DIRECT AND INVERSE RESULTS FOR POPULAR DIFFERENCES IN TREES OF POSITIVE DIMENSION

ALEXANDER FISH AND LEO JIANG;
WITH A JOINT APPENDIX WITH ILYA D. SHKREDOV

Abstract. We establish analogues for trees of results relating the density of a set $E \subset \mathbb{N}$, the density of its set of popular differences, and the structure of E. To obtain our results, we formalise a correspondence principle of Furstenberg and Weiss which relates combinatorial data on a tree to the dynamics of a Markov process. Our main tools are Kneser-type inverse theorems for sets of return times in measure-preserving systems. In the ergodic setting we use a recent result of the first author with Björklund and Shkredov and a stability-type extension (proved jointly with Shkredov); we also prove a new result for non-ergodic systems.

1. Introduction

In [FW03] Furstenberg and Weiss initiated the use of dynamical methods in the study of Ramsey theoretic questions for trees. They proved a Szemerédi-type theorem using a multiple recurrence result for a class of Markov processes (a purely combinatorial proof was later given by Pach, Solymosi, and Tardos [PST12]). More precisely, they showed that finite replicas of the full binary tree could always be found in (infinite) trees of positive growth rate. It is then a natural question to quantify the abundance of finite configurations in a tree in relation to its size as measured by its upper Minkowski and Hausdorff dimensions.

To begin, we review the analogous question in the integer setting. Specifically, we consider the abundance of configurations in a subset $E \subset \mathbb{N}$ of positive density. Recall that the *upper density* and *upper Banach density* of E are

$$\overline{d}(E) = \limsup_{N\to\infty} \frac{|E \cap \{1,\dots,N\}|}{N}, \quad d^*(E) = \limsup_{N-M\to\infty} \frac{|E\cap\{M,\dots,N-1\}|}{N-M}.$$

It is well-known that the abundance of 2-term arithmetic progressions (2-APs) in E can be related to the density of E in the following way. Consider the sets of *popular differences of* E with respect to $\overline{d}$ and d^* defined by

$$\overline{\Delta}_0(E) = \{m \in \mathbb{N} \colon \overline{d}(E\cap(E-m)) > 0\}, \quad \Delta_0^*(E) = \{m\in\mathbb{N}\colon d^*(E\cap(E-m))>0\}.$$

If $\overline{d}(E) > 0$, Furstenberg's correspondence principle [Fur77] states that there exists a measure-preserving system $(X, \mathscr{B}, \nu, S)$ and $A \in \mathscr{B}$ with $\nu(A) = \overline{d}(E)$ such that for all $k \in \mathbb{N}$ and $n_1, \dots, n_k \in \mathbb{N}$

$$\overline{d}(E\cap(E-n_1)\cap\cdots\cap(E-n_k)) \geqslant \nu(A\cap S^{-n_1}A\cap\cdots\cap S^{-n_k}A).$$

Taking $k=1$, it follows that $\overline{\Delta}_0(E)$ contains

$$\mathcal{R} = \mathcal{R}(A) = \{n\in\mathbb{N}\colon \nu(A\cap S^{-n}A)>0\},$$

the set of *return times of* A. Applying the mean ergodic theorem then gives

$$\underline{d}(\overline{\Delta}_0(E)) \geqslant \underline{d}(\mathcal{R}) \geqslant \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\frac{\nu(A\cap S^{-n}A)}{\nu(A)} \geqslant \nu(A) \geqslant \overline{d}(E), \tag{1}$$

where the lower density $\underline{d}$ is defined for $E\subset\mathbb{N}$ by

$$\underline{d}(E) = \liminf_{N\to\infty}\frac{|E\cap\{1,\dots,N\}|}{N}.$$

Date: August 23, 2020.
Key words and phrases. Ramsey theory on trees, return times, inverse theorems, popular difference sets.

If in the above the upper density is replaced by the upper Banach density then ν can further be chosen to be ergodic [Fur81].

We briefly describe how popular difference sets are generalised to trees. Interpreting $m \in \mathbb{N}$ as a parameter determining a configuration $\{a, a+m\}$, we can view elements of $\Delta_0(E)$ as parameters such that the corresponding configurations occur with positive density in E. The tree analogue of $\Delta_0(E)$ is the set $G \subset \mathbb{N}$ of "generic" parameters of certain configurations, depending on a notion of density for sets of vertices in a tree.

