 2r\right)^{s_0}\right).$$

Proof Let

$$A := \{t_\omega : \omega \in \mathcal{W}_{k,r}, \lambda \cdot g'(t_\omega) \cdot r \in [n, n+1]\} = (g')^{-1}\left(\left[\frac{n}{r \cdot \lambda}, \frac{n+1}{r \cdot \lambda}\right]\right)$$
$$\bigcap \{t_\omega : \omega \in \mathcal{W}_{b,r}\}.$$

Then $\operatorname{diam}(g'(A)) \leq \frac{1}{r \cdot \lambda}$. Using the exponent $\epsilon = \frac{s_0}{2}$, we deduce from the δ-non-flatness of g' that

$$\mathcal{N}_{(r \cdot \lambda)^{-\delta}}(A) < (r \cdot \lambda)^{\frac{\delta \cdot s_0}{2}}.$$

Note that here we are assuming λ is sufficiently large. Recalling that $\pi : \mathcal{A}^{\mathbb{N}} \to K$ is the coding map, for every $\omega \in \mathcal{W}_b$ and $\eta \in [\omega]$ we have

$$|\pi(\eta) - t_\omega| = |\pi(\eta) - f_\omega(0)| \leq r.$$

So, if $t_\omega \in I$, where I is an interval of length $(r \cdot \lambda)^{-\delta}$, then $f_\omega([0,1])$ belongs to an interval of size $(r \cdot \lambda)^{-\delta} + 2r$. Thus, combining with our previous estimates,

$$\mathcal{N}_{(r \cdot \lambda)^{-\delta} + 2r}\left(\bigcup_{\omega \in \mathcal{W}_{k,r} :\ \lambda \cdot g'(t_\omega) \cdot r \in [n, n+1]} f_\omega([0,1])\right) < (r \cdot \lambda)^{\frac{\delta \cdot s_0}{2}}.$$

Applying Proposition 2.1 and Boole's inequality, we obtain

$$\mathbb{P}\left([\omega] : \omega \in \mathcal{W}_{k,r}, \lambda \cdot g'(t_\omega) \cdot r \in [n, n+1]\right) = O\left((r \cdot \lambda)^{\frac{\delta \cdot s_0}{2}} \cdot \left((r \cdot \lambda)^{-\delta} + 2r\right)^{s_0}\right).$$

This proves the lemma. □

Step 3: Collecting the error terms. First, recalling (18) and the terms $c_0, C > 0$ from (19), we partition $\mathcal{W}_{k,r}$ into two subsets:

- $\mathcal{W}_{b,r,1}$ is the collection of all $\omega \in \mathcal{W}_{b,r}$ such that

$$|\widehat{\mu}(g'(t_\omega) \cdot r_\omega \cdot \lambda)| \geq |\lambda \cdot r \cdot C|^{-c_0}.$$

- $\mathcal{W}_{b,r,2}$ is the collection of all the other $\omega \in \mathcal{W}_{k,r}$, that is,

$$\mathcal{W}_{b,r,2} = \mathcal{W}_{b,r} \setminus \mathcal{W}_{b,r,1}.$$

Now, for every $r \in \mathcal{C}_k$, applying (19) and Lemma 3.1 with the notation $\mathbb{P}$ as in that lemma yields

$$\begin{aligned}\sum_{\omega \in \mathcal{W}_{b,r}} p_\omega \cdot |\widehat{\mu}(g'(t_\omega)\cdot r\cdot\lambda)| &= \sum_{\omega \in \mathcal{W}_{b,r,1}} p_\omega \cdot |\widehat{\mu}(g'(t_\omega)\cdot r\cdot\lambda)| + \sum_{\omega \in \mathcal{W}_{b,r,2}} p_\omega \cdot |\widehat{\mu}(g'(t_\omega)\cdot r\cdot\lambda)| \\ &\le \sum_{n\in\mathbb{Z}} \mathbb{P}([\omega] : \omega \in \mathcal{W}_{b,r,1} \text{ with } \lambda\cdot t_\omega \cdot r \in [n,n+1]) + \\ &\quad \mathbb{P}([\omega] : \omega \in \mathcal{W}_{b,r,2}) \cdot |\lambda\cdot r\cdot C|^{-c_0} \\ &\le |\lambda\cdot r\cdot C|^{\epsilon} \cdot O\left((r\cdot\lambda)^{\frac{\delta\cdot s_0}{2}} \cdot \left((r\cdot\lambda)^{-\delta} + 2r\right)^{s_0}\right) + |\lambda\cdot r\cdot C|^{-c_0}.\end{aligned}$$

Combining this with (17) and Lemma 2.2 Part (1), noting that $r = \Theta(b)$ and omitting big-O notations, we obtain

$$\begin{aligned}|\widehat{g\mu}(\lambda)| &\le \sum_{\omega\in\mathcal{W}_b} p_\omega \cdot |\widehat{\mu}(g'(t_\omega)\cdot r_\omega\cdot\lambda)| + \left|b^{1+\alpha}\cdot\lambda\right| \\ &= \sum_{r\in\mathcal{C}_b}\sum_{\omega\in\mathcal{W}_{b,r}} p_\omega \cdot |\widehat{\mu}(g'(t_\omega)\cdot r\cdot\lambda)| + \left|b^{1+\alpha}\cdot\lambda\right| \\ &\lesssim |\mathcal{C}_b| \cdot \left(|\lambda\cdot b\cdot C|^{\epsilon} \cdot \left|(b\cdot\lambda)^{\frac{\delta\cdot s_0}{2}} \cdot \left((b\cdot\lambda)^{-\delta} + 2b\right|^{s_0}\right) + |\lambda\cdot b\cdot C|^{-c_0}\right) + \left|b^{1+\alpha}\cdot\lambda\right| \\ &\lesssim \log(b) \cdot \left(|\lambda\cdot b\cdot C|^{\epsilon} \cdot \left|(b\cdot\lambda)^{\frac{\delta\cdot s_0}{2}} \cdot \left((b\cdot\lambda)^{-\delta} + 2b\right)^{s_0}\right) + |\lambda\cdot b\cdot C|^{-c_0}\right) + \left|b^{1+\alpha}\cdot\lambda\right|.\end{aligned} \tag{20}$$

Here, we use the notation $A \lesssim B$ to indicate that B is larger than A up to a uniform multiplicative constant.

Step 4: Choice of parameters and conclusion of the proof. Recall that we were given some λ, and that we may assume $\lambda \gg 1$. We are now in a position to determine b and ϵ; As for the other parameters in (20), note that $c_0 = c_0(\epsilon)$, $s_0 = s_0(\mu)$, and that $C = C(g)$ is a global constant (independent of λ). We also recall that before Lemma 3.1 we added the conditions that as $b \to 0$, uniformly for $\omega \in \mathcal{W}_b$,

$$|\lambda\cdot r_\omega| \to \infty \text{ and } |\lambda\cdot r_\omega^{1+\alpha}| \to 0,$$

which is the same as asking that, as $r_\omega = \Theta(b)$,

$$|b\cdot\lambda| \to \infty \text{ and } |b^{1+\alpha}\cdot\lambda| \to 0 \text{ as } \lambda \to \infty.$$

Recalling that $\alpha > 0$ by our assumption, find any γ such that

$$\frac{1}{1+\alpha} < \gamma < 1.$$

We pick

$$b = \lambda^{-\gamma}, \quad \epsilon = \frac{\delta\cdot s_0}{4}.$$

Note that, as $0 < \gamma, \delta, s_0 \leq 1$,

$$\epsilon(1-\gamma) + \frac{\delta \cdot s_0}{2}(1-\gamma) - s_0 \cdot \gamma \leq \frac{3s_0}{4}(1-\gamma) - s_0 \cdot \gamma < 1.$$

Thus, omitting global multiplicative constants (in particular, the big-O notation, absolute values, and multiplicative constants), via (20) we have

$$\begin{aligned} |\widehat{g\mu}(\lambda)| &\lesssim \log(\lambda) \cdot \left(|\lambda \cdot b|^{\epsilon} \cdot \left((b\cdot\lambda)^{\frac{\delta \cdot s_0}{2}} \cdot \left((b\cdot\lambda)^{-\delta\cdot s_0} + b^{s_0}\right)\right) + |\lambda \cdot b|^{-c_0}\right) + \left|b^{1+\alpha} \cdot \lambda\right| \\ &\lesssim \log(\lambda) \cdot \left(\lambda^{-(1-\gamma)\cdot\frac{s_0 \cdot \delta}{4}} + \lambda^{\epsilon(1-\gamma) + \frac{\delta \cdot s_0}{2}(1-\gamma) - s_0\cdot\gamma} + \lambda^{-c_0\cdot(1-\gamma)}\right) + \lambda^{1-(1+\alpha)\cdot\gamma}. \end{aligned}$$

This proves the existence of τ as in Theorem 1.1. As a constant C can be worked out from this argument by re-considering the implicit constants in the omitted big-O and the $\lesssim$ notations, Theorem 1.1 is proved.

4 Proof of Theorem 1.2

In this section we prove Theorem 1.2. The general idea is to prove a quantitative van der Corput type estimate for a family of polynomials that arise naturally when considering the k-level correlation sums (4), and the Poisson summation formula. Once this theorem is established, we will derive Theorem 1.2 from it via standard arguments.

Let us state the main technical theorem of this section:

Theorem 4.1 *Let $\mu \in \mathcal{P}([a,b])$ be a non-atomic self-similar measure, where $1 < a < b$. Let $k \geq 2$ be an integer, $N \gg 1$ and $0 < \epsilon \ll 1$, and let $g \in \mathbb{Z}[x]$ be a polynomial with integer coefficients of the form*

$$g(x) = \sum_{j=1}^{k-1} l_j (x^{u_j} - x^{u_{j+1}}) + \sum_{j=1}^{k-1} m_j (x^{v_j} - x^{v_{j+1}}), \tag{21}$$

where $|l_j|, |m_j| \in (0, N^{1+\epsilon}]$, $1 \leq u_k < \ldots < u_1 \leq N$, $1 \leq v_k < \ldots < v_1 \leq N$.

Suppose that either $u_1 > v_1$, or $u_1 = v_1$ but $l_1 + m_1 \neq 0$. If $u_1 = \deg(g) \geq N^{\frac{1}{2k}}$, then there exists $\tau > 0$ such that

$$\left| \int e^{2\pi i g(x)} d\mu(x) \right| \leq a^{-N^{\tau}}, \tag{22}$$

where τ is independent of g.

The proof of Theorem 4.1 will take up the majority of this section. We remark that its proof is based upon the scheme of proof of Theorem 1.6. Also, a significant part of our argument is devoted to the study of the polynomials as in (21). This discussion takes place in the next section.

4.1 A covering lemma for a class of polynomials

Let us first define a family of polynomials that will play an essential role in the proof of Theorem 4.1:

Definition 4.2 Let $k \geq 2$ be an integer, and let $N \geq 2$. Define

$$\mathcal{F}_{k,N} := \left\{ \sum_{j=1}^{k} l_j x^{u_j} \,:\, l_j \in \mathbb{Z},\, l_j \neq 0,\, |l_j| \leq N^4, u_j \in \mathbb{Z},\, 0 \leq u_k < \ldots < u_1 \leq N \right\} \subseteq \mathbb{Z}[x]. \tag{23}$$

By Descartes' sign rule, each polynomial in $\mathcal{F}_{k,N}$ has at most $2k$ distinct roots in $\mathbb{R}$ [15, Lemma 3]. We have the following quantitative covering Lemma concerning the elements of $\mathcal{F}_{k,N}$.

Lemma 4.3 *Fix an integer $k \geq 2$ and reals $1 < a < b$. For any $\epsilon > 0$ there exists $\sigma \in (0, 1)$ such that for all large enough N and $h \in \mathcal{F}_{k,N}$, there exist at most $2k + 2$ sub-intervals $I_i \subset [a, b]$ satisfying:*

1. $|I_i| \leq a^{-N^{\sigma}}$ for each i, and
2. $|h(x)| \geq a^{u_1 - N^{\epsilon}}$ for all $x \in [a, b] \setminus \bigcup_i I_i$.

Our proof of Lemma 4.3 relies on the following lemma.

Lemma 4.4 *Fix an integer $k \geq 2$ and reals $1 < a < b$. Let $m \geq 1$ and let $P(x) \in \mathbb{R}[x]$ be a polynomial such that* $\deg(P) = m$ *and:*

$$P(x) = \sum_{j=0}^{m} l_j x^j, \quad |l_m| \geq 1, \quad \sharp\{0 \leq j \leq m : l_j \neq 0\} \leq k.$$

Then for any $q > 2(1 + \log_a 2 + \log_a 3)$ there exist at most $2k + 2$ sub-intervals $I_i \subset [a, b]$ such that:

1. $|I_i| \leq a^{-mq}$ for each i;
2. $|P(x)| \geq a^{-(mq)^2}$ for all $x \in [a, b] \setminus \bigcup_i I_i$.

Proof Let

$$A = \left\{ x \in [a, b] : |P(x)| < a^{-(mq)^2} \right\}.$$

Since P is a polynomial with less than k terms, both P and P' have at most $2k$ real roots. Thus A is the union of at most $2k+2$ intervals. Suppose that there exists an interval $I \subseteq [a,b]$ satisfying

$$|I| > a^{-mq} \text{ and } |P(x)| < a^{-(mq)^2} \text{ for all } x \in I. \tag{24}$$

We divide the interval I into 3^m consecutive sub-intervals of equal length: Initially, the interval I is divided into 3 sub-intervals of equal length, I_1, I_2, I_3, labeled according to their Euclidean order. Then, for each $i \in \{1,2,3\}$, I_i has a similar decomposition $I_1 = I_{i1} \cup I_{i2} \cup I_{i3}$. Continuing inductively, we decompose I into sub-intervals $I_\omega, \omega \in \{1,2,3\}^m$. For each $\omega \in \{1,2,3\}^m$, let us choose an arbitrary point $x_\omega \in I_\omega$. For each $\omega \in \{1,2,3\}^{m-1}$ we have

$$\frac{P(x_{\omega*1}) - P(x_{\omega*3})}{x_{\omega*1} - x_{\omega*3}} = P'(z_\omega)$$

for some point $z_\omega \in \mathcal{I}_\omega$. By (24),

$$|P'(z_\omega)| < \frac{2 \cdot a^{-(mq)^2}}{(a^{-mq} \cdot 3^{-m})}.$$

Similarly, we see via the triangle inequality that for each $\omega \in \{1,2,3\}^{m-2}$, there exists $z_\omega \in I_\omega$ such that

$$|P^{(2)}(z_\omega)| < \frac{2^2 \cdot a^{-(mq)^2}}{(a^{-mq} \cdot 3^{-m})(a^{-mq} \cdot 3^{-m+1})}.$$

Via m iterations of this process, we can find a point $x_0 \in I$ such that

$$|P^{(m)}(x_0)| < \frac{2^m \cdot a^{-(mq)^2}}{(a^{-mq})^m \cdot 3^{-m} \cdot 3^{-m+1} \cdots 3^{-1}} = a^{-(mq)^2 + qm^2 + m\log_a 2 + \frac{m(m+1)}{2}\log_a 3} < 1,$$

where the last inequality follows from the assumptions that $q > 2(1 + \log_a 2 + \log_a 3)$ and $m \geq 1$. This is impossible, since $|l_m| \geq 1$ implies

$$|P^{(m)}(x)| = |l_m| \cdot m! \geq 1, \ \forall x \in [a,b].$$

It follows that each of the $2k+2$ intervals that cover A has diameter at most a^{-mq}, as claimed. □

Proof of Lemma 4.3 Let $\epsilon > 0$ be small, and let $\delta = \epsilon \cdot 3^{-(k+1)}$. We say that a j-tuple $(u_1, \ldots, u_j)$ with $2 \leq j \leq k$ and $0 \leq u_j < \ldots < u_1 \leq N$ is *admissible* if

$$u_i - u_{i+1} \leq N^{3^{i-1}\delta} \text{ for all } 1 \leq i \leq j-1.$$

Additionally, (u_1) is *admissible* by convention.

Now, let $h(x)=\sum_{j=1}^{k} l_j x^{u_i} \in \mathcal{F}_{k,N}$. Let j be the largest integer such that $(u_1,\cdots,u_j)$ is admissible but $(u_1,\cdots,u_j,u_{j+1})$ is not admissible. Note that it might happen that $j=k$, and in such a case, we just put $u_{j+1}=-\infty$. Thus, we always have

$$u_j - u_{j+1} > N^{3^{j-1}\delta}.$$

Since $(u_1,\cdots,u_j)$ is admissible

$$u_i - u_{i+1} \le N^{3^{i-1}\delta},\ \forall 1\le i\le j-1.$$

Let

$$T_1(x) := \sum_{i=1}^{j} l_i x^{u_i} \text{ and } T_2(x) := \sum_{i=j+1}^{k} l_i x^{u_i}, \text{ so that } h(x)=T_1(x)+T_2(x).$$

We rewrite T_1 as

$$T_1(x) = x^{u_j}\cdot\left(l_1 x^{u_1-u_j}+\ldots+l_j\right) = x^{u_j-N^{3^{j-1}\delta}}\cdot x^{N^{3^{j-1}\delta}}\cdot\left(l_1 x^{u_1-u_j}+\ldots+l_j\right). \tag{25}$$

Let $q\in\mathbb{R}$ be such that

$$\left((u_1-u_j)\,q\right)^2 + 5\log_a N = N^{3^{j-1}\delta}. \tag{26}$$

Note that we have

$$u_1-u_j \le N^{\delta}+\cdots+N^{3^{j-2}\delta} \le 2N^{3^{j-2}\delta}.$$

Therefore, as $N\gg 1$,

$$g > 2+2\log_a 2 + 2\log_a 3.$$

Applying Lemma 4.4 to the polynomial $P(x)=l_1x^{u_1-u_j}+\ldots+l_j$, it follows that there are at most $2k+2$ intervals $I_i\subset[a,b]$ satisfying

$$|I_i| \le a^{-(u_1-u_j)q} \text{ for each } i, \tag{27}$$

and

$$|P(x)| \ge a^{-(u_1-u_j)^2q^2} \text{ for all } x\in[a,b]\setminus\cup_i I_i. \tag{28}$$

By (25), (26) and (28), we have

$$\begin{aligned}|T_1(x)| \geq& x^{u_j - N^{3^{j-1}\delta}} a^{N^{3^{j-1}\delta}} a^{-(u_1-u_j)^2 q^2} \\ =& x^{u_j - N^{3^{j-1}\delta}} a^{4\log_a N} \text{ for all } x \in [a,b] \setminus \cup_i I_i.\end{aligned}$$

Since $|l_{j+1}|, \ldots, |l_k| \leq N^4$ and $u_j - u_{j+1} > N^{3^{j-1}\delta}$, for $N \gg k$, we have

$$|T_2(x)| \leq k a^{\log_a N^4} x^{u_i+1} \leq \frac{|T_1(x)|}{2}.$$

It follows that

$$|h(x)| = |T_1(x) + T_2(x)| \geq \frac{1}{2} a^{u_j - N^{3^{j-1}\delta}} a^{5\log_a N} \text{ for all } x \in [a,b] \setminus \cup_i I_i.$$

Recall that $3^{n+1}\delta = \epsilon$, when $N \gg k$, we get

$$|h(x)| \geq a^{u_1 - N^{\epsilon}} \text{ for all } x \in [a,b] \setminus \cup_i \mathcal{I}_i.$$

This is the second assertion.

Finally, the first assertion follows by combining (27), and

$$u_1 - u_j \leq 2N^{3^{j-2}\delta}, \quad \text{and } (u_1 - u_j)q = (N^{3^{j-1}\delta} - 5\log_a N)^{1/2}$$

from which a suitable choice of N^{σ} can be made. □

4.2 Proof of Theorem 4.1

Let $\mu = \mu_{\mathrm{p}}$ be our non-atomic self-similar measure on $[a, b]$, and recall our other notations from the previous sections. Let $g \in \mathbb{Z}[x]$ be a polynomial as in (21), and recall the other parameters and notations from Theorem 4.1. We aim to prove the existence of some $\tau > 0$ such that

$$\left| \int e^{2\pi i g(x)} d\mu(x) \right| \leq a^{-N^{\tau}}, \text{ where } \tau \text{ is independent of } g.$$

We begin with the following reduction:

Lemma 4.5 *If (22) holds whenever $1 < a < b < a^2$, then it holds for arbitrary $1 < a < b$.*

Proof Since Φ is uniformly contracting, there exists some N such that: for every $\omega \in \mathcal{A}^N$ there exists $1 < a(\omega) < b(\omega) < (a(\omega))^2$ that satisfy

$$\text{supp } (f_\omega \mu) \subseteq [a(\omega), b(\omega)].$$

It follows that the self-similar measure $f_\omega \mu$ satisfies (22), possibly with $\tau = \tau(\omega)$ varying across the different $\omega \in \mathcal{A}^N$. By self-similarity, for all g as in Theorem 4.1,

$$\left| \int e^{2\pi i g(x)} d\mu(x) \right|$$

is bounded above by a convex combination over

$$\left\{ \left| \int e^{2\pi i g(x)} \, df_\omega \mu(x) \right| : \omega \in \mathcal{A}^N \right\}.$$

So, Theorem 4.1 holds for μ (with some choice of $\tau > 0$). □

By Lemma 4.5, we may assume that there exists some $0 < \delta_0 < 1$ such that

$$\frac{b}{a^{2\delta_0}} < 1. \tag{29}$$

We fix a parameter $q = q(N)$ to be chosen later, and roughly follow the proof outline of Theorem 1.6:

Step 1: Linearization and choice of q. Recall the definition of the cut-set (10)

$$\mathcal{W}_q = \{\omega \in \mathcal{A}^* : |r_\omega| \le q < |r_{\omega^-}|\}.$$

We have the following linearization lemma:

Lemma 4.6 *We have*

$$\left| \int e^{2\pi i g(x)} d\mu(x) \right| \le \sum_{\omega \in \mathcal{W}_q} p_\omega \left| \widehat{\mu}\left(g'(t_\omega) \cdot r_\omega\right) \right| + O\left(N^2 \cdot u_1^2 \cdot b^{u_1} \cdot q^2\right). \tag{30}$$

Proof By (a minor variation of) Lemma 2.2, as Φ is a self-similar IFS

$$\int e^{2\pi i g(x)} d\mu(x) = \sum_{\omega \in \mathcal{W}_q} p_\omega \int e^{2\pi i g(f_\omega(x))} d\mu(x). \tag{31}$$

Recall that

$$g(x) = \sum_{j=1}^{k-1} l_j (x^{u_j} - x^{u_{j+1}}) + \sum_{j=1}^{k-1} m_j (x^{v_j} - x^{v_{j+1}}),$$

with $l_j, m_j \in \mathbb{Z}, 0 < |l_j|, |m_j| \le N^{1+\epsilon}$ and $1 \le u_k < \ldots < u_1 \le N, 1 \le v_k < \ldots < v_1 \le N$, where $k \ge 2, N \gg 1, 0 < \epsilon \ll 1$ are fixed and the coefficient of the leading term in $g(x)$ is non-zero. Note that $g(x)$

consists of at most $2k$ distinct terms. Moreover, we have $g(x), g'(x), g''(x) \in \mathcal{F}_{2k,N}$. Using the Taylor expansion and the assumption on g, for every $\omega \in \mathcal{A}^*$ we have

$$\begin{aligned} g(f_\omega(x)) =& g(f_\omega(0)) + g'(f_\omega(0))(f_\omega(x) - f_\omega(0)) + O(||g''||_\infty)(f_\omega(x) - f_\omega(0))^2) \\ =& g(t_\omega) + g'(t_\omega) \cdot r_\omega \cdot x + O(N^2 u_1^2 b^{u_1} r_\omega^2). \end{aligned} \tag{32}$$

The lemma follows by combining (31), (32), the Lipschitz property of the complex exponent, and the fact that $r_\omega \leq q$ for $\omega \in \mathcal{W}_q$, similarly to the first step in the proof of Theorem 1.6. □

We now choose

$$q = \left(N^2 a^{u_1}\right)^{-\delta_0}. \tag{33}$$

So, by (29) we can find some $\delta_1 > 0$ such that

$$N^2 u_1^2 b^{u_1} q^2 = O\left(a^{-u_1 \delta_1}\right).$$

Thus, as a corollary of Lemma 4.6,

$$\left| \int e^{2\pi i g(x)} d\mu(x) \right| \leq \sum_{\omega \in \mathcal{W}_q} p_\omega \left| \widehat{\mu}\left(g'(t_\omega) \cdot r_\omega\right) \right| + O\left(a^{-u_1 \delta_1}\right). \tag{34}$$

Step 2: Applying Tsujii's Theorem and the covering lemma (Lemma 4.3). Recall definitions (10) and (12). Fix $r \in \mathcal{C}_q$ and define $\mathcal{W}_{q,r}$ as in (18),

$$\mathcal{W}_{q,r} = \{\omega \in \mathcal{W}_q : r_\omega = r\}.$$

Next, fix some $\beta > 0$ to be determined later. Applying the Corollary (14) of Tsujii's Theorem 2.4 with this β, there exists some $c_0 = c_0(\beta) > 0$ such that for all b large enough, putting $t = \log_a(2u_1\delta_0)$ such that it grows to ∞ by the assumption $u_1 \geq N^{\frac{1}{2k}}$ as $N \gg 1$, we have

$$\left| \left\{ n \in \mathbb{Z} : \exists \lambda \in [n, n+1] \cap [-a^{2\delta_0 u_1}, a^{2\delta_0 u_1}] \text{ s.t. } |\widehat{\mu}(\lambda)| \geq a^{-c_0 u_1} \right\} \right| \leq a^{\beta u_1}. \tag{35}$$

We now deal, as before, with the $\mathbb{P}$ mass of the bad frequencies (we use the notation as in Lemma 3.1). This is where we use our work from Section 4.1: Applying Lemma 4.3 to $g'(x)$ and $g''(x)$, we know that there exist $\sigma > 0$ and at most $8k+8$ intervals $I_i \subset [a,b]$ with $|I_i| \leq a^{-N^\sigma}$ such that

$$|g'(x)| \geq a^{u_1 - N^\epsilon} \text{ and } |g''(x)| \geq a^{u_1 - N^\epsilon} \text{ for all } x \in [a,b] \setminus \cup_i I_i. \tag{36}$$

Let us now define:

$$\mathcal{L}_{q,r} = \mathcal{W}_{q,r} \setminus \{\omega \in \mathcal{W}_{q,r} : t_\omega \in \bigcup_{i=1}^{8k+8} I_i\}. \tag{37}$$

Recall that $s_0 > 0$ is the Frostman exponent of the measure μ as in Proposition 2.1.

Lemma 4.7 *For every fixed* $n \in \mathbb{Z}$,

$$\mathbb{P}\left([\omega] : \omega \in \mathcal{L}_{k,r}, g'(t_\omega) \cdot r \in [n, n+1]\right) = O\left(\left(a^{-u_1(1-\delta_0)+N^\epsilon} \cdot N^{2\delta_0} + r\right)^{s_0}\right).$$

Proof First, we claim that for any connected component J of $[a, b] \setminus \left(\bigcup_{i=1}^{8k+8} I_i\right)$, the set

$$\{t_\omega : \omega \in \mathcal{L}_{q,r},\ t_\omega \in J, \text{ and } g'(t_\omega)r \in [n,\ n+1]\}$$

is contained in an interval of length at most $\frac{2}{r_{\min}} \cdot a^{-u_1(1-\delta_0)+N^\epsilon} \cdot N^{2\delta_0}$.

Indeed, it follows from (36) that for $\omega, \eta \in \mathcal{L}_{k,r}$, if

$$|t_\omega - t_\eta| > c \text{ for some } c > 0, \text{ and } [t_\omega, t_\eta] \cap \cup_i I_i = \emptyset \text{ assuming } t_\omega < t_\eta,$$

then

$$|g'(t_\omega)r - g'(t_\eta)r| \geq a^{u_1 - N^\epsilon} c|r|. \tag{38}$$

So, if $c = \frac{2}{r_{\min}} \cdot a^{-u_1(1-\delta_0)+N^\epsilon} \cdot N^{2\delta_0}$, we see that, by (33) and since $|r| \geq q \cdot r_{\min}$

$$|g'(t_\omega)r - g'(t_\eta)r| \geq 2. \tag{39}$$

This is impossible since $\operatorname{diam}([n,\ n+1]) = 1$ and $g'(t_\omega)r,\ g'(t_\eta)r \in [n, n+1]$ by our assumptions.

Thus, invoking Proposition 2.1 and arguing similarly as in Lemma 3.1, we obtain

$$\mathbb{P}\left([\omega] : \omega \in \mathcal{L}_{k,r}, g'(t_\omega) \cdot r \in [n, n+1]\right) = O\left(\left(a^{-u_1(1-\delta_0)+N^\epsilon} \cdot N^{2\delta_0} + r\right)^{s_0}\right).$$

□

Step 3: Collecting the error terms and conclusion of the proof. Recalling (37) and the term $c_0 > 0$ from (35), we partition $\mathcal{W}_{k,r}$ into three subsets:

- $\mathcal{W}_{q,r,1}$ is the collection of all $\omega \in \mathcal{L}_{q,r}$ such that

$$|\widehat{\mu}\left(g'(t_\omega) \cdot r\right)| \geq a^{-c_0 \cdot u_1}.$$

- $\mathcal{W}_{q,r,2}$ is the collection of all the other $\omega \in \mathcal{L}_{q,r}$, that is,

$$\mathcal{W}_{q,r,2} = \mathcal{L}_{q,r} \setminus \mathcal{W}_{q,r,1}.$$

- $\mathcal{W}_{q,r,3}$ is the family of all the other $\omega \in \mathcal{W}_{q,r}$,

$$\mathcal{W}_{q,r,3} = \mathcal{W}_{q,r} \setminus \mathcal{L}_{q,r}.$$

Now, for any $\omega \in \mathcal{W}_{q,r}$ and N large enough,

$$|g'(t_I)r_I| \leq 4kN^{1+\epsilon}b^{u_1} \leq a^{2\delta_0 u_1}.$$

This is important, since we are now in a position to apply (35) to our $\mathcal{W}_{q,r,1}$, as we do below. Note that, due to the properties of the coding map π (as in e.g. Lemma 3.1), $\{\pi([\omega])\}_{\omega\in\mathcal{W}_{q,r,3}}$ is contained in at most $8k+8$ intervals of length less than $a^{-N^\sigma}+r$. Applying (34), and omitting global multiplicative constants from notation, using Lemma 2.2 to estimate the size of $\mathcal{C}_q$, we get

$$\begin{aligned}
\left|\int e^{2\pi i g(x)}d\mu(x)\right| &\leq \sum_{\omega\in\mathcal{W}_b} p_\omega \left|\widehat{\mu}\left(g'(t_\omega)\cdot r_\omega\right)\right| + a^{-u_1\delta_1} \\
&\lesssim \sum_{r\in\mathcal{C}_q}\left(\left|\sum_{\omega\in\mathcal{W}_{q,r,1}} p_\omega\widehat{\mu}\left(g'(t_\omega)r\right)\right| + \left|\sum_{\omega\in\mathcal{W}_{q,r,2}} p_\omega\widehat{\mu}\left(g'(t_\omega)r\right)\right| + \left|\sum_{\omega\in\mathcal{W}_{q,r,3}} p_\omega\widehat{\mu}\left(g'(t_\omega)r\right)\right|\right) \\
&\quad + a^{-u_1\delta_1} \\
&\lesssim \sum_{r\in\mathcal{C}_q}\left(\sum_{n\in\mathbb{Z}}\left(\sum_{\substack{\omega\in\mathcal{W}_{q,r,1}\\ g'(t_\omega)r_\omega\in[n,n+1]}} p_\omega\right) + \sum_{\omega\in\mathcal{W}_{q,r,2}} p_\omega a^{-c_0u_1} + \sum_{\omega\in\mathcal{W}_{q,r,3}} p_\omega\right) + a^{-u_1\delta_1} \\
&\lesssim \log(q^{-1})\cdot\left(a^{\beta u_1}\cdot\left(a^{-u_1(1-\delta_0)+N^\epsilon}\cdot N^{2\delta_0} + \left(N^2a^{u_1}\right)^{-\delta_0}\right)^{s_0}\right. \\
&\quad \left. + a^{-c_0u_1} + (8k+8)\cdot\left(a^{-N^\sigma} + \left(N^2a^{u_1}\right)^{-\delta_0}\right)^{s_0}\right) + a^{-u_1\delta_1}.
\end{aligned}$$

Here we have used the notation $\lesssim$ similarly to (20). Putting $\beta < \min\{\delta_0,\, 1-\delta_0\}$ and recalling that $u_1 \geq N^{\frac{1}{2k}}$ and $\epsilon \ll 1$, we get

$$\left|\int e^{2\pi i g(x)}d\mu(x)\right| \lesssim a^{-N^\tau}$$

for some constant τ which is independent of the polynomial g. The proof is complete. □

4.3 Proof of Theorem 1.2

Theorem 1.2 will be deduced from Theorem 4.1 via the following Proposition. It works in significantly greater generality, and is due to Technau and Yesha [39]. Recall the definitions and the notations as in Section 1.2 and Theorem 1.2. In particular, μ is our non-atomic self-similar measure.