Following Furstenberg and Weiss [FW03], we formulate a correspondence principle for arbitrary finite configurations in a tree and use it to obtain analogues of the inequality (1). We then analyse the case of equality in (1) and its analogues for trees using inverse theorems for the set of return times. In the ergodic situation we use a result of Björklund, the first author, and Shkredov [BFS19] and a stability-type extension proved jointly with Shkredov in Appendix A, while in the general case we prove a slightly weaker statement (Theorem 1.2). Using these we obtain inverse theorems for inequality (1): a tree for which equality holds must contain arbitrarily long "arithmetic progressions" with a fixed common difference.

1.1. **Main Results.**

1.1.1. *Tree analogues of popular difference sets.* To describe our results, we first summarise the necessary definitions (see Section 2 for precise formulations).

Fix an integer $q \geqslant 2$. In this paper a tree can be thought of as a directed graph T with a distinguished vertex (the root) having no incoming edges, such that each vertex has between 1 and q outgoing edges and each nonroot vertex has exactly one incoming edge. The "size" of T can be quantified by its upper Minkowski and Hausdorff dimensions $\overline{\dim}_M T$ and $\dim T$.

An n-AP in $E \subset \mathbb{N}$ can be viewed as an affine map $\{1, \ldots, n\} \to E$. Similarly, we consider maps from configurations ("finite trees") C into T satisfying some conditions. In particular, certain configurations F^r for $2 \leqslant r \leqslant q$ are analogues of 2-APs. We say that a vertex $v \in T$ is in $F^r_m = F^r_m(T)$ if there is a directed path of length m from v to a vertex w such that v and w both have at least r outgoing edges. The statement $v \in F^r_m$ corresponds to the statement $n \in E \cap (E-m)$ in the integer setting. Using definitions of upper and upper Banach density for sets of vertices in a tree (denoted by $\overline{d}_T$ and d^*_T), we can define sets of "generic parameters" as analogues of popular difference sets:

$$\overline{G}(F^r) = \{m \in \mathbb{N} \colon \overline{d}_T(F^r_m) > 0\}, \qquad G^*(F^r) = \{m \in \mathbb{N} \colon d^*_T(F^r_m) > 0\}.$$

Our first result is an analogue of (1) for trees:

Theorem A. For any tree T we have

$$\underline{d}(\overline{G}(F^r)) \geqslant \frac{\overline{\dim}_M T - \log_q (r-1)}{1 - \log_q (r-1)} \quad \text{and} \quad \underline{d}(G^*(F^r)) \geqslant \frac{\dim T - \log_q (r-1)}{1 - \log_q (r-1)}.$$

Remark 1.1. The conclusion of Theorem A also holds when F^r is replaced by the full r-ary tree of height 2, which we denote by $D^{r,2}$. See Theorem 4.1.

1.1.2. *Inverse theorems for sets of return times.* Given the direct result Theorem A, we are interested in characterising trees such that equality holds (or almost holds). To illustrate the ideas we consider here the situation when equality is (almost) achieved in (1), which is the analogous question for subsets of $\mathbb{N}$. Observe that the density of the set of return times of A is then close to the measure of A. It is natural to expect in this situation that the dynamics of A under S is structured in some way, and this is indeed the case.

Let $(X, \mathscr{B}, \nu, S)$ be a measure-preserving system, and let A be a measurable set with $\nu(A) > 0$ and set of return times $\mathcal{R}$. Using a theorem of Kneser we prove the following result:

Theorem 1.2. If $\overline{d}(\mathcal{R}) = \nu(A) > 0$, then there exists an integer $k \geqslant 1$ such that up to ν-null sets

$$X = \bigsqcup_{i=0}^{k-1} S^{-i} A.$$

Question 1.3. Does the assumption $\underline{d}(\mathcal{R}) = \nu(A)$ suffice to prove the statement of Theorem 1.2?

If ν is ergodic then Question 1.3 has an affirmative answer, and further there is an inverse result for cases of almost equality. The following theorem is an easy corollary of results by Björklund, the first author, and Shkredov in [BFS19]:

Theorem 1.4. If $(X, \mathscr{B}, \nu, S)$ is ergodic and

$$0 < \underline{d}(\mathcal{R}) < \frac{3}{2}\nu(A),$$

then there exists an integer $k \geqslant 1$ such that $\mathcal{R} = k\mathbb{N}$ and $X = \bigsqcup_{i=0}^{k-1} S^{-i}\left(\bigcup_{j=0}^{\infty} S^{-jk}A\right)$ up to ν-null sets.