Proposition 4.8 *[39, Proposition 7.1] Let $k \geq 2$ and let $C_k(N)$ be a sequence such that $\lim_{N\to\infty} C_k(N) = 1$. Suppose that there exists some $\tau > 0$ such that for all $f \in C_c^\infty\left(\mathbb{R}^{k-1}\right)$ we have*

$$\int_J \left(R_k(f, (x^n \mod 1)_{n\geq 1}, N) - C_k(N) \cdot \int_{\mathbb{R}^{k-1}} f(\mathrm{x})d\mathrm{x} \right)^2 d\mu(x) = O\left(N^{-\tau}\right), \, asN \to \infty,$$

where μ is a supported on a bounded interval *J*.

Then for μ-a.e. x the sequence $(x^n \mod 1)_{n\geq 1}$ has Poissonian k-point correlations.

We note that Proposition 4.8 is stated and proved in [39, Proposition 7.1] for the Lebesgue measure on *J*, but it is not hard to see that the argument works for any Borel probability measure on *J*.

Proof of Theorem 1.2 With no loss of generality, we assume $\xi = 1$, and the other cases are similar. For any $f \in C_c^\infty(\mathbb{R}^{k-1})$ let $\widehat{f}$ denote its Fourier transform, and, recalling the definitions from Section 1.2, put

$$C_k(N) := \frac{\sharp\mathcal{U}_k}{N^k} = \left(1 - \frac{1}{N}\right) \cdots \left(1 - \frac{k-1}{N}\right).$$

Let $\epsilon > 0$ be an auxiliary parameter to be decided later. By Poisson summation we have

$$\begin{aligned}
R_k(f, \{x^n\})_{n\geq 1}, N) =& \frac{1}{N^k} \sum_{\mathrm{l}\in\mathbb{Z}^{k-1}} \widehat{f}\left(\frac{\mathrm{l}}{N}\right) \sum_{\mathrm{u}\in\mathcal{U}_k} e^{-2\pi i \sum_{j=1}^{k-1} l_j (x^{u_j} - x^{u_{j+1}})} \\
=& C_k(N)\widehat{f}(0) + \frac{1}{N^k} \sum_{\mathrm{l}\in\mathbb{Z}^{k-1}\setminus\{0\}} \widehat{f}\left(\frac{\mathrm{l}}{N}\right) \sum_{\mathrm{u}\in\mathcal{U}_k} e^{-2\pi i \sum_{j=1}^{k-1} l_j (x^{u_j} - x^{u_{j+1}})} \\
=& C_k(N)\widehat{f}(0) + \frac{1}{N^k} \sum_{\substack{\mathrm{l} \in \mathbb{Z}^{k-1} \\ 0 < \|\mathrm{l}\|_\infty \leq N^{1+\epsilon}}} \widehat{f}\left(\frac{\mathrm{l}}{N}\right) \sum_{\mathrm{u}\in\mathcal{U}_k} e^{-2\pi i \sum_{j=1}^{k-1} l_j (x^{u_j} - x^{u_{j+1}})} \\
&+ \frac{1}{N^k} \sum_{\substack{\mathrm{l} \in \mathbb{Z}^{k-1} \\ \|\mathrm{l}\|_\infty > N^{1+\epsilon}}} \widehat{f}\left(\frac{\mathrm{l}}{N}\right) \sum_{\mathrm{u}\in\mathcal{U}_k} e^{-2\pi i \sum_{j=1}^{k-1} l_j (x^{u_j} - x^{u_{j+1}})},
\end{aligned}$$

where $0 < \|\mathrm{l}\|_\infty \leq N^{1+\epsilon}$ means $0 < |l_j| \leq N^{1+\epsilon}$ for each *j*. We rewrite the last displayed equation as

$$C_k(N) \cdot \widehat{f}(0) + X_N(x) + Y_N(x).$$

Since $f \in C^\infty(\mathbb{R})$, $\widehat{f}$ decays faster than any polynomial, so there exists $\alpha > 0$ such that $|Y_N(x)| = O(N^{-\alpha})$ uniformly in $x \in \operatorname{spt}(\mu)$. Next, we have

$$\widehat{f}(0) = \int_{\mathbb{R}^{k-1}} f(\mathrm{x})d\mathrm{x}, \text{ and } \lim_{N\to\infty} C_k(N) = 1.$$

Thus, applying Proposition 4.8, it remains to show that $\lim_{N\to\infty} X_N = 0$ in $L^2(\mu)$ with a polynomial rate of convergence. This is proved below.

Let

$$\tilde{\mathcal{U}}_k = \{\mathrm{u} = (u_1, \dots, u_k) : u_i \in \{1, ..., N\},\ 1 \le u_k < \dots < u_1 \le N\}.$$

By symmetry we have

$$\begin{aligned}
&\int |X_N(x)|^2 d\mu(x) \\
&= \int \Big| \frac{1}{N^k} \sum_{0<\|\mathrm{l}\|_\infty \le N^{1+\epsilon}} \widehat{f}\left(\frac{\mathrm{l}}{N}\right) \sum_{\mathrm{u}\in\mathcal{U}_k} e^{-2\pi i \sum_{j=1}^{k-1} l_j (x^{u_j} - x^{u_{j+1}})} \Big|^2 d\mu(x) \\
&= \frac{1}{N^{2k}} \sum_{0<\|\mathrm{l}\|_\infty, \|\mathrm{m}\|_\infty \le N^{1+\epsilon}} \widehat{f}\left(\frac{\mathrm{l}}{N}\right) \widehat{f}\left(\frac{\mathrm{m}}{N}\right) \sum_{\mathrm{u},\mathrm{v}\in\mathcal{U}_k} \int e^{2\pi i g_{\mathrm{l},\mathrm{m},\mathrm{u},\mathrm{v}}(x)} d\mu(x) \\
&\le \frac{2(k!)^2}{N^{2k}} \sum_{0<\|\mathrm{l}\|_\infty, \|\mathrm{m}\|_\infty \le N^{1+\epsilon}} \widehat{f}\left(\frac{\mathrm{l}}{N}\right) \widehat{f}\left(\frac{\mathrm{m}}{N}\right) \sum_{u_1 \ge v_1, \mathrm{u},\mathrm{v}\in\tilde{\mathcal{U}}_k} \int e^{2\pi i g_{\mathrm{l},\mathrm{m},\mathrm{u},\mathrm{v}}(x)} d\mu(x),
\end{aligned} \tag{40}$$

where

$$g_{\mathrm{l},\mathrm{m},\mathrm{u},\mathrm{v}}(x) =: \sum_{j=1}^{k-1} l_j (x^{u_j} - x^{u_{j+1}}) + \sum_{j=1}^{k-1} m_j (x^{v_j} - x^{v_{j+1}}).$$

We decompose the region of of parameters

$$\Omega = \left\{(\mathrm{l}, \mathrm{m}, \mathrm{u}, \mathrm{v}) : 0 < \|\mathrm{l}\|_\infty, \|\mathrm{m}\|_\infty \le N^{1+\epsilon}, u_1 \ge v_1, \mathrm{u}, \mathrm{v} \in \tilde{\mathcal{U}}_k\right\}$$

into three parts:

$$\begin{aligned}
\Omega_1 &= \left\{(\mathrm{l}, \mathrm{m}, \mathrm{u}, \mathrm{v}) \in \Omega : u_1 \le N^{\frac{1}{2k}}\right\}, \\
\Omega_2 &= \left\{(\mathrm{l}, \mathrm{m}, \mathrm{u}, \mathrm{v}) \in \Omega : u_1 > N^{\frac{1}{2k}}, u_1 = v_1, l_1 + m_1 = 0\right\}, \\
\Omega_3 &= \Omega \setminus \cup_{i=1}^2 \Omega_i.
\end{aligned}$$

Then

$$\frac{1}{N^{2k}}\left|\sum_{(\mathrm{l,m,u,v})\in\Omega_1}\widehat{f}\left(\frac{\mathrm{l}}{N}\right)\widehat{f}\left(\frac{\mathrm{m}}{N}\right)\int e^{2\pi i g_{\mathrm{l,m,u,v}}(x)}d\mu(x)\right|\leq\frac{1}{N^{2k}}$$
$$\cdot N^{2(k-1)(1+\epsilon)}\cdot(N^{\frac{1}{2k}})^{2k}=\frac{1}{N^{1-2k\epsilon}}. \tag{41}$$

By taking $\epsilon<\frac{1}{2k}$ we see that the term corresponding to summation over Ω_1 decays polynomially. Next, by Theorem 4.1,

$$\frac{1}{N^{2k}}\left|\sum_{(\mathrm{l,m,u,v})\in\Omega_3}\widehat{f}\left(\frac{\mathrm{l}}{N}\right)\widehat{f}\left(\frac{\mathrm{m}}{N}\right)\int e^{2\pi i g_{\mathrm{l,m,u,v}}(x)}d\mu(x)\right|\leq\frac{1}{N^{2k}}$$
$$\cdot N^{2(k-1)(1+\epsilon)}\cdot N^{2k}\cdot a^{-N^{\tau}}=O\left(a^{-N^{\tau/2}}\right). \tag{42}$$

For the term corresponding to summation over Ω_2, we can compare the values of u_2, v_2 and proceed similarly as in the estimates (41) and (42) to conclude that

$$\frac{1}{N^{2k}}\left|\sum_{(\mathrm{l,m,u,v})\in\Omega_2}\widehat{f}\left(\frac{\mathrm{l}}{N}\right)\widehat{f}\left(\frac{\mathrm{m}}{N}\right)\int e^{2\pi i g_{\mathrm{l,m,u,v}}(x)}d\mu(x)\right|=O\left(N^{-c}\right),\ \text{for some } c\in(0,1). \tag{43}$$

Plugging (41)-(43) into (40), we see that $\int|X_N(x)|^2d\mu(x)$ decays at a polynomial speed. Theorem 1.2 now follows directly from Proposition 4.8. □

Acknowledgements We thank Federico Rodriguez Hertz, Zhiren Wang, Simon Baker, Amlan Banaji, Hongfei Cui, Osama Khalil, and the anonymous referee, for many helpful suggestions and comments. The first author is supported by Grant No. 2022034 from the United States - Israel Binational Science Foundation (BSF), Jerusalem, Israel. The last three authors are supported by the Academy of Finland, project grant No. 318217. The second author was partially supported by NSFC 11901204, 12271418, and 12481540195.

Funding Open access funding provided by University of Haifa.

Data availability There is no data associated with this manuscript.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no Conflict of interest.

References

1. Aistleitner, C., Baker, S.: On the pair correlations of powers of real numbers. Israel J. Math. **242**(1), 243–268 (2021)
2. Aistleitner, C., Baker, S., Technau, N., Yesha, N.: Gap statistics and higher correlations for geometric progressions modulo one. Math. Ann. **385**(1–2), 845–861 (2023)
3. Aistleitner, C., Lachmann, T., Pausinger, F.: Pair correlations and equidistribution. J. Number Theory **182**, 206–220 (2018)
4. Algom, A., Baker, S., Shmerkin, P.: On normal numbers and self-similar measures. Adv. Math. **399**, 108276 (2022)
5. Algom, A., Hertz, F.R., Wang, Z.: Logarithmic Fourier decay for self conformal measures. J. Lond. Math. Soc. **106**(2), 1628–1661 (2022)
6. Algom, A., Hertz, F.R., Wang, Z.: Polynomial Fourier decay and a cocycle version of Dolgopyat's method for self conformal measures. *arXiv preprint* arXiv:2306.01275, (2023)
7. Algom, A., Hertz, F.R., Wang, Z.: Pointwise normality and Fourier decay for self-conformal measures. Adv. Math. **393**, 108096 (2021)
8. Avila, A., Bochi, J., Yoccoz, J.-C.: Uniformly hyperbolic finite-valued [CDATA[\text{ SL }(2,\ mathbb{R})]] $\mathrm{SL}\ (2,\mathbb{R})$-cocycles. Comment. Math. Helv. **85**(4), 813–884 (2010)
9. Baker, S.: Equidistribution results for self-similar measures. Int. Math. Res. Not. IMRN **16**, 12378–12401 (2022)
10. Baker, S., Banaji, A.: Polynomial Fourier decay for fractal measures and their pushforwards. Math. Ann. **392**(1), 209–261 (2025)
11. Baker, S., Sahlsten, T.: Spectral gaps and Fourier dimension for self-conformal sets with overlaps. *arXiv preprint* arXiv:2306.01389, (2023)
12. Bourgain, J., Dyatlov, S.: Fourier dimension and spectral gaps for hyperbolic surfaces. Geom. Funct. Anal. **27**(4), 744–771 (2017)
13. Brémont, J.: Self-similar measures and the Rajchman property. Ann. H. Lebesgue **4**, 973–1004 (2021)
14. Bufetov, A.I., Solomyak, B.: On the modulus of continuity for spectral measures in substitution dynamics. Adv. Math. **260**, 84–129 (2014)
15. Cucker, F., Koiran, P., Smale, S.: A polynomial time algorithm for Diophantine equations in one variable. J. Symbolic Comput. **27**(1), 21–29 (1999)
16. Dai, X.-R.: When does a Bernoulli convolution admit a spectrum? Adv. Math. **231**(3–4), 1681–1693 (2012)
17. Dai, X.-R., Feng, D.-J., Wang, Y.: Refinable functions with non-integer dilations. J. Funct. Anal. **250**(1), 1–20 (2007)
18. de Melo, W., van Strien, S.: One-dimensional dynamics. Springer-Verlag, Berlin (1993)
19. Feng, D.-J., Lau, K.-S.: Multifractal formalism for self-similar measures with weak separation condition. J. Math. Pures Appl. **92**(4), 407–428 (2009)
20. Grepstad, S., Larcher, G.: On pair correlation and discrepancy. Arch. Math. (Basel) **109**(2), 143–149 (2017)
21. Jordan, T., Sahlsten, T.: Fourier transforms of Gibbs measures for the Gauss map. Math. Ann. **364**(3–4), 983–1023 (2016)
22. Kaufman, R.: On Bernoulli convolutions. In: Conference in modern analysis and probability (New Haven, Conn., 1982), volume 26 of Contemp. Math., pages 217–222. Amer. Math. Soc., Providence, RI, 1984
23. Osama Khalil. Exponential mixing via additive combinatorics. *arXiv preprint* arXiv:2305.00527, 2023
24. Kurlberg, P., Rudnick, Z.: The distribution of spacings between quadratic residues. Duke Math. J. **100**(2), 211–242 (1999)
25. Li, J.: Decrease of Fourier coefficients of stationary measures. Math. Ann. **372**(3–4), 1189–1238 (2018)
26. Li, J.: Fourier decay, renewal theorem and spectral gaps for random walks on split semisimple Lie groups. Ann. Sci. Éc. Norm. Supér. **55**(6), 1613–1686 (2022)
27. Li, J., Naud, F., Pan, W.: Kleinian Schottky groups, patterson-sullivan measures, and fourier decay. Duke Math. J. **170**(4), 775–825 (2021). (**With an appendix by Li**)

28. Li, J., Sahlsten, T.: Trigonometric series and self-similar sets. J. Eur. Math. Soc. (JEMS) **24**(1), 341–368 (2022)
29. Marklof, J.: Pair correlation and equidistribution on manifolds. Monatsh. Math. **191**(2), 279–294 (2020)
30. Mattila, P.: Fourier analysis and Hausdorff dimension. Cambridge Studies in Advanced Mathematics, vol. 150. Cambridge University Press, Cambridge (2015)
31. Mosquera, C.A., Shmerkin, P.S.: Self-similar measures: asymptotic bounds for the dimension and Fourier decay of smooth images. Ann. Acad. Sci. Fenn. Math. **43**(2), 823–834 (2018)
32. Naud, F.: Expanding maps on Cantor sets and analytic continuation of zeta functions. Ann. Sci. École Norm. Sup. **38**(1), 116–153 (2005)
33. Rapaport, A.: On the Rajchman property for self-similar measures on $\mathbb{R}^d$[CDATA[{\mathbb{R}}^d]]. Adv. Math. **403**, 108375 (2022)
34. Sahlsten, T.: Fourier transforms and iterated function systems. *arXiv preprint* arXiv:2311.00585, (2023)
35. Sahlsten, T., Stevens, C.: Fourier transform and expanding maps on Cantor sets. Am. J. Math. **146**(4), 945–982 (2024)
36. Solomyak, B.: Fourier decay for self-similar measures. Proc. Amer. Math. Soc. **149**(8), 3277–3291 (2021)
37. Stein, E.M.: Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals, volume 43 of Princeton Mathematical Series. Princeton University Press, Princeton, NJ, (1993). With the assistance of Timothy S. Murphy, Monographs in Harmonic Analysis, III
38. Streck, L.: On absolute continuity and maximal garsia entropy for self-similar measures with algebraic contraction ratio. *arXiv preprint* arXiv:2303.07785, (2023)
39. Technau, N., Yesha, N.: On the correlations of $n^\alpha {\rm mod}\;\;1$[CDATA[n^\alpha {\rm mod }\;\;1]]. J. Eur. Math. Soc. (JEMS) **25**(10), 4123–4154 (2023)
40. Tsujii, M.: On the Fourier transforms of self-similar measures. Dyn. Syst. **30**(4), 468–484 (2015)
41. Varjú, P.P., Han, Yu.: Fourier decay of self-similar measures and self-similar sets of uniqueness. Anal. PDE **15**(3), 843–858 (2022)
42. Yoccoz, J.-C.: Some questions and remarks about $\text{ SL }(2,{\mathbb{R}})$[CDATA[\text{ SL }(2,{\mathbb{R}})]] cocycles. In: Modern dynamical systems and applications, pages 447–458. Cambridge Univ. Press, Cambridge, (2004)

V

SCALING PROPERTIES OF INVARIANT MEASURES FOR CONFORMAL ITERATED FUNCTION SYSTEMS

MENG WU AND YU-LIANG WU

ABSTRACT. Let $\{f_i\}_{i\in\Lambda}$ be a finite conformal iterated function system (IFS) on $\mathbb{R}^d$. Let $(\Lambda^{\mathbb{N}}, \sigma)$ denote the full shift over the alphabet Λ, and let $\pi : \Lambda^{\mathbb{N}} \to \mathbb{R}^d$ be the coding map associated with the IFS. We prove that the projection of an ergodic σ-invariant measure on $\Lambda^{\mathbb{N}}$ under π is always scaling, and the associated tangent distributions are ergodic. Additionally, it is shown that the limits of their scenery flows exhibit a structured form involving orthogonal transformations. This complements the recent advancements in the study of quasi-Bernoulli and self-conformal measures.

1. INTRODUCTION

Let μ be a Radon measure on $\mathbb{R}^d$ and $B_r = B(0, r)$ denote the closed r-ball centered at the origin. Define the *magnification* of x by e^t by

$$\mu_{x,t}(A) = \frac{\mu(e^{-t}(A \cap B_1) + x)}{\mu(B(x, e^{-t}))}.$$

The family $(\mu_{x,t})_{t\geq 0}$ is called the *scenery* of μ at x, associated with which is a family of measures

$$\frac{1}{T}\int_0^T \delta_{\mu_{x,t}} dt, T \geq 0$$

called the *scenery flow* of μ at x. The limit of the scenery flow needs not exist in general, but if it were to exist for μ-a.e. x, such a measure μ is said to be a *scaling measure*, and it is said to be a *uniformly scaling measure* if such a limit exists independently of the choice of the point.

Uniform scaling measures were introduced by Gavish [10], though related concepts and examples had been explored earlier by many other authors, including [2, 11, 19]. A central notion in this area is that of fractal distributions—the limits of scenery flows—which finds its roots in Furstenberg's CP-process theory [8], concerning the dynamics of fractal measure rescaling.

The scenery flow has become a fundamental tool in analyzing the local geometric structure of fractal measures thanks to seminal contributions by Furstenberg [8, 9] and Hochman and Shmerkin [13, 15], among others. In particular, measures whose scenery flow converge at almost every point exhibit significantly more regular geometric behavior than the others. For example, it has been proved recently in [23] that if μ is a scaling measure on $\mathbb{R}^2$, then it satisfies a strong form of Marstrand's projection theorem: There exists a countable exceptional set E of orthogonal projections from $\mathbb{R}^2$ to $\mathbb{R}$ such that for all $\pi \in E^c$,

$$\dim_H \pi\mu = \min(1, \dim_H \mu).$$

The scaling properties of measures also have applications beyond geometry, such as in the study of normal numbers with respect to fractal measures (see, e.g., [16]).

Motivated by these advances, it is natural to seek a deeper understanding of the scenery flow for various classes of fractal measures. Notable examples of scaling measures include $\times m$-invariant measures on the torus $\mathbb{T}$ [12, 7], self-similar measures satisfying the open set condition, the occupation measure of Brownian motion in dimensions $d \geq 3$ [10], and natural measures on random fractals [13].

In this paper we shall focus on measures arising from iterated function systems (IFS). Recall that an IFS $\Phi = (f_i)_{i\in\Lambda}$ is a finite collection of contracting maps on $\mathbb{R}^d$. The IFS Φ is called self-conformal (respectively, self-similar) if the maps $(f_i)_{i\in\Lambda}$ are conformal maps (respectively, similitude). Let $(\Lambda^{\mathbb{N}}, \sigma)$ denote the full shift over the alphabet Λ, and let $\pi : \Lambda^{\mathbb{N}} \to \mathbb{R}^d$ be the coding map associated with the IFS Φ. We refer to Section 2.2 for details about these notions.

Theorem 1.1. *Suppose Φ is a conformal iterated function system on $\mathbb{R}^d$. If μ is an ergodic σ-invariant measure on $\Lambda^{\mathbb{N}}$, then $\pi\mu$ is a scaling measure. Moreover, there exists an ergodic fractal distribution $P \in \mathcal{P}(\mathcal{P}(\mathbb{R}^d))$ such that the limit of the scenery flow P_x at πx is of the form $A_x P$ for some orthogonal matrix A_x.*

Let us review relevant literature on the scaling properties of measures associated with conformal iterated function systems (IFS). When μ is a Bernoulli measure on $(\Lambda^{\mathbb{N}}, \sigma)$, The projected measure $\pi\mu$ is called a self-conformal measure. If in addition the IFS Φ is self-similar, then $\pi\mu$ is called a self-similar measure.

Feng and Sahlsten [5] showed that homogeneous self-similar measures satisfying the finite type condition are uniform scaling. Subsequently, Pyörälä [20] extended this result to self-similar measures satisfying the weak separation condition. Most recently, Bárány, Käenmäki, Pyörälä and Wu [3] proved that *all* self-similar measures are uniform scaling, without requiring any separation condition. Their work also established the uniform scaling property for all self-conformal measures on $\mathbb{R}$.

The arguments in [3] apply not only to Bernoulli measures, but also to quasi-Bernoulli measures. However, their methods do not extend to more general invariant measures. A key step in their approach involves studying *micromeasures* (or tangents) of self-conformal measures and applying a deep result of Hochman [13], which connects micromeasures with tangent distributions. Crucially, for self-conformal measures, micromeasures are closely related to the original measure—they are approximately given by convolutions of the original measure with low-dimensional measures. This structural relationship plays an essential role in the proof. In contrast, such a property does *not* hold for projections of general invariant measures. To overcome this obstacle in the more general setting, we develop a new strategy. Below, we briefly outline our approach in the case where Φ is a $C^{1+\delta}$ IFS on $\mathbb{R}$ with positive derivatives.

Outline of proof. Let $\Phi = \{f_i\}_{i\in\Lambda}$ be a $C^{1+\delta}$ iterated function system on $\mathbb{R}$ with positive derivatives and μ be an ergodic measure on $\Lambda^{\mathbb{N}}$. We aim to show that the measure $\pi\mu$ is scaling.

The collection of Borel probability measures on a metric space X is denoted by $\mathcal{P}(X)$. For $\nu_1, \nu_2 \in \mathcal{P}(X)$, we denote by $d_{\mathrm{LP}}(\nu_1, \nu_2)$ the Lévy-Prohorov distance between ν_1 and ν_2. We use $\mathcal{D}_n$ to denote the collection of dyadic intervals of length 2^{-n} and $\mathcal{D}_n(x)$ the dyadic interval containing $x \in \mathbb{R}$. For probability measure ν on

$\mathbb{R}$ and $x \in \mathrm{supp}(\nu)$, we denote

$$\nu^{\mathcal{D}_n(x)} = \frac{1}{\mu(\mathcal{D}_n(x))} S_{\mathcal{D}_n(\pi x)}(\mu|_{\mathcal{D}_n(x)}),$$

where $\nu|_{\mathcal{D}_n(x)}$ is the restriction of ν on $\mathcal{D}_n(x)$ and $S_{\mathcal{D}_n(x)}$ is the unique homothety sending $\mathcal{D}_n(x)$ to $[0,1]$. For $\nu \in \mathcal{P}(X)$ and a partition $\mathcal{A}$ of X, $H(\nu, \mathcal{A})$ is the Shannon entropy of ν with respect to $\mathcal{A}$.

In what follows, let us take $\nu = \pi\mu$ and set $\alpha = \dim \nu$.

Step (1). We first show that ν is uniformly dimensional in the sense that for ν-a.e. x, we have

$$\limsup_{\epsilon\to 0} \limsup_{\ell\to\infty} \limsup_{n\to\infty} \frac{1}{n} \left| \left\{ 1 \le k \le n : \left| \frac{H(\nu^{\mathcal{D}_n(x)}, \mathcal{D}_\ell)}{\ell \log 2} - \alpha \right| \ge \epsilon \right\} \right| = 0. \tag{1}$$

To prove the above estimate, the following fact is used: For μ-a.e. x, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \delta_{\mu^{[x_1^n]}} = Q, \tag{2}$$

where Q is some ergodic CP-distribution on $(\Lambda^{\mathbb{N}}, \sigma)$. Moreover, we have

$$\mu = \int \eta dQ(\eta). \tag{3}$$

Let $A \subset \Lambda^{\mathbb{N}}$ be a compact set such that $\mu(A) > 1-\epsilon$. Let A_1 be the ρ-neighbourhood of A. Then, using (2) and (3), we can show that for μ-a.e. x,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mu^{[x_1^n]}(A_1) \ge 1 - \epsilon.$$

In later use of this fact, we shall take $\rho = 2^{-n}$ for large n. Using the last estimate, one is able to prove (1).

Step (2). Use "local entropy average" techniques of Hochman-Shmerkin [15] to show that (combining with Step (1)) for any fixed ℓ (with $\ell \gg 1$, ℓ is very large relative to the contraction ratios of the maps of $\mathcal{F}$) the following holds for μ-a.e. x:

$$\limsup_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \frac{1}{\ell} \int_0^\ell \mathrm{dist}\left((\pi\mu_{[x_1^n]})^{B(\pi x, e^{-t-\lambda_n(x)})}, (\pi\mu)^{B(\pi x, e^{-t-\lambda_n(x)})} \right) dt = o_\ell(1),$$

where $\lambda_n(x)$ is the Lyapunov exponent of $f_{x_0^n}$. Recall that $\mu_{[x_0^n]} = \mu|_{[x_1^n]}/\mu[x_1^n]$, it is a probability measure on the cylinder $[x_0^n]$.

Step (3). We "guess" the limiting fractal distribution. Let $\tilde{\mu}$ be the natural extension of μ. Thus $\tilde{\mu}$ is a shift-invariant ergodic measure on $\Lambda^{\mathbb{Z}}$ with $\mathrm{Proj}_+(\tilde{\mu}) = \mu$, where $\mathrm{Proj}_+ : \Lambda^{\mathbb{Z}} \to \Lambda^{\mathbb{N}}$ is the nature projection $\mathrm{Proj}_+(x_{-\infty}^{\infty}) = x_1^{\infty}$.

For $\tilde{\mu}$-a.e. $x = x_{-\infty}^{\infty}$, the following limit exists

$$\lim_{n\to\infty} \mu^{[x_{-n}^0]}.$$

We denote this limit as $\mu^{[x_{-\infty}^0]}$. Note that $\mu^{[x_{-\infty}^0]}$ is a measure defined on $\Lambda^{\mathbb{N}}$. Let

$$P_\ell = \int_{\Lambda^{\mathbb{Z}}} \left(\frac{1}{\ell} \int_1^\ell \delta_{(\pi\mu^{[x_{-\infty}^0]})^{B(\pi x_1^\infty, e^{-t})}} dt \right) d\tilde{\mu}(x).$$

Claim 1.2. *For large $\ell \gg 1$, the following holds true: For μ-a.e. x,*

$$\limsup_{T\to\infty} \operatorname{dist}\left(P_\ell,\ \frac{1}{T}\int_1^T \delta_{(\pi\mu)^{B(\pi x, e^{-t})}}\,dt\right) = o_\ell(1).$$

From the above Claim, we deduce that

(i) the limit $P := \lim_{\ell\to\infty} P_\ell$ exists;
(ii) for μ-a.e. x, we have

$$\lim_{T\to\infty} \frac{1}{T}\int_1^T \delta_{(\pi\mu)^{B(\pi x, e^{-t})}}\,dt = P.$$

These imply that the measure $\pi\mu$ is scaling.

2. Preliminaries

Suppose (X, d_X) is a metric space. Denote by $\mathcal{B}(X)$ the family of Borel sets, and by $\mathcal{P}(X)$ the space of Borel probability measures. Implicitly, $\mathcal{P}(X)$ is endowed with the weak* topology, which, when X is separable, is compatible with the *Lévy-Prohorov metric*:

$$d_{\mathrm{LP}}(\mu,\nu) := \inf\{\epsilon : \mu(A) \le \nu(A^{(\epsilon)}) \text{ and } \nu(A) \le \mu(A^{(\epsilon)}), A \in \mathcal{B}(X)\},$$

where $A^{(\epsilon)} = \{x \in A : d_X(x, A) < \epsilon\}$ means the set of points in the ϵ-neighborhood of A. As a well-known fact, if the metric space (X, d_X) is complete (respectively, compact), then so is $(\mathcal{P}(X), d_{\mathrm{LP}})$ (respectively, compact).

Let $\mu, \nu \in \mathcal{P}(X)$ and $\mathcal{A} \subset \mathcal{B}(X)$ be a countable partition. The *Shannon entropy* of μ with respect to $\mathcal{A}$ is

$$H(\mu, \mathcal{A}) = -\sum_{A\in\mathcal{A}} \mu(A)\log\mu(A),$$

and the *Kullback–Leibler divergence* between μ and ν with respect to $\mathcal{A}$ is

$$D_{\mathrm{KL}}(\mu||\nu, \mathcal{A}) = \sum_{A\in\mathcal{A}} \mu(A)\log\frac{\mu(A)}{\nu(A)},$$

where the partition $\mathcal{A}$ is omitted if it is clear from the context. We recall in the following a number of entropy-related inequalities.

Lemma 2.1. *Let $\mu, \nu \in \mathcal{P}(X)$, and $\mathcal{A}, \mathcal{A}'$ be a countable partition of X. Then, the following hold.*

- *The function $\mu \mapsto H(\mu, \mathcal{A})$ is concave.*
- *If $N := \sup_{A\in\mathcal{A}} \#\{A' \in \mathcal{A}' : A' \cap A \ne \emptyset\} < \infty$, then*

$$0 \le H(\mu, \mathcal{A}\vee\mathcal{A}') - H(\mu, \mathcal{A}) \le \log N.$$

The inequalities above are fundamental, and proofs can be found, for example, in [14, Section 11]. The following is due to Pinsker, and a proof can be found in [21, Lemma 2.5].

Proposition 2.2 (Pinsker)**.** *Let $\mu, \nu \in \mathcal{P}(X)$, and $\mathcal{A}$ be a countable partition of X into Borel sets. Then,*

$$0 \le \operatorname{dist}((\mu(A))_{A\in\mathcal{A}}, (\nu(A))_{A\in\mathcal{A}}) \le \sqrt{1/2 \cdot D_{\mathrm{KL}}(\mu||\nu, \mathcal{A})} \tag{4}$$

where 'dist' is the total variation distance *between probability vectors:*

$$\mathrm{dist}((p_I)_{i\in I}, (q_I)_{i\in I}) := \sup_{I'\subseteq I} \sum_{i\in I'} p_i - q_i.$$

Proposition 2.2 alludes to an interplay between entropy, total variation distance, and Lévy-Prohorov distance, as can be illustrated by considering $X = \mathbb{R}^d$ as follows: Letting

$$\mathcal{D}_n := \left\{ \prod_{j=1}^{d} [i_j \cdot 2^{-\lfloor n \rfloor}, (i_j+1)\cdot 2^{-\lfloor n \rfloor}) : i_j \in \mathbb{Z}, \forall j = 1, 2, \cdots, d \right\}$$

and $\mathcal{D}_n(x)$ be the unique cube in $\mathcal{D}_n$ containing $x \in \mathbb{R}^d$, inequality (4) yields

$$\begin{aligned} d_{\mathrm{LP}}(\mu,\nu) &\le \max\{\mathrm{dist}(\mu(D))_{D\in\mathcal{D}_n}, (\mu(D))_{D\in\mathcal{D}_n}), \sqrt{d}\cdot 2^{-n}\} \\ &\le \max\{\sqrt{2^{-1}\cdot D_{\mathrm{KL}}(\mu\|\nu,\mathcal{D}_n)}, \sqrt{d}\cdot 2^{-n}\}. \end{aligned} \tag{5}$$

In this context, we define for every $\mu \in \mathcal{P}(\mathbb{R}^d)$ and every $D_n \in \mathcal{D}_n$ the *dyadic magnification* of μ at D_n by

$$\mu^{D_n} := \frac{(f_{D_n})_*(\mu|_{D_n})}{\mu(D_n)},$$

where f_{D_n} is the unique orientation-preserving homothety mapping D_n to $[0,1)^d$.