Remark 1.5. Example 1.2 in [BFS19] shows that for every $\beta > 1$ there exists a non-ergodic measure-preserving system $(X, \mathscr{B}, \nu, S)$ and $A \in \mathscr{B}$ of arbitrarily small measure such that $\overline{d}(\mathcal{R}) \leqslant \beta\nu(A)$ and there is no $k \geqslant 1$ such that $\mathcal{R} = k\mathbb{N}$.

1.1.3. *Inverse results for popular difference sets.* As a corollary of Theorem 1.2 and Furstenberg's correspondence principle we immediately obtain the following inverse-type result for (1):

Proposition 1.6. Assume that $E \subset \mathbb{N}$ satisfies $\overline{d}(\overline{\Delta}_0(E)) = \overline{d}(E) > 0$. Then there exists $k \geqslant 1$ such that $k\mathbb{N} \subset \overline{\Delta}_0(E)$ and $\overline{d}(\overline{\Delta}_0(E)) = \overline{d}(E) = k^{-1}$. Moreover, for every $m \geqslant 2$

$$\overline{d}\left(\{n \in \mathbb{N} \colon \{n, n+k, \ldots, n+(m-1)k\} \subset E\}\right) = \overline{d}(E).$$

If we consider $\Delta_0^*(E)$ and $d^*(E)$ in place of $\overline{\Delta}_0(E)$ and $\overline{d}(E)$, we can apply Theorem 1.4 to obtain the following inverse result:

Proposition 1.7. Let $1 \leqslant \beta < 3/2$. Assume that $E \subset \mathbb{N}$ satisfies

$$0 < \underline{d}(\Delta_0^*(E)) = \beta \cdot d^*(E).$$

Then there exists $k \geqslant 1$ such that $k\mathbb{N} \subset \Delta_0^*(E)$. Moreover, for every $m \geqslant 2$ that satisfies $(1 - \beta^{-1})m < 1$ we have

$$d^*\left(\{n \in \mathbb{N} \colon \{n, n+k, \ldots, n+(m-1)k\} \subset E\}\right) > 0.$$

1.1.4. *Inverse results for $\overline{G}(F^r)$ and $G^*(F^r)$.* Propositions 1.6 and 1.7 can be interpreted as saying that (almost) equality holds in (1) for a subset $E \subset \mathbb{N}$ only if E is "similar" to the periodic set $k\mathbb{N}$. In the tree setting we prove analogous results.

For every $k \geqslant 1$ and $2 \leqslant r \leqslant q$, define $T^r_{k\mathbb{N}}$ to be the tree such that $v \in T^r_{k\mathbb{N}}$ has q outgoing edges if the directed path from the root to v has length a multiple of k, and $r-1$ outgoing edges otherwise. The inequalities in Theorem A are equalities for $T^r_{k\mathbb{N}}$ (see Subsection 2.0.1).

For every $n \geqslant 1$, define the configuration $V^{r,k,n}$ to be the first n levels of $T^r_{k\mathbb{N}}$. The following two theorems are analogues of Proposition 1.6 and Proposition 1.7 respectively.

Theorem B. Assume that

$$\overline{d}(\overline{G}(F^r)) = \frac{\overline{\dim}_M T - \log_q(r-1)}{1 - \log_q(r-1)} > 0.$$

Then there exists an integer $k \geqslant 1$ such that $\overline{\dim}_M T = k^{-1}(1 - \log_q(r-1)) + \log_q(r-1)$, and $\overline{d}_T(V^{r,k,n}) > 0$ for every $n \geqslant 1$.

Denote by $D^{r,n}$ the full r-ary tree up to height n. Let $D^{r,n}_k$ be the set of vertices in T such that there exists an "affine embedding" $D^{r,n} \to T$ parametrised by k sending the root of $D^{r,n}$ to v (see Section 2 for the definition).

Theorem C. Assume that

$$\underline{d}(G^*(F^r)) = \frac{\dim T - \log_q(r-1)}{1 - \log_q(r-1)} > 0 \quad \text{or} \quad \overline{d}(G^*(D^{r,2})) = \frac{\dim T - \log_q(r-1)}{1 - \log_q(r-1)} > 0.$$

Then there exists an integer $k \geqslant 1$ such that $\dim T = k^{-1}(1 - \log_q(r-1)) + \log_q(r-1)$, and $d^*_T(V^{r,k,n}) > 0$ for every $n \geqslant 1$.

Remark 1.8. We show in Subsection 2.0.2 that Theorem C cannot be improved. Indeed, for every $\varepsilon > 0$ there exists a tree T_ε such that

$$0 < \underline{d}(G^*(F^r)) < (1+\varepsilon)\frac{\dim T - \log_q(r-1)}{1-\log_q(r-1)}$$

and the configuration $V^{r,k,n}$ does not appear at all in T_ε for some large n.

Our final theorem is another partial analogue of Proposition 1.7.

Theorem 1.9. Assume that for $\beta < 3/2$

$$0 < \underline{d}(G^*(F^r)) = \beta \cdot \frac{\dim T - \log_q(r-1)}{1-\log_q(r-1)}.$$

Then there exists $k \geqslant 1$ such that $k\mathbb{N} \subset G^*(F^r)$.

Remark 1.10. It follows from the work of Furstenberg and Weiss in [FW03] that for every n there exists m such that $d^*_T(D^{2,n}_m) > 0$ provided that $\dim T > 0$.

Question 1.11. Under the assumptions of Theorem 1.9, is it true that $d^*_T(D^{r,n}_k) > 0$ for every n satisfying $(1-\beta^{-1})n < 1$?

Organisation of the paper. In Section 2 we describe the combinatorial and dynamical setup. In Section 3 we establish Furstenberg–Weiss correspondence principles relating the density of a finite configuration appearing in T and quantities defined on the associated dynamical system. In Section 4 we prove Theorem A and its extension to $D^{r,2}$. In Section 5 we prove inverse theorems for sets of return times in measure-preserving systems. In Section 6 we prove inverse theorems for popular difference sets for trees (Theorems B, C, and 1.9). In Appendix A we prove a stability-type extension of Theorem 1.4.

Acknowledgments. We thank Michael Björklund and James Parkinson. The current paper is a sequel of joint works of A.F. and I.S. with them. A.F. is grateful to Itai Benjamini for inspiring conversations on the subject during his visit to the Weizmann Institute hosted by Omri Sarig. He thanks Omri and the Weizmann Institute for hospitality. He also thanks Haotian Wu and Cecilia Gonzalez Tokman for fruitful discussions. I.S. is grateful to SMRI and the School of Mathematics and Statistics at the University of Sydney for funding his visit and for their hospitality.

2. Trees and Markov processes

Fix an integer $q \geqslant 2$, and for $2 \leqslant r \leqslant q$ define $\Lambda_r = \{0, \ldots, r-1\}$ and $\Lambda = \Lambda_q$. We set $\mathbb{N} = \{0, 1, \ldots\}$.

Combinatorial setup. In this paper a *tree* is a pruned tree on Λ (in the terminology of descriptive set theory). More explicitly, let $\Lambda^* = \cup_{n=0}^\infty \Lambda^n$ be the set of finite words over Λ. Consider the partial order $\leqslant$ on Λ^* defined by $v \leqslant w$ if w is the concatenation vu of v and some $u \in \Lambda^*$. A tree is then a nonempty subset $T \subset \Lambda^*$ closed under predecessors and having no maximal elements with respect to $\leqslant$. We refer to elements of T as *vertices* (using the natural graph-theoretic terminology), and write $l(v) = n$ if $v \in T(n) = T \cap \Lambda^n$. Every tree contains the empty word $\emptyset$, which is called the *root*, and for every $v \in T$ there is a tree $T^v = \{w \in \Lambda^* \colon vw \in T\}$.

Remark 2.1. Trees are combinatorial realisations of closed sets in $\Lambda^\mathbb{N}$, a symbolic analogue of $[0, 1]$. Given a tree T, the set

$$\{(a_i)_{i\geqslant 0} \in \Lambda^\mathbb{N} \colon (a_0, \ldots, a_n) \in T \text{ for all } n \in \mathbb{N}\}$$

is closed in $\Lambda^\mathbb{N}$ (with the product of discrete topologies on Λ), and there is an inverse map sending a closed subset $A \subset \Lambda^\mathbb{N}$ to the tree

$$\{v \in \Lambda^* \colon vw \in A \text{ for some } w \in \Lambda^\mathbb{N}\}.$$

This motivates several definitions we give below.