The symbols 'o' and 'O' are the usual little-o and big-O notation, with the dependent variables displayed in subscript, e.g., $o_{a,b}(1)$ denotes the error estimates depending on variables a and b.

2.1. **Shift spaces.** Let Λ be an alphabet. A *one-sided full-shift* is a topological dynamical system $(\Lambda^{\mathbb{N}}, \sigma)$ equipped with a metric $d_{\Lambda^{\mathbb{N}}}(x,y) = e^{-\min\{i: x_i \neq y_i\}}$ and a continuous transformation $\sigma : \Lambda^{\mathbb{N}} \to \Lambda^{\mathbb{N}}$ defined by $(\sigma x)_i = x_{i+1}$ for all $i \in \mathbb{N}$. A *word* w of length n is an element of Λ^n, and its associated *n-cylinder set* is defined by

$$[w] := \{x \in \Lambda^{\mathbb{N}} : x_1^n = w_1^n\},$$

where

$$y_i^j = (y_i y_{i+1} \cdots y_j) \quad \text{for } i \le j, y \in \Lambda^{\mathbb{N}} \cup (\cup_{n\ge j}\Lambda^n).$$

Write $\mathcal{C}_n = \{[w] : w \in \Lambda^n\}$ for the collection of n-cylinder sets, and $\mathcal{C}_{m:n} = \sigma^{-m}\mathcal{C}_{n-m+1}$. The collection of all σ-invariant (respectively, ergodic) Borel probability measures on $(\Lambda^{\mathbb{N}}, \sigma)$ is denoted by $\mathcal{P}_\sigma(\Lambda^{\mathbb{N}})$ (respectively, $\mathcal{P}_{erg}(\Lambda^{\mathbb{N}})$). For the purpose of the measure-theoretic study, such a system is further equipped with the Borel Σ-algebra and a σ-invariant Borel probability measures μ^+ to become a measure-preserving system $(\Lambda^{\mathbb{N}}, \sigma, \mu^+)$. In what follows, discussions are at times carried out with the aid of the natural extension of the $(\Lambda^{\mathbb{N}}, \sigma)$ in the form of the *two-sided full-shift* $(\Lambda^{\mathbb{Z}}, \sigma)$, where the transformation σ is extended to the space $\Lambda^{\mathbb{Z}}$ by setting $(\sigma x)_i = x_{i+1}$ for $i \in \mathbb{Z}$. Such an extension is inherited by its measure-theoretic counterpart, yielding an extended measure-preserving system and, in particular, a (uniquely) extended σ-invariant measure $\mu \in \mathcal{P}_\sigma(\Lambda^{\mathbb{Z}})$ from $\mu^+ \in \mathcal{P}_\sigma(\Lambda^{\mathbb{N}})$. As a readily checked fact, $\mu \in \mathcal{P}_{erg}(\Lambda^{\mathbb{Z}})$ if and only if $\mu^+ \in \mathcal{P}_{erg}(\Lambda^{\mathbb{N}})$.

Facts regarding the entropy of symbolic spaces are collected as follows. For a countable partition $\mathcal{A}$ of the one-sided subshift $\Lambda^{\mathbb{N}}$, the *measure-theoretic entropy* with respect to a finite partition $\mathcal{A}$ is

$$h_\mu(\mathcal{A}) := \lim_{n\to\infty} \frac{1}{n} H\left(\mu, \vee_{i=0}^{n-1}\sigma^{-i}\mathcal{A}\right),$$

where $\vee_{i=0}^{n-1}\sigma^{-i}\mathcal{A} := \{\cap_{i=0}^{n-1}\sigma^{-i}A_i : A_i \in \mathcal{A}\}$, and the *Kolmogorov-Sinai entropy* is defined to be the supremum of it over all partitions

$$h_\mu := \sup_{\mathcal{A}} h_\mu(\mathcal{A}),$$

which, in the context of symbolic spaces, is known to be attained by the partition $\mathcal{C}_1$ formed by 1-cylinder sets, or more explicitly, $h_\mu = \lim_{n\to\infty} \frac{1}{n} H(\mu, \mathcal{C}_n)$.

2.2. **Iterated function systems.** Let X be a complete metric space. An *iterated function system* (IFS) is a collection of contractions $\Phi = \{f_i : X \to X : i \in \Lambda\}$. Associated with the IFS is a coding map $\pi : \Lambda^{\mathbb{N}} \to X$ (or $\pi : \Lambda^{\mathbb{Z}} \to X$ analogously) defined by

$$\pi x = \lim_{n\to\infty} f_{x_1^k}(*) \quad \text{with} \quad f_{x_1^k} := f_{x_1} \circ f_{x_2} \circ \cdots \circ f_{x_k},$$

with $*$ an arbitrary but given point in X. Being continuous, the map π projects $\Lambda^{\mathbb{N}}$ onto a compact set $K = \pi(\Lambda^{\mathbb{N}})$ and hence induces a pushforward map $\pi : \mathcal{P}(\Lambda^{\mathbb{N}}) \mapsto \mathcal{P}(K)$ defined by $\pi\mu^+ := \mu^+ \circ \pi^{-1}$. We denote by $B^\pi(x, r) := \pi^{-1}B(\pi x, r)$.

For the scope of this paper, we consider the iterated function systems on $\mathbb{R}^d$ which are *conformal* in the following sense. Given any open domain $U \subseteq \mathbb{R}^d$, a C^1-map $f : U \to \mathbb{R}^d$ is said to be *conformal* if $x \mapsto D_x f$ is Hölder continuous and $\|D_x f\|^{-1} \cdot D_x f$ is an orthogonal matrix, where $D_x f$ is the derivative of f at x and $\|\cdot\|$ is the operator norm. A collection of maps $\Phi = (f_i)_{i\in\Lambda}$ is called a *conformal iterated function system* if f_i's are injective conformal maps defined on a bounded open convex domain $U \subset \mathbb{R}^d$ such that $\overline{f_i(U)} \subset U$ for all $i \in \Lambda$ and that $\sup_{i\in\Lambda, x\in\mathbb{R}} \|D_x f_i\| < 1$. Notably, [1, Lemma 6.1] asserts that a conformal IFS is indeed an IFS in the sense that it admits a bounded open convex set $V \subset U$ such that

$$f_i(\overline{V}) \subset V \subset \overline{V} \subset U \quad \text{for all } i \in \Lambda, \tag{6}$$

Upon a conformal IFS and a measure $\mu^+ \in \mathcal{P}(\Lambda^{\mathbb{N}})$, we define the *upper* and *lower Lyapunov exponent* at $x \in \Lambda^{\mathbb{N}}$ to be

$$\overline{\lambda}(x) = \limsup_{n\to\infty} -\frac{\log \|D_{\pi\sigma^k x} f_{x_1^k}\|}{n} \quad \text{and} \quad \underline{\lambda}(x) = \liminf_{n\to\infty} \frac{\log \|(D_{\pi\sigma^k x} f_{x_1^k})^{-1}\|}{n},$$

In general, $\overline{\lambda}(x) \geq \underline{\lambda}(x)$; nevertheless, the limits concinde at μ^+-a.e. x if $\mu^+ \in \mathcal{P}_\sigma(\Lambda^{\mathbb{N}})$, in which case we write $\lambda(x) = \overline{\lambda}(x) = \underline{\lambda}(x)$. In addition, if we further require $\mu \in \mathcal{P}_{erg}(\Lambda^{\mathbb{N}})$, the Lyapunov exponenet is constant μ-almost everywhere, and we denote $\lambda = \lambda(x)$.

Here we include a folklore lemma which later comes in handy.

Lemma 2.3 (bounded distortion)**.** *Let $\Phi = (f_i)_{i\in\Lambda}$ be a conformal IFS with θ-Hölder continuous derivatives, and let V be as in* (6)*. Then, there exist constants $C > 0$ such that*

$$\|(D_x f_w)^{-1} D_y f_w - \mathrm{Id}\| \leq C\|x - y\|^\theta \quad \textit{for all } x, y \in V, w \in \cup_{i=1}^\infty \Lambda^i,$$

In particular, putting $C' = (1 + C \cdot \mathrm{diam}(V)^\theta)^{-1}$,

$$\frac{\|f_w(x) - f_w(y)\|}{\sup_{z\in V} \|D_z f_w\| \cdot \|x - y\|} \in [C', 1] \quad \textit{for all } x \neq y \in V, w \in \cup_{i=1}^\infty \Lambda^i.$$

Proof. Let $C', \theta > 0$ be the constants such that

$$\|D_x f_i - D_y f_i\| \le C'\|x-y\|^\theta \quad \text{for all } x, y \in V, i \in \Lambda.$$

Further, take $r_- = \inf_{i\in\Lambda, x\in V} \|D_x f_i\| < 1$ and $r_+ = \sup_{i\in\Lambda, x\in V} \|D_x f_i\| < 1$, so that

$$\|f_w(x) - f_w(y)\| \le r_+^k \|x-y\| \quad \text{for all } x, y \in V, w \in \Lambda^k, k \in \mathbb{N}.$$

Then, for any word w of length n and any $x, y \in V$, we write

$$x_k = \begin{cases} x & \text{if } k = n, \\ f_{w_{k+1}^n}(x) & \text{if } 0 \le k < n, \end{cases} \quad \text{and } y_k = \begin{cases} y & \text{if } k = n, \\ f_{w_{k+1}^n}(y) & \text{if } 0 \le k < n, \end{cases},$$

so that

$$\begin{aligned}
&\|(D_x f_w)^{-1} D_y f_w - \mathrm{Id}\| = \|D_x f_w\|^{-1} \|D_y f_w - D_x f_w\| \\
\le& \frac{\sum_{k=1}^n \prod_{i=1}^{k-1} \|D_{y_i} f_{w_i}\| \cdot \|D_{x_k} f_{w_k} - D_{y_k} f_{w_k}\| \cdot \prod_{j=k+1}^n \|D_{x_j} f_{w_j}\|}{\prod_{k=1}^n \|D_{x_k} f_{w_k}\|} \\
=& \sum_{k=1}^n \prod_{i=1}^{k-1} \frac{\|D_{y_i} f_{w_i}\|}{\|D_{x_i} f_{w_i}\|} \cdot \|D_{x_k} f_{w_k} - D_{y_k} f_{w_k}\|
\end{aligned}$$

Applying the mean value theorem to the function $x \mapsto \log x$, the above is bounded from above by

$$\begin{aligned}
&\|(D_x f_w)^{-1} D_y f_w - \mathrm{Id}\| \\
\le& \sum_{k=1}^n \exp\left(r_-^{-1} \sum_{i=1}^{k-1} \|D_{y_i} f_{w_i} - D_{x_i} f_{w_i}\| \right) \cdot \|D_{x_k} f_{w_k} - D_{y_k} f_{w_k}\| \\
\le& \sum_{k=1}^n \exp\left(\frac{C'}{r_-(1-r_+^\theta)} \|x-y\|^\theta \right) \cdot C' r_+^{\theta(n-k)} \|x-y\|^\theta \\
\le& \frac{C'\|x-y\|^\theta}{1-r_+^\theta} \exp\left(\frac{C'\|x-y\|^\theta}{r_-(1-r_+^\theta)} \right) \le \frac{C'\|x-y\|^\theta}{1-r_+^\theta} \exp\left(\frac{C' \cdot \mathrm{diam}(V)^\theta}{r_-(1-r_+^\theta)} \right),
\end{aligned}$$

This wraps up the proof. □

2.3. **Dimensions.** Let $\mu \in \mathcal{P}(X)$. The *upper* and *lower local dimensions* of μ at a point $x \in X$ are respectively defined by

$$\overline{\dim}_{\mathrm{loc}}(\mu, x) = \limsup_{r\to 0^+} \frac{\log \mu(B(x,r))}{\log r} \quad \text{and} \quad \underline{\dim}_{\mathrm{loc}}(\mu, x) = \liminf_{r\to 0^+} \frac{\log \mu(B(x,r))}{\log r}.$$

If $\overline{\dim}_{\mathrm{loc}}(\mu, x) = \underline{\dim}_{\mathrm{loc}}(\mu, x)$ coincide, their common value is denoted as $\dim_{\mathrm{loc}}(\mu, x)$; if further $\dim_{\mathrm{loc}}(\mu, x)$ is constant μ-amost everywhere (μ-a.e.), the measure μ is said to be *exactly dimensional* and we set $\dim \mu$ as the dimension of the measure. When $X = \mathbb{R}^d$, the local dimensions, as functions that agree almost everywhere, can alternatively be determined via dyadic partitions of $\mathbb{R}^d$: For μ-a.e. x,

$$\overline{\dim}_{\mathrm{loc}}(\mu, x) = \limsup_{n\to\infty} \frac{\log \mu(\mathcal{D}_n(x))}{n \log 2} \quad \text{and} \quad \underline{\dim}_{\mathrm{loc}}(\mu, x) = \liminf_{n\to\infty} \frac{\log \mu(\mathcal{D}_n(x))}{n \log 2}.$$

In particular, Shannon-McMillan-Breiman theorem and [6, Theorem 2.8], respectively, assert that upper and lower local dimensions coincide almost everywhere for μ^+ and $\pi\mu^+$ when π is associated with a conformal IFS and $\mu^+ \in \mathcal{P}_\sigma(\Lambda^{\mathbb{N}})$, and that both measures are exact dimensional if $\mu^+ \in \mathcal{P}_{erg}(\Lambda^{\mathbb{N}})$.

In this article, we also consider the *upper* and *lower Hausdorff dimension* of a measure μ, which are defined respectively as

$$\underline{\dim}_{\mathrm{H}}\mu = \inf\{\dim_{\mathrm{H}} E : E \in \mathcal{B}(X), \mu(E) = 1\}$$

and

$$\underline{\dim}_{\mathrm{H}}\mu = \inf\{\dim_{\mathrm{H}} E : E \in \mathcal{B}(X), \mu(E) > 0\},$$

and we can safely write $\dim_{\mathrm{H}} \mu$ whenever the dimensions above coincide. Equivalently, the dimensions are sometimes interpreted in terms of local dimensions (see for example [4, Proposition 10.3]) as

$$\underline{\dim}_{\mathrm{H}}\mu = \operatorname*{ess\,inf}_{x\sim\mu} \dim_{\mathrm{loc}} \dim_{\mathrm{loc}}(\mu, x) \text{ and } \overline{\dim}_{\mathrm{H}}\mu = \operatorname*{ess\,sup}_{x\sim\mu} \dim_{\mathrm{loc}} \dim_{\mathrm{loc}}(\mu, x).$$

2.4. **Fractal distributions.** Let $X = \{\mu \in \mathcal{P}(B_1) : 0 \in \mathrm{supp}\mu\}$, which is a Borel-measurable subset in the space $\mathcal{P}(B_1)$, and $S : \mathbb{R} \times X \to X$ be the scaling flow, a measurable flow defined by $S_t\mu = \mu_{0,t}$. We recall some definitions in the following.

Definition 2.4 (fractal distribution)**.** *Let (X, S) be as defined.*

- *$P \in \mathcal{P}(X)$ is said to be quasi-Palm the following is satisfied: If $\mathcal{A}$ is a P-full Borel set, then P-a.e $\nu \in \mathcal{A}$ and ν-a.e. x satisfying $B(x, e^{-t}) \subset B_1$, we have $\nu_{x,t} \in \mathcal{A}$.*
- *A distribution $P \in \mathcal{P}(X)$ is called a (restricted)* fractal distribution *if P is S-invariant and quasi-Palm.*
- *An* ergodic fractal distribution *P is a fractal distribution (c.f. [13, Theorem 1.3]) that renders (X, S, P) ergodic.*

It is noteworthy that although the results in this paper are stated only for restricted version of fractal distributions, cited theorems are sometimes stated in the form *extended fractal distributions* ([13, Definition 1.2]). Such transplants are, throughout the paper, consistently justified by the one-to-one correspondence between the two classes of distributions [13, Lemma 3.1]. In particular, the characterization of ergodic decomposition for extended fractal distributions [13, Theorem 1.3] passes naturally to its restricted counterparts due to the aforementioned correspondence.

2.5. **CP-Processes.** The idea of CP-processes was first proposed by Furstenberg in [9]. Herein, we focus on CP-processes under a symbolic setting.

Definition 2.5 (CP-distribution)**.** *Consider a topological dynamical system (X, M) defined by*

$$X = \{(\mu, x) \in \mathcal{P}(\Lambda^{\mathbb{N}}) \times \Lambda^{\mathbb{N}} : x \in \mathrm{supp}(\mu)\} \quad \textit{and} \quad M(x, \mu) = (\sigma x, \mu^{[x_1]}),$$

where X is equipped with the product topology and $\mu^{[x_1]} := \sigma(\mu|_{[x_1]})/\mu[x_1]$. A CP-distribution *is a measure $P \in \mathcal{P}(X)$ which is M-invariant and adapted, i.e., for every $f \in C(\mathcal{P}(\Lambda^{\mathbb{N}}), \Lambda^{\mathbb{N}})$,*

$$\int f(\mu, x) dP(\mu, x) = \int f(\mu, x) d\mu(x) dP^1(\mu), \tag{7}$$

where P^1 is the marginal of P on the first coordinate.

Remark 2.6. *It is readily checked from the definition above that $M^n(x, \mu) = (\sigma^n x, \mu^{[x_1^n]})$, where $\mu^{[x_1^n]} := \sigma^n(\mu|_{[x_1^n]})/\mu([x_1^n])$.*

Some properties of CP-distributions are recalled in the following proposition.

Proposition 2.7. *Suppose $\mu^+ \in \mathcal{P}_{erg}(\Lambda^{\mathbb{N}})$ with a natural extension $\mu \in \mathcal{P}_{erg}(\Lambda^{\mathbb{Z}})$. Then,*

(1) For μ-a.e. x, the prediction measure

$$\mu^{[x^0_{-\infty}]} := \lim_{n\to\infty} \mu^{[x^0_{-n}]}$$

is well-defined and is measurable as a function of $x \in \Lambda^{\mathbb{Z}}$.

(2) There exists an ergodic CP-distribution $P \in \mathcal{P}(\mathcal{P}(\Lambda^{\mathbb{N}}) \times \Lambda^{\mathbb{N}})$ such that

$$P_{x,n} := \frac{1}{n}\sum_{k=1}^{n} \delta_{(\mu^{[x^k_{-\infty}]}, x^\infty_{k+1})} \xrightarrow{weak^*} P \quad \text{for } \mu\text{-a.e. } x.$$

(3) For for every $\epsilon > 0$, there exists $\ell = \ell(\epsilon)$ such that μ-a.e. x,

$$\limsup_{n\to\infty} \frac{1}{n}\#\left\{k \in [1,n] : \left|\frac{1}{\ell}H(\mu^{[x^k_{-\infty}]}, \mathcal{C}_\ell) - h\right| > \epsilon\right\} < \epsilon,$$

where $h = \dim \mu$.

Proof. (1) This follows from the Martingale convergence theorem.

(2) Let $(\Lambda^{\mathbb{Z}}, \sigma, \mu)$ be the natural extension of $(\Lambda^{\mathbb{N}}, \sigma, \mu^+)$. Then, inasmuch as μ is ergodic, Birkhoff ergodic theorem gurantees μ-almost sure convergence of $P_{x,n} \to P$ to a constant distribution $P \in \mathcal{P}(\mathcal{P}(\Lambda^{\mathbb{N}}) \times \Lambda^{\mathbb{N}})$. The adaptedness and the ergodicity of P follow from a routine check and are thus omitted here.

(4) The exact dimensionality of μ follows from the Shannon-McMillan-Breiman theorem, whereas the stated estimate follows from an identical argument as in [22, Proposition 3.7]. □

Lemma 2.8. *Let $\mu \in \mathcal{P}_{erg}(\Lambda^{\mathbb{Z}})$. Then, for every $\epsilon > 0$, there exists $\ell(\epsilon) \in \mathbb{N}$ such that for μ-a.e. x and all $\ell \geq \ell(\epsilon)$,*

$$\limsup_{n\to\infty} \frac{1}{n}\#\left\{k \in [1,n] : \frac{H(\pi\mu^{[x^k_{-\infty}]}, \mathcal{D}_\ell)}{\ell \log 2} \leq \dim \pi\mu - \epsilon\right\} < \epsilon.$$

Proof. As the preparation step, recall the exact dimensionality of $\pi\mu^{[x^0_{-\infty}]}$ for μ-a.e. x given that $\mu \in \mathcal{P}_{erg}(\Lambda^{\mathbb{Z}})$ (which follows essentially from [6, Theorem 2.11]). Under the circumstances, we can safely write $\dim \pi\mu = \alpha$. Fix $0 < \epsilon < \min\{\lambda/2, \alpha\}$, where $\lambda > 0$ denotes the Lyapunov exponent. According to Proposition 2.7, one is entitled to choose $\ell(\epsilon)$ such that

$$\limsup_{n\to\infty} \frac{1}{n}\#\left\{k \in [1,n] : \left|\frac{H(\mu^{[x^k_{-\infty}]}, \mathcal{C}_\ell)}{\ell \log 2} - h\right| > \epsilon\right\} < \epsilon \text{ for all } \ell \geq \ell(\epsilon), \tag{8}$$

Exploiting the exact dimensionality of the measures $\mu^{[x^0_{-\infty}]}$ and $\pi\mu^{[x^0_{-\infty}]}$ and applying the Egorov's theorem, there exist a compact set $A \subset \text{supp}(\mu^{[x^0_{-\infty}]})$ with $\mu^{[x^0_{-\infty}]}(A) > 1 - \epsilon^2$ and an integer $\ell'(\epsilon) \geq \max\{\epsilon^{-1}, \ell(\epsilon), \sqrt{d} + 1\}$ such that for all

$x \in A$ and all $\ell \geq \ell'(\epsilon)$

$$\text{(A)} \qquad \left| \frac{\log \left\| D_{\pi\sigma^\ell x} f_{x_1^\ell} \right\|}{\ell \log 2} - \lambda \right| < \epsilon,$$

$$\text{(B)} \qquad \left| \frac{\log \mu^{[x_{-\infty}^0]}[x_1^\ell]}{\ell \log 2} - h \right| < \epsilon,$$

$$\text{(C)} \qquad \left| \frac{\log \mu^{[x_{-\infty}^0]}(B^\pi(x, 2^{-\ell(\lambda-2\epsilon)}))}{\ell(\lambda - 2\epsilon)\log 2} - \alpha \right| < \epsilon,$$

where (B) follows from the ergodicity as in Proposition 2.7, and (C) follows from an adaptation of [6, Theorem 2.2]. Moreover, combining Lemma 2.3 and (A), we assume without loss of generality that the following also holds for all $x \in A$ and $\ell \geq \ell'(\epsilon)$:

$$\text{(A')} \qquad [x_1^\ell] \subset B^\pi(x, 2^{-\ell(\lambda-\epsilon)}).$$

For each $\ell \in \mathbb{N}$, denote by $N_\ell(S)$ the number of ℓ-cylinder sets in $\mathcal{C}_\ell$ intersecting $S \subset \Lambda^{\mathbb{Z}}$ and enlarge A to its open neighborhood $A^{(\ell)} := \cup_{w\in\Lambda^\ell:[w]\cap A\neq\emptyset}[w]$. We first claim the following:

$$(9) \qquad \sup_{D\in\mathcal{D}_{\ell\lambda}} \frac{\log N_\ell(A^{(\ell)} \cap \pi^{-1}D_{\ell\lambda})}{\ell \log 2} < h - \alpha\lambda + O_d(\epsilon).$$

Indeed, for each $\ell \geq \ell'(\epsilon)$ and each $A^{(\ell)} \cap \pi^{-1}D_{\ell\lambda}$, there exists $x \in A$ such that

$$A^{(\ell)} \cap \pi^{-1}D_{\ell\lambda} \subset B^\pi(x, (\sqrt{d}+1)\cdot 2^{-\ell(\lambda-\epsilon)}) \subset B^\pi(x, 2^{-\ell(\lambda-2\epsilon)}).$$

The assumptions (A')(B)(C) altogether then imply

$$N_\ell(A^{(\ell)} \cap \pi^{-1}D_{\ell\lambda})\cdot 2^{-\ell(h+\epsilon)} \leq \sum_{\substack{w\in\Lambda^\ell: \\ [w]\cap A\cap B^\pi(x,2^{-\ell(\lambda-2\epsilon)})\neq\emptyset}} \mu^{[x_{-\infty}^0]}[w]$$
$$\leq \mu^{[x_{-\infty}^0]}(B^\pi(x, 2^{-\ell(\lambda-2\epsilon)})) \leq 2^{-\ell(\lambda-2\epsilon)(\alpha-\epsilon)},$$

from which the claim follows immediately. To prove the lemma, utilize Proposition 2.7 (3) to deduce that

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \mu^{[x_{-\infty}^k]}(A^{(\ell)}) > 1 - \epsilon^2.$$

Combining this identity with (8), one obtains a set $\mathcal{N}_\ell \subset \mathbb{N}$ with $\underline{\mathtt{d}}(\mathcal{N}_\ell) \geq 1 - 2\epsilon$ that satisfies the following property:

$$k \in \mathcal{N}_\ell \Rightarrow \frac{H(\mu^{[x_{-\infty}^k]}, \mathcal{C}_\ell)}{\ell \log 2} > h - \epsilon \text{ and } \mu^{[x_{-\infty}^k]}(A^{(\ell)}) > 1 - \epsilon.$$

Consequently, by expressing $\mu^{[x_{-\infty}^k]} = (1-\delta)\nu_1 + \delta\nu_2$ with $\delta = \mu^{[x_{-\infty}^k]}(\Lambda^{\mathbb{N}} \setminus A^{(\ell)})$ and

$$\nu_1 = (1-\delta)^{-1}\cdot \mu^{[x_{-\infty}^k]}|_{A^{(\ell)}} \quad \text{and} \quad \nu_2 = \delta^{-1}\cdot \mu^{[x_{-\infty}^k]}|_{\Lambda^{\mathbb{N}}\setminus A^{(\ell)}},$$

we deduce from (9) and from Lemma 2.1 that for $k \in \mathcal{N}_\ell$

$$\frac{H(\mu^{[x^k_{-\infty}]}, \pi^{-1}\mathcal{D}_{\ell\lambda})}{\ell\lambda \log 2} \geq (1-\delta) \cdot \frac{H(\nu_1, \pi^{-1}\mathcal{D}_{\ell\lambda})}{\ell\lambda \log 2} + \delta \cdot \frac{H(\nu_2, \pi^{-1}\mathcal{D}_{\ell\lambda})}{\ell\lambda \log 2}$$
$$\geq (1-\delta)\left[\frac{H(\nu_1, \pi^{-1}\mathcal{D}_{\ell\lambda} \vee \mathcal{C}_\ell)}{\ell\lambda \log 2} - (h - \alpha\lambda + O_d(\epsilon))\right]$$
$$\geq (1-\delta)\left[\frac{H(\nu_1, \mathcal{C}_\ell)}{\ell\lambda \log 2} - (h - \alpha\lambda + O_d(\epsilon))\right] = \alpha + O_d(\epsilon).$$

This concludes the proof. □

3. Technical estimates

In this section, we establish necessary lemmas in preparation for the proof of Theorem 1.1. Throughout this section, μ^+ is reserved for a measure in $\mathcal{P}_{erg}(\Lambda^{\mathbb{N}})$ which is associated with its natural extension $\mu \in \mathcal{P}_{erg}(\Lambda^{\mathbb{N}})$, and π is always the projection map associated with a conformal IFS $(f_i)_{i\in\Lambda}$. In addition, we consistently denote by $\overline{\mathsf{d}}(\mathcal{N})$ and $\underline{\mathsf{d}}(\mathcal{N})$ the upper and lower asymptotic density of a set of indices $\mathcal{N} \subseteq \mathbb{N}$, respectively.

Lemma 3.1. *Suppose $f_j : \mathbb{N} \to \mathbb{N}$, $j = 1, \cdots, J$, are at most n-to-one increasing sequence with bounded gap $\max_{k\in\mathbb{N}} f_j(k+1) - f_j(k) \leq m$. Then, the following hold.*

- *If $\underline{\mathsf{d}}(\mathcal{N}) \geq 1-\delta$, then $\underline{\mathsf{d}}(f_j^{-1}(\mathcal{N})) \geq 1 - nm\delta$.*
- *If $\#\mathbb{N} \setminus \bigcup_{j=1}^J f_j(\mathbb{N}) < \infty$ and $\underline{\mathsf{d}}(f_j^{-1}(\mathcal{N})) \geq 1-\delta$, then $\underline{\mathsf{d}}(\mathcal{N}) \geq 1 - nJ\delta$.*

Proof. The first estimate follows from the following observation:

$$f_j([1,N]) \cap \mathcal{N}^c \subset \mathcal{N}^c \cap [1, mN].$$

For the second, assume $\mathbb{N} \setminus \bigcup_{j=1}^J f_j(\mathbb{N}) \subset [1, N_0]$ and observe

$$(N_0, N] \cap \mathcal{N}^c \subset \cup_{j=1}^J f_j([1, nmN]) \cap \mathcal{N}^c$$

as desired. □

Lemma 3.2. *Suppose $\dim \pi\mu > 0$. Then, for μ-a.e. x,*

$$\limsup_{\ell\to\infty} \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^n \frac{1}{\ell}\int_0^\ell d_{\mathrm{LP}}\left((\pi\mu^{[x^-]}|_{[x_1^k]})_{\pi x, t+\lambda_k(x)}, (\pi\mu)_{\pi x, t+\lambda_k(x)}\right) dt = 0,$$

where $x^- = x^0_{-\infty}$ and $\lambda_k(x) = -\log \|D_{\pi x^\infty_{k+1}} f_{x_1^k}\|$.

Proof. To prove the lemma, it suffices to show that for μ-a.e. x, every $\epsilon > 0$, and all sufficiently large $\ell \in \mathbb{N}$, there exists a set of indices $\mathcal{N} = \mathcal{N}(x, \epsilon, \ell) \subseteq \mathbb{N}$ with $\overline{\mathsf{d}}(\mathcal{N}^c) \leq 1 - \epsilon$ such that

$$k \in \mathcal{N} \Rightarrow \frac{1}{\ell}\int_0^\ell d_{\mathrm{LP}}\left((\pi\mu^{[x^-]}|_{[x_1^k]})_{\pi x, t+\lambda_k(x)}, (\pi\mu)_{\pi x, t+\lambda_k(x)}\right) dt < \epsilon,$$

whence the lemma results from the Borel-Catelli lemma.

The general assumptions and conventions are as follows. Firstly, $\pi\mu$ is assumed to have its support $\mathrm{supp}\pi\mu$ contained in $\subset V \subset \overline{V} \subset (0,1)^d$, where V is the bounded open convex set given in (6). Secondly, for μ-a.e. x, $(2^n\pi(x) \pmod 2)$ is assumed to be equidistributed, meaning that $n^{-1} \cdot \sum_{k=1}^n \delta_{2^k x \pmod 2}$ converges weak* to the

Lebesgue measure $\mathcal{L}^d$ on $\mathbb{R}^d$. This could always be achieved by a proper translation of the measure, as $\mathcal{L}^d$-a.e. $(2^n y \pmod 2)$ is equidistributed and thus

$$\pi\mu \times \mathcal{L}^d(\{(y,t) \in \mathbb{R}^d \times \mathbb{R}^d : (2^n(y+t) \pmod 2)_{n\in\mathbb{N}} \text{ is not equidistributed}\}) = 0.$$

Both assumptions immediately pass on to $\pi\mu^{[x^-]}$ as it is absolutely continuous with respect to $\pi\mu$ by definition. Once the claim is proved with these assumptions, the remaining case then follows naturally. We write $\alpha = \dim \pi\mu$ for short throughout.