The *(upper) Minkowski dimension* of T is

$$\overline{\dim}_M T = \limsup_{n\to\infty} \frac{\log_q |T(n)|}{n}.$$

To define the Hausdorff dimension of a tree, we first define the analogue of an irredundant open cover for trees. A *section* of a tree T is a finite subset $\Pi \subset T$ such that $|\Pi \cap \{w \in T : w \leqslant v\}| = 1$ for all but finitely many $v \in T$. Define also $l(\Pi) = \min\{l(v) : v \in \Pi\}$. Then the *Hausdorff dimension* of T is

$$\dim T = \inf\left\{\lambda > 0 \colon \liminf_{n\to\infty} \inf_{l(\Pi)=n} \sum_{v\in\Pi} q^{-\lambda l(v)} < 1\right\}$$

Example 2.2. Given $E \subset \mathbb{N}$ and $2 \leqslant r \leqslant q$, define the tree

$$T_E^r = \{\emptyset\} \cup \bigcup_{i=0}^{\infty} \prod_{0\leqslant j\leqslant i} \Gamma_j, \qquad \text{where } \Gamma_j = \begin{cases} \Lambda & \text{if } j \in E \\ \Lambda_{r-1} & \text{otherwise.} \end{cases}$$

It is straightforward that

$$\begin{aligned} \overline{\dim}_M T_E^r &= \limsup_{N\to\infty} \frac{\log_q q^{|E\cap\{0,\ldots,N-1\}|}(r-1)^{|E^c\cap\{0,\ldots,N-1\}|}}{N} \\ &= \overline{d}(E) + \log_q(r-1)(1-\overline{d}(E)). \end{aligned}$$

If E is a "periodic" set (such as $k\mathbb{N}$) then T_E^r is "self-similar" and $\overline{\dim}_M T_E^r = \dim T_E^r$.

Elements of Λ^* correspond to cylinder sets of $\Lambda^{\mathbb{N}}$. By the Carathéodory extension theorem, Borel probability measures on $\Lambda^{\mathbb{N}}$ are in bijection with functions $\tau \colon \Lambda^* \to [0,1]$ such that $\tau(\emptyset) = 1$ and $\tau(v) = \sum_{a\in\Lambda} \tau(va)$ for all $v \in \Lambda^*$. We call such functions *Markov trees*, since the support $|\tau| = \{v \in \Lambda^* : \tau(v) > 0\}$ of such a function is a tree. The set of Markov trees is a closed subspace of the compact space $[0,1]^{\Lambda^*}$ with metric $d(\tau_1,\tau_2) = \sum_{v\in\Lambda^*} q^{-l(v)}|\tau_1(v) - \tau_2(v)|$. By abuse of notation we denote it by $\mathcal{P}(\Lambda^{\mathbb{N}})$, since it is homeomorphic to the space of Borel probability measures on $\Lambda^{\mathbb{N}}$ with the weak-* topology.

The *dimension* of a Markov tree [Fur70, Definition 7] is

$$\dim \tau = \liminf_{\substack{l(\Pi)\to\infty \\ \Pi \text{ section of } |\tau|}} \frac{-\sum_{v\in\Pi} \tau(v) \log_q \tau(v)}{\sum_{v\in\Pi} l(v)\tau(v)}.$$

Given a subset $V \subset T$ we define its *upper density*

$$\overline{d_T}(V) = \limsup_{n\to\infty} \frac{1}{|T(n)|} \sum_{v\in T(n)} \frac{|V \cap \{w \in T : w \leqslant v\}|}{n+1}$$

and its *upper Banach density*

$$d_T^*(V) = \sup_{\substack{|\tau|\subset T \\ \dim\tau>0}} \limsup_{n\to\infty} \sup_{v\in|\tau|} \frac{1}{n+1} \sum_{l(w)\leqslant n} \frac{\tau(vw)}{\tau(v)} 1_V(vw).$$

Warning. The upper Banach density $d_T^*(V)$ of $V \subset T$ is not necessarily greater than or equal to the upper density $\overline{d}(V)$.

Example 2.3. If $V = \{v \in T_E^r : l(v) \in E\}$ is the set of q-splitting vertices in T_E^r then $\overline{d_{T_E^r}}(V) = \overline{d}(E)$ and $d^*_{T_E^r}(V) = d^*(E)$ provided $\dim T_E^r > 0$.

We use the term *configuration* to refer to a a nonempty finite subset $C \subset \Lambda^*$ closed under predecessors (a finite tree). Terminology and notation defined above for trees are used for configurations as appropriate without comment. A configuration C is *nonbranching* if $|C(n)| \leqslant 1$ for all $n \in \mathbb{N}$ and *branching* otherwise.