Let $\hat{\ell} = \lfloor \ell^{1/2} \rfloor$, and $L_k(x) = \min\{n : \sup_{y\in V} \|D_y f_{x_1^n}\| < 2^{-k}\}$. We first prove that for μ-a.e. x, every $\epsilon > 0$ and all sufficiently large $\ell \in \mathbb{N}$, there exists $\mathcal{N}_\ell \subseteq \mathbb{N}$ such that $\overline{\mathsf{d}}((\mathcal{N}_\ell)^c) < \epsilon$ and that

$$k\ell + i \in \mathcal{N}_\ell \Rightarrow d_{\mathrm{LP}}\left(\left(\pi\mu^{[x^-]}|_{[x_1^{L_{k+\hat{\ell}}(x)}]}\right)^{\mathcal{D}_{k+i}(\pi x)}, (\pi\mu)^{\mathcal{D}_{k+i}(\pi x)}\right) < \epsilon, \tag{10}$$

where the criterion should be interpreted with $k \in \mathbb{N}$ and $0 \le i < \ell$. To this end, fix ℓ_0 to be the smallest integer such that $d^{1/2} \cdot 2^{-\ell_0} < \epsilon$ and

$$\mathcal{E}_k = \left\{[w] : \sup_{y\in V} \|D_y f_{w'}\| \ge 2^{-k} > \sup_{y\in V} \|D_y f_w\|\right\},$$

where w' is the subword of w with the last letter removed. Then, define the information function by

$$I^\mu_{k,\ell}(x) = -\log \frac{\pi\mu(\mathcal{D}_{k+\ell}(\pi x))}{\pi\mu(\mathcal{D}_k(\pi x))}.$$

In terms of the information function, the local dimension of $\pi\mu$ is almost surely expressed as

$$\alpha \equiv \dim_{\mathrm{loc}}(\pi\mu, \pi x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\ell} \sum_{i=1}^{\ell} \frac{I^\mu_{k+i,\ell_0}(x)}{\ell_0 \log 2},$$

Furthermore, we can utilize the law of large numbers for Martingale difference sequences to deduce its almost sure coincidence with

$$\begin{aligned} \alpha &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\ell} \sum_{i=1}^{\ell} \mathbb{E}\left[\frac{I^\mu_{k+i,\ell_0}}{\ell_0 \log 2} \middle| \Sigma(\mathcal{C}_{-\infty:0}, \mathcal{E}_{k+\hat{\ell}}, \pi^{-1}\mathcal{D}_{k+i})\right](x) \\ &= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\ell} \sum_{i=1}^{\ell} \frac{1}{\ell_0 \log 2} \cdot \frac{\nu_{x,k,\ell}((\pi^{-1}\mathcal{D}_{k+\ell_0+i})(x))}{\nu_{x,k,\ell}((\pi^{-1}\mathcal{D}_{k+i})(x))} \cdot I^\mu_{k+i,\ell_0}(x). \end{aligned} \tag{11}$$

where $\Sigma(\cdot)$ stands for the Σ-algebra generated by the specified collections of sets and

$$\nu_{x,k,\ell} := \mu^{[x^-]}|_{[x_1^{L_{k+\hat{\ell}}(x)}]}.$$

By Gibbs's inequality (i.e., the first inequality in (4)), (11) serves as a term-wise upper bound of

$$\begin{aligned}
&\liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}\frac{1}{\ell}\sum_{i=1}^{\ell}\frac{1}{\ell_0\log 2}\cdot\frac{\nu_{x,k,\ell}((\pi^{-1}\mathcal{D}_{k+\ell_0+i})(x))}{\nu_{x,k,\ell}((\pi^{-1}\mathcal{D}_{k+i})(x))}\cdot I_{k+i,\ell_0}^{\nu_{x,k,\ell}}(x)\\
(12)\qquad &=\liminf_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n}\frac{1}{\ell}\sum_{i=1}^{\ell}\frac{1}{\ell_0\log 2}\cdot I_{k+i,\ell_0}^{\nu_{x,k,\ell}}(x)\\
&=\liminf_{n\to\infty} \frac{1}{n}\sum_{j=1}^{\ell_0}\frac{N_j\ell_0}{\ell}\sum_{k=1}^{n}\frac{H((\pi\nu_{x,k,\ell})^{\mathcal{D}_{k+j}(\pi x)},\mathcal{D}_{k+j+N_j\ell_0})}{N_j\ell_0\log 2},
\end{aligned}$$

where $N_j := \lceil\frac{\ell+1-j}{\ell_0}\rceil$ and the equalities are once again derived using Martingale difference sequences. Given this last expression, a lower bound of (12) could be obtained by applying Lemma 2.8: Due to the equidistribution property of πx, the index set

$$\mathcal{N}'_{\ell,\ell_0} := \{k\in\mathbb{N} : B(\pi x, 2^{-k-\hat{\ell}})\subset\mathcal{D}_{k+\ell_0}(\pi x)\}$$

has the asymptotic density

$$\underline{\mathtt{d}}(\mathcal{N}'_{\ell,\ell_0}) = \mathcal{L}^d[2^{\ell_0-\hat{\ell}}, 1-2^{\ell_0-\hat{\ell}})^d \geq 1 - d\cdot 2^{\ell_0-\hat{\ell}+1}.$$

Moreover,

$$\begin{aligned}
&k\in\mathcal{N}'_{\ell,\ell_0}\\
&\Rightarrow \operatorname{supp}(\pi\nu_{x,k,\ell})\subset B(\pi x,2^{-k-\hat{\ell}})\subset\mathcal{D}_{k+\ell_0}(\pi x)\\
&\Rightarrow H((\pi\nu_{x,k,\ell})^{\mathcal{D}_{k+j}(\pi x)},\mathcal{D}_{N_j\ell_0}) \geq H(\pi\mu^{[x_{-\infty}^{L_{k+\hat{\ell}}(x)}]},\mathcal{D}_{N_j\ell_0+j-\hat{\ell}}) - C
\end{aligned}$$

for some constant $C>0$ depending solely on the IFS but not on ℓ_0, ℓ, or j. Thus, by Lemma 2.8,

$$\alpha \geq (11) \geq (12) \geq \alpha - o_{\ell,\ell_0}(1) \tag{13}$$

has its error $o_{\ell,\ell_0}(1)\to 0$ as $\ell\to\infty$. Writing

$$\begin{aligned}
p_{k,i}(x) &= \left((\pi\nu_{x,k,\ell})^{\mathcal{D}_{k+i}(\pi x)}(D)\right)_{D\in\mathcal{D}_{\ell_0}},\\
q_{k,i}(x) &= \left((\pi\mu)^{\mathcal{D}_{k+i}(\pi x)}(D)\right)_{D\in\mathcal{D}_{\ell_0}},
\end{aligned}$$

one may rephrase (13) as

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\frac{1}{\ell}\sum_{i=1}^{\ell}\frac{D_{\mathrm{KL}}(p_{k,i}(x)||q_{k,i}(x))}{\ell_0\log 2} \leq o_{\ell,\ell_0}(1). \tag{14}$$

Combining this with the observation (5), we may apply Markov's inequality to obtain an index set $\mathcal{N}_{\ell,\ell_0}\subseteq\mathbb{N}$ with $\overline{\mathtt{d}}((\mathcal{N}_{\ell,\ell_0})^c)\leq(o_{\ell,\ell_0}(1))^{1/2}$ such that

$$\begin{aligned}
&k\ell+i\in\mathcal{N}_\ell\\
&\Rightarrow D_{\mathrm{KL}}(p_{k,i}(x)||q_{k,i}(x)) \leq (o_{\ell,\ell_0}(1))^{1/2}\cdot\ell_0\log 2.\\
&\Rightarrow d_{\mathrm{LP}}\left((\pi\nu_{x,k,\ell})^{\mathcal{D}_{k+i}(\pi x)},(\pi\mu)^{\mathcal{D}_{k+i}(\pi x)}\right) \leq \max\{o'_{\ell,\ell_0}(1), d^{1/2}\cdot 2^{-\ell_0}\},
\end{aligned}$$

where $o'_{\ell,\ell_0}(1) = (o_{\ell,\ell_0}(1))^{1/4}\cdot(\ell_0/2\cdot\log 2)^{1/2}$. This finishes the proof of the claim.

The proof of the lemma follows similarly to that of [3, Proposition 3.4]. Let $1 > \epsilon > 0$ be fixed. Take $\ell_1 \in \mathbb{N}$ to be the maximal integer satisfying $2^{-\alpha \ell_1 2^{\ell_1 d}} \geq \epsilon$. Firstly, due to equidistribution property of πx, there exist $\mathcal{N}_\ell^{(1)} \subseteq \mathbb{N}$ with $\overline{\mathsf{d}}((\mathcal{N}_\ell^{(1)})^c) \leq C \cdot 2^{(-\ell_1+1)d} = o_\epsilon(1)$ such that

$$k\ell + i \in \mathcal{N}_\ell^{(1)} \Rightarrow B(\pi x, 2^{-k-\ell_1-i}) \subset \mathcal{D}_{k+i}(\pi x).$$

Next, apply Markov's inequality to (13) (with ℓ_0 replaced by ℓ_1 and lim inf by lim sup) to find a set $\mathcal{N}_\ell^{(2)} \subseteq \mathbb{N}$ with $\overline{\mathsf{d}}((\mathcal{N}_\ell^{(2)})^c) \leq 2 \cdot 2^{-\ell_1 d} = o_\epsilon(1)$ such that

$$k\ell + i \in \mathcal{N}_\ell^{(2)} \Rightarrow \frac{\pi\mu(\mathcal{D}_{k+\ell_1+i}(\pi x))}{\pi\mu(\mathcal{D}_{k+i}(\pi x)))}, \frac{\pi\nu_{x,k,\ell}(\mathcal{D}_{k+\ell_1+i}(\pi x))}{\pi\nu_{x,k,\ell}(\mathcal{D}_{k+i}(\pi x)))} \geq 2^{-\alpha \ell_1 2^{\ell_1 d}} \geq \epsilon.$$

Then, by $\alpha = \dim \pi\mu > 0$, [3, Lemma 3.2], and Markov's inequality, there exist $\rho > 0$ and $\mathcal{N}_\ell^{(3)} \subseteq \mathbb{N}$ with $\overline{\mathsf{d}}((\mathcal{N}_\ell^{(3)})^c) \leq \epsilon^2$ such that

$$k\ell + i \in \mathcal{N}_\ell^{(3)} \Rightarrow \sup_{\rho \leq r \leq 1} (\pi\mu)_{\pi x, (-k-\ell_1-i)\log 2} \left(B(\pi x, r+\rho) \setminus B(\pi x, r-\rho)\right) < \epsilon^2.$$

Finally, for all sufficiently large ℓ, one can find $\mathcal{N}_\ell^{(4)}$, according to our claim, such that $\overline{\mathsf{d}}((\mathcal{N}_\ell^{(4)})^c) < \epsilon$ and

$$k\ell + i \in \mathcal{N}_\ell^{(4)} \Rightarrow d_{\mathrm{LP}}\left((\pi\nu_{x,k,\ell})^{\mathcal{D}_{k+i}(\pi x)}, (\pi\mu)^{\mathcal{D}_{k+i}(\pi x)}\right) < \rho,$$

By [3, Lemma 3.3], the above imply that $\overline{\mathsf{d}}((\cap_{j=1}^4 \mathcal{N}_\ell^{(j)})^c) = o_\epsilon(1)$ and that

$$\begin{aligned} & k\ell + i \in \cap_{j=1}^4 \mathcal{N}_\ell^{(j)} \\ \Rightarrow & \int_0^{\ell \log 2} d_{\mathrm{LP}}\left((\pi\nu_{x,k,\ell})_{\pi x,(k+\ell_1+t)\log 2}, (\pi\mu)_{\pi x,(k+\ell_1+t)\log 2}\right) dt < O(\epsilon), \end{aligned}$$

Altogether, the above guarantee a set $\mathcal{N}_\ell'' \subset \mathbb{N}$ with $\overline{\mathsf{d}}(\mathcal{N}_\ell''^c) \leq (\overline{\mathsf{d}}((\cap_{j=1}^4 \mathcal{N}_\ell^{(j)})^c)^{1/2}$ such that

$$\begin{aligned} & k \in \mathcal{N}_\ell'' \\ \Rightarrow & \frac{1}{\ell \log 2} \int_0^{\ell \log 2} d_{\mathrm{LP}}\left((\pi\nu_{x,k,\ell})_{\pi x,(k+\ell_1+t)\log 2}, (\pi\mu)_{\pi x,(k+\ell_1+t)\log 2}\right) dt < o_{\ell,\ell_0}(1). \end{aligned}$$

At this point, the missing piece of the puzzle lies in that $\{L_{k+\hat{\ell}}(x) : k \in \mathbb{N}\}$ might not agree with the full index set $\mathbb{N}$. Nevertheless, this could be easily remedied by noting that the estimate (10) holds with $L_{k+\hat{\ell}}(x)$ replaced by $L_{k+\hat{\ell}}(x) + j$ for $0 \leq j \leq \max_{x \in \Lambda} \lceil -\log_2 \|D_{\pi\sigma x} f_{x_1}\| \rceil =: J$, from which the remaining argument follows accordingly. Since $\{L_{k+\hat{\ell}}(x) + j : k \in \mathbb{N}, j \in J\}$ contains all sufficiently large integers and the obtained estimate remains valid for each $L_{k+\hat{\ell}}(x) + j$, the proof is concluded by an application of Lemma 3.1. □

Lemma 3.3. *Suppose $\mu \in \mathcal{P}_{erg}(\Lambda^{\mathbb{Z}})$. Then, for μ-a.e. x,*

$$\limsup_{n\to\infty} d_{\mathrm{LP}}\left(\frac{1}{n}\sum_{k=1}^n \frac{1}{\ell\lambda}\int_0^{\ell\lambda} \delta_{(\pi\mu)_{\pi x, t+\lambda_k(x)}} dt, \frac{1}{n\lambda}\int_0^{n\lambda} \delta_{(\pi\mu)_{\pi x, t+\lambda_k(x)}} dt\right) = o_\ell(1).$$

Proof. Due to ergodicity, $\lim_{\ell\to\infty}\frac{\lambda_\ell(x)}{\ell}=\lambda$ for μ-a.e. x, and thus

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}\left|\frac{\lambda_\ell(\sigma^k x)}{\ell}-\lambda\right|=\int\left|\frac{\lambda_\ell(\hat{y})}{\ell}-\lambda\right|d\hat{\mu}(\hat{y})\quad\text{for }\mu\text{-a.e. }x,$$

which diminishes to 0 as $\ell\to\infty$. This proves

$$\limsup_{n\to\infty} d_{\mathrm{LP}}\left(\frac{1}{n}\sum_{k=1}^{n}\frac{1}{\ell\lambda}\int_0^{\ell\lambda}\delta_{(\pi\mu)_{\pi x,t+\lambda_k(x)}}dt,\right.$$
$$\left.\frac{1}{n}\sum_{k=1}^{n}\frac{1}{\ell\lambda}\int_0^{\lambda_\ell(\sigma^k x)}\delta_{(\pi\mu)_{\pi x,t+\lambda_k(x)}}dt\right)=o_\ell(1)$$

with the latter summation immediately yielding the following estimate

$$\lim_{n\to\infty} d_{\mathrm{LP}}\left(\frac{1}{n}\sum_{k=1}^{n}\frac{1}{\ell\lambda}\int_0^{\ell\lambda_\ell(\sigma^k x)}\delta_{(\pi\mu)_{\pi x,t+\lambda_k(x)}}dt,\frac{1}{n\lambda}\int_0^{n\lambda}\delta_{(\pi\mu)_{\pi x,t+\lambda_k(x)}}dt\right)=0.$$

The lemma is hence proved. □

Lemma 3.4. *Suppose $\mu\in\mathcal{P}(\Lambda^{\mathbb{Z}})$ and $\dim\pi\mu>0$. Then, for μ-a.e. x,*

$$\limsup_{n\to\infty} d_{\mathrm{LP}}\left(\frac{1}{n}\sum_{k=1}^{n}\frac{1}{\ell}\int_0^{\ell}\delta_{(\pi\mu)_{\pi x,t+\lambda_k(x)}}dt,\frac{1}{n}\sum_{k=1}^{n}\frac{1}{\ell}\int_0^{\ell}\delta_{\phi(k,x)(\pi\mu^{[x^k_{-\infty}]})_{\pi x,t}}dt\right)=o_\ell(1),$$

where $\phi(k,x)=\|D_{\pi\sigma^n x}f_{x_1^n}\|^{-1}D_{\pi\sigma^n x}f_{x_1^n}$.

Proof. As before, we write $S^{(\epsilon)}$ to mean the ϵ-neighborhood of the set S.

We recall a well-known result in the literature (see, for example, [16, Lemma 4.11] for a discrete version): For every $\epsilon>0$ and μ-a.e. x, there exists $\delta>0$ such that

$$\limsup_{L\to\infty}\frac{1}{L}\mathcal{L}(\{t\in[0,L]:(\pi\mu)_{\pi x,t}(\mathrm{cl}(B_1\setminus B_{1-\delta}))\geq\epsilon\})\leq\epsilon. \tag{15}$$

We herein include a proof for completeness. Suppose to the contrary that (15) fails for some $\epsilon>0$ and some set $S\subset\Lambda^{\mathbb{Z}}$ with $\mu(S)>0$. Then, for every $\delta>0$ and every $x\in S$, there exists a sequence $L_{\delta,i}$ such that the left hand side of (15) converges and is at least ϵ, and that μ generates a tangent distribution P_δ at x along $L_{\delta,i}$, which is a fractal distribution for μ-a.e. $x\in S$ according to [13, Theorem 1.7] and, under this assumption, admits some limiting fractal distribution P of P_δ (see [18, Theorem 1.1] for closedness of the set of fractal distributions). Now that

$$F_\delta:=\{\nu\in\mathcal{P}(B_1):\nu(\mathrm{cl}(B_1\setminus B_{1-\delta}))\geq\epsilon\},\quad\delta>0.$$

is a monotone class of closed sets, $P(F_\delta)\geq\liminf_{\delta'\to0}P_{\delta'}(F_\delta)\geq\epsilon$ for every δ. Consequently,

$$P(\{\nu\in\mathcal{P}(B_1):\nu(\partial B_1)\geq\epsilon\})=P(\cap_{\delta>0}F_\delta)\geq\epsilon,$$

a contradiction to [17, Theorem 3.22], since $\dim P=\dim\mu>0$ according to [13, Proposition 1.19].

The rest of proof is an adaptation of [13, Proposition 1.9] with the underpinning machinery replaced by (15) and Lemma 2.3. To begin with, let $\epsilon>0$ be fixed and take $\delta \in (0, \epsilon)$ so that our claim, Lemma 3.2, and Lemma 3.3 altogether guarantee that for all sufficiently large ℓ,

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \frac{1}{\ell}\mathcal{L}(\{t \in [0,\ell] : (\pi\mu^{[x^-]}|_{[x_1^k]})_{\pi x, t+\lambda_k(x)}(\mathrm{cl}(B_1 \setminus B_{1-\delta})) \geq \epsilon\}) \leq \epsilon. \tag{16}$$

For conciseness, we rephrase

$$\begin{cases} (\pi\mu^{[x^-]}|_{[x_1^k]})_{\pi x, t+\lambda_k(x)} & =: (F_k\nu_k)_{\pi x, t+\lambda_k(x)}, \\ \phi(k,x)(\pi\mu^{[x^-]}|_{[x_1^k]})_{\pi\sigma^k x, t} & =: (G_k\nu_k)_{\pi x, t+\lambda_k(x)}, \end{cases}$$

where $\nu_k = \pi\mu^{[x_{-\infty}^k]}$ and, by writing $T_x(y) = y - x$ for $x, y \in \mathbb{R}^d$,

$$F_k = f_{x_1^k} \quad \text{and} \quad G_k = T_{-\pi x} \circ D_{\pi\sigma^k x} f_{x_1^k} \circ T_{\pi\sigma^k x}.$$

Observe, by Lemma 2.3 and the mean value theorem, that

$$\frac{\|F_k(y) - G_k(y)\|}{\|F_k(y) - F_k(\pi\sigma^k x)\|}, \frac{\|F_k(y) - G_k(y)\|}{\|G_k(y) - G_k(\pi\sigma^k x)\|} \leq C\|y - \pi\sigma^k x\|^\theta \quad \text{for all } y \in V, k \in \mathbb{N},$$

which yields a constant $t_0 = t_0(C,\theta) > 0$ independent of k and ℓ such that for all $t' \geq t_0 + \lambda_k(x)$ and $A \subset B(\pi x, e^{-t'})$,

$$\begin{cases} (G_k\nu_k)(A) \leq (F_k\nu_k)(A^{(e^{-t'}\delta/2)}), \\ (F_k\nu_k)(A) \leq (G_k\nu_k)(A^{(e^{-t'}\delta/2)}). \end{cases} \tag{17}$$

If in addition $(F_k\nu_k)_{\pi x, t'}(\mathrm{cl}(B_1 \setminus B_{1-\delta})) < \epsilon$, then

$$\begin{aligned} (F_k\nu_k)_{\pi x, t'}(A) &\leq (F_k\nu_k)_{\pi x, t'}(A \cap B_{1-\delta}) + \epsilon \\ &\leq \frac{F_k\nu_k(e^{-t'}(A \cap B_{1-\delta}) + \pi x)}{F_k\nu_k(B(\pi x, e^{-t'}))} + \epsilon \\ &\leq \frac{G_k\nu_k(e^{-t'}(1-\delta/2)(A^{(\delta)} \cap B_1) + \pi x)}{G_k\nu_k(B(\pi x, e^{-t'}(1-\delta/2)))} + \epsilon \\ &= (G_k\nu_k)_{\pi x, t' - \log(1-\delta/2)}(A^{(\delta)}) + o_\epsilon(1) \end{aligned} \tag{18}$$

and, similarly,

$$\begin{aligned} (G_k\nu_k)_{\pi x, t'-\log(1-\delta/2)}(A) &= \frac{G_k\nu_k(e^{-t'}(1-\delta/2)A + \pi x)}{G_k\nu_k(B(\pi x, (1-\delta/2)e^{-t'}))} + \epsilon \\ \leq \frac{F_k\nu_k(e^{-t'}(A^{(\delta)} \cap B_1) + \pi x)}{F_k\nu_k(B(\pi x, e^{-t'}(1-\delta)))} &+ \epsilon = (F_k\nu_k)_{\pi x, t'}(A^{(\delta)}) + o_\epsilon(1) \end{aligned} \tag{19}$$

In summary, equations (17)-(19) imply

$$\begin{aligned} & t' \geq t_0 + \lambda_k(x) \text{ and } (F_k\nu_k)_{\pi x, t'}(\mathrm{cl}(B_1 \setminus B_{1-\delta})) < \epsilon \\ \Rightarrow\ & d_{\mathrm{LP}}((F_k\nu_k)_{\pi x, t'}, (G_k\nu_k)_{\pi x, t'-\log(1-\delta/2)}) = o_\epsilon(1). \end{aligned} \tag{20}$$

Given that (16) holds for all sufficiently large ℓ however small $\epsilon > 0$ is and that t_0 is independent of k and ℓ, the desired result follows immediately from (20). $\square$

4. Proof of Theorem 1.1

To begin with, a number of related definitions and notations of group extensions for invertible symbolic systems $(\Lambda^{\mathbb{Z}}, \sigma, \mu)$ are recollected. Recall that a cocycle for the σ-action with value in a group G is a measurable map $\phi : \mathbb{Z} \times \Lambda^{\mathbb{Z}} \to G$ obeying the cocycle equation:

$$\phi(m+n, x) = \phi(m, x) \cdot \phi(n, \sigma^m x) \quad \text{for all } x \in \Lambda^{\mathbb{Z}}, n, m \in \mathbb{Z}.$$

This naturally lifts the measurable system $(\Lambda^{\mathbb{Z}}, \sigma)$ to its *group extension* $(\Lambda^{\mathbb{Z}} \times G, T_\phi)$ with

$$T_\phi^k(x, g) = (\sigma^k x, \phi(k, x)) \quad \text{for all } (x, g) \in \Lambda^{\mathbb{Z}} \times G, k \in \mathbb{Z}.$$

To suit our needs, we fix henceforth G to be the minimal closed subgroup of the orthogonal group $\mathcal{O}(\mathbb{R}^d)$ containing the normalized derivatives of the IFS:

$$\{\|D_{\pi\sigma^n x} f_{x_1^n}\|^{-1} D_{\pi\sigma^n x} f_{x_1^n} : x \in \Lambda^{\mathbb{Z}}, n \in \mathbb{N}\} \subset \mathcal{O}(\mathbb{R}^d) \tag{21}$$

and $\phi : \mathbb{Z} \times \Lambda^{\mathbb{Z}} \to G$ to be the continuous cocycle induced by the conformal IFS, namely,

$$\phi(n, x) = \begin{cases} \|D_{\pi\sigma^n x} f_{x_1^n}\|^{-1} D_{\pi\sigma^n x} f_{x_1^n} & \text{if } n \geq 1, \\ \mathrm{Id} & \text{if } n = 0, \\ \|(D_{\pi\sigma^{n+1} x} f_{x_{n+1}^0})^{-1}\|^{-1} (D_{\pi\sigma^{n+1} x} f_{x_{n+1}^0})^{-1} & \text{if } n \leq -1, \end{cases}$$

which made its debut in Lemma 3.4. Under this specific setting, the ergodic system $(\Lambda^{\mathbb{Z}}, \sigma, \mu)$ can be lifted to an ergodic group extension $(\Lambda^{\mathbb{Z}} \times G, T_\phi, \eta)$ by choosing a typical ergodic component of T_ϕ-invariant measure $\mu \times m_G$, where m_G is the normalized Haar measure on G.

Proof of Theorem 1.1. We first prove that $\pi\mu$ is a scaling measure. As our starting point, consider the approximate scenery flow of $\pi\mu$ at πx:

$$P_{\ell,x,n} = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{\ell} \int_0^\ell \delta_{\phi(k,x)(\pi\mu^{[x_{-\infty}^k]})_{\pi\sigma^k x, t}} dt,$$

for which we claim that for η-a.e. (x, A),

$$P_\ell \equiv AP_{\ell,x} := \lim_{n\to\infty} AP_{\ell,x,n} \quad \text{and} \quad P := \lim_{\ell\to\infty} P_\ell \tag{22}$$

both exist. Indeed, the former limit in (22) enjoys a convergence as a consequence of Birkhoff ergodic theorem, or more explicitly, for η-a.e. (x, A):

$$\lim_{n\to\infty} AP_{\ell,x,n} \equiv \int \frac{1}{\ell} \int_0^\ell \delta_{A'(\pi\mu)_{\pi x', t}} \, dt d\eta(x', A') =: P_\ell. \tag{23}$$

With (23), Lemmas 3.2-3.4 result in that for η-a.e. (x, A),

$$\limsup_{n\to\infty} d_{\mathrm{LP}} \left(P_\ell, \frac{1}{n/\lambda} \int_0^{n/\lambda} \delta_{A(\pi\mu)_{\pi x, t}} dt \right) = o_\ell(1),$$

and thus the existence of P results from the completeness of $\mathcal{P}(\mathcal{P}(B_1))$. In particular, this proves that $\pi\mu$ is a scaling measure.

To prove the ergodicity of P, we assume, for the sake of convenience, that $P = P_x$ for some $x \in \Lambda^{\mathbb{N}}$, which is achieved by a proper choice of η. We then show that P-a.e. ν generates at the origin $A_\nu P$ for some $A_\nu \in \mathcal{O}(\mathbb{R}^d)$, so that the ergodicity of P follows from [13, Theorem 1.6] and the quasi-Palm property. For the described purposes, write $F_\ell(x) = G_\ell((\pi\mu)_{\pi x,0})$ for $x \in \Lambda^{\mathbb{Z}}$ and $\ell \in \mathbb{N}$, where $G_\ell : \mathcal{P}(B_1) \to \mathbb{R}$ is a function defined by

$$G_\ell(\nu) = d_{\mathrm{LP}} \left(\frac{1}{\ell} \int_0^\ell \delta_{\nu_{0,t}} dt, \bigcup_{A' \in \mathcal{O}(\mathbb{R}^d)} A'P \right).$$

As readily checked facts, $\bigcup_{A' \in \mathcal{O}(\mathbb{R}^d)} A'P$ is a closed set, and both F_ℓ and G_ℓ are continuous functions. The first part of theorem asserts that $F_\ell(x) \to 0$ for μ-a.e. x. According to Egorov's theorem, for every $\epsilon > 0$, there exists a compact set K with $P(K) > 1 - \epsilon$ and an $\ell(\epsilon, K) \in \mathbb{N}$ such that

$$x \in K, \ell \geq \ell(\epsilon, K) \Rightarrow F_\ell(x) < \epsilon.$$

By continuity of F_ℓ, there exists an open set $U_\ell \supset K$ such that $F_\ell(x) < \epsilon$ if and only if $x \in U_\ell$. Hence, for μ-a.e. x,

$$1 - \epsilon < P(U_\ell) \leq \liminf_{k \to \infty} \frac{1}{N} \# \left\{ k \in [1, N] : G_\ell(\phi(k, x)(\pi\mu^{[x_{-\infty}^k]})_{\pi\sigma^k x, t}) < \epsilon \right\}.$$

This together with Lemma 3.2 indicates that for every $\epsilon > 0$, there exists $\ell'(\epsilon) \in \mathbb{N}$ such that for μ-a.e. x and all $\ell \geq \ell'(\epsilon)$,

$$\liminf_{k \to \infty} \frac{1}{N} \# \left\{ k \in [1, N] : G_\ell((\pi\mu)_{\pi x, t + \lambda_k(x)}) < \epsilon \right\} > 1 - \epsilon. \tag{24}$$

The proof is finally wrapped up utilizing the estimate

$$\begin{aligned} & s \in [\lambda_k(x), \lambda_{k+1}(x)], \nu \in \mathcal{P}(X) \\ \Rightarrow\ & d_{\mathrm{LP}} \left(\frac{1}{\ell} \int_{\lambda_k(x)}^{\lambda_k(x)+\ell} \delta_{\nu_{0,t}} dt, \frac{1}{\ell} \int_s^{s+\ell} \delta_{\nu_{0,t}} dt \right) \leq \frac{2(\lambda_{k+1}(x) - \lambda_k(x))}{\ell}, \end{aligned}$$

which combined with (4) validates

$$\int G_\ell(\nu) dP(\nu) = \lim_{L \to \infty} \frac{1}{L} \int_0^L G_\ell((\pi\mu)_{\pi x, s}) ds = o_\ell(1),$$

and promises a subsequence $(\ell_i)_{i \in \mathbb{N}}$ satisfying

$$P\left(\left\{ \nu : \limsup_{i \to \infty} G_{\ell_i}(\nu) = 0 \right\} \right) = 1.$$

The proof of our claim is concluded by the uniform scaling property for P-a.e. ν. □

References

1. J. Angelevska, A. Käenmäki, and S. Troscheit, *Self-conformal sets with positive Hausdorff measure*, Bull. Lond. Math. Soc. **52** (2020), no. 1, 200–223.
2. C. Bandt, *The tangent distribution for self-similar measures.*, Lecture at the Fifth Conference on Real Analysis (1992).
3. B. Bárány, A. Käenmäki, A. Pyörälä, and M. Wu, *Scaling limits of self-conformal measures*, 2023, arXiv:2308.11399.
4. K. Falconer, *Techniques in fractal geometry*, John Wiley & Sons, Chichester, England, March 1997 (en).
5. D.-J. Feng and T. Sahlsten, Private communication.
6. D.-J. Feng and H. Hu, *Dimension theory of iterated function systems*, Comm. Pure Appl. Math. **62** (2009), no. 11, 1435–1500.

7. J. Fraser and M. Pollicott, *Uniform scaling limits for ergodic measures*, J. Fractal Geom. **4** (2017), 1–19.
8. H. Furstenberg, *Intersections of Cantor sets and transversality of semigroups*, Problems in analysis (Sympos. Salomon Bochner, Princeton Univ., Princeton, N.J., 1969), Princeton Univ. Press, Princeton, N.J. (1970), 41–59.
9. ———, *Intersections of Cantor sets and transversality of semigroups*, p. 41–60, Princeton University Press, December 2015.
10. M. Gavish, *Measures with uniform scaling scenery*, Ergodic Theory Dynam. Systems **31** (2010), no. 1, 33–48.
11. S. Graf, *On Bandt's tangential distribution for self-similar measures*, Monatsh. Math. **1209** (1995), 223–246.
12. M. Hochman, *Geometric rigidity of $\times m$ invariant measures.*, J. Eur. Math. Soc. **14 (5)** (2012), 1539–1563.
13. ———, *Dynamics on fractals and fractal distributions*, 2013, arXiv:1008.3731.
14. ———, *Lectures on fractal geometry*, Unpublished lecture notes (2023), available online at `http://math.huji.ac.il/~mhochman/courses/fractals-2023/notes.pdf`.
15. M. Hochman and P. Shmerkin, *Local entropy averages and projections of fractal measures*, Ann. Math. **175** (2012), no. 3, 1001–1059.
16. ———, *Equidistribution from fractal measures*, Invent. Math. **202** (2015), no. 1, 427–479.
17. A. Käenmäki, T. Sahlsten, and P. Shmerkin, *Dynamics of the scenery flow and geometry of measures*, Proc. Lond. Math. Soc. **110** (2015), no. 5, 1248–1280.
18. ———, *Structure of distributions generated by the scenery flow*, J. Lond. Math. Soc. **91** (2015), no. 2, 464–494.
19. N. Patzschke and U. Zähle, *Self-similar random measures. IV – The recursive construction model of Falconer, Graf, and Mauldin and Williams*, Math. Nachr. **149** (1990), no. 1, 285–302.
20. A. Pyörälä, *The scenery flow of self-similar measures with weak separation condition*, Ergodic Theory Dynam. Systems (2021), 1–24.
21. A. Tsybakov, *Introduction to nonparametric estimation*, Springer New York, 2009.
22. M. Wu, *A proof of Furstenberg's conjecture on the intersections of $\times p$- and $\times q$-invariant sets*, Ann. Math. **189** (2019), no. 3.
23. M. Wu, *Projection theorems with countably many exceptions and applications to the exact overlaps conjecture*, (2025), arXiv:2503.21923.