By analogy with arithmetic progressions in $\mathbb{N}$, we consider "affine embeddings" of C in a tree T. More precisely, for a vertex $v \in T$ and $m \in \mathbb{N}$ we say $v \in C_m = C_m(T)$ if there exists an infimum-preserving map of posets $\iota \colon C \to T$ such that $\iota(\emptyset) = v$ and $l(\iota(w)) = l(v) + ml(w)$ for all $w \in C$. Equivalently, we say the configuration C appears at v (with parameter m). Observe that trivially every configuration appears at every vertex with parameter 0.

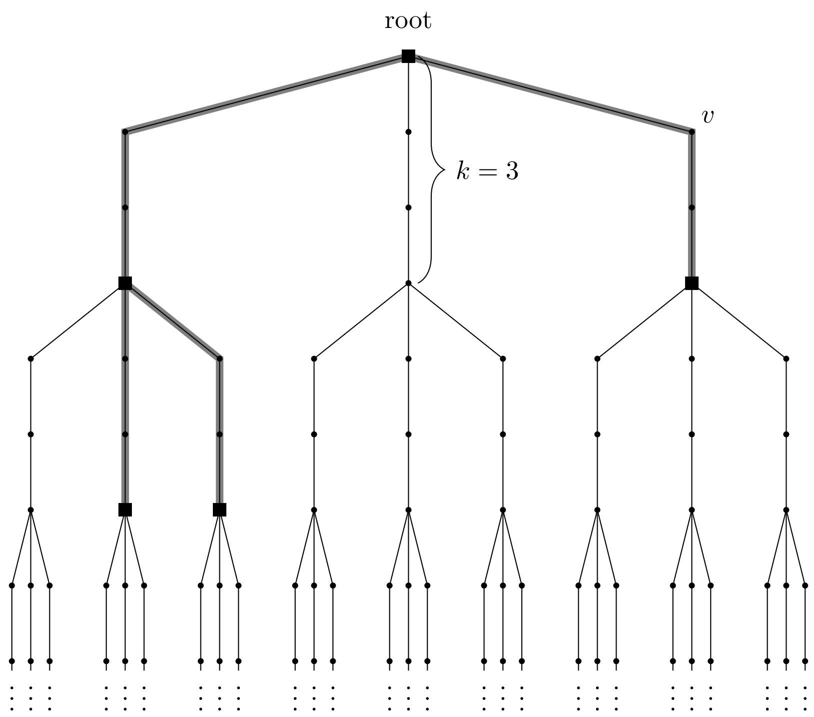

FIGURE 1. The configuration F^2 appears at the root of $T^2_{3\mathbb{N}}$ with parameter 3, while $v \notin F^2_m$ for any $m \geqslant 1$.

We will be concerned with the following configurations (see Figure 1):

$$F^r = \{\emptyset\} \cup \Lambda_r \cup 0\Lambda_r, \;\; D^{r,n} = \bigcup_{i=0}^{n} \Lambda_r^i, \;\; V^{r,k,n} = \{v \in T^r_{k\mathbb{N}} \colon l(v) \leqslant n+1\}.$$

For every configuration C and tree T we define the sets of generic parameters

$$\overline{G}(C) = \overline{G}(C,T) = \{m \in \mathbb{N} \colon \overline{d_T}(C_m) > 0\},$$

$$G^*(C) = G^*(C,T) = \{m \in \mathbb{N} \colon d^*_T(C_m) > 0\}.$$

Remark 2.4. Notice that F^r_m appears in T^r_E if and only if $D^{r,2}_m$ appears in T^r_E if and only if $m \in E - E$. This is why $\overline{G}(F^r)$ and $\overline{G}(D^{r,2})$ are analogues of $\overline{\Delta}_0$ for trees (and similarly for $G^*(F^r)$ and $G^*(D^{r,2})$ with Δ^*_0).

2.0.1. *Equality in Theorems B and C.* The tree $T^r_{k\mathbb{N}}$ achieves equality in Theorems B and C. Indeed,

$$\overline{\dim}_M T^r_{k\mathbb{N}} = \frac{1}{k} + \log_q(r-1)\left(1 - \frac{1}{k}\right).$$

It follows from the self-similarity of $T^r_{k\mathbb{N}}$ that $\dim T^r_{k\mathbb{N}} = \overline{\dim}_M T^r_{k\mathbb{N}}$. Also,

$$\overline{G}(F^r, T^r_{k\mathbb{N}}) = G^*(F^r, T^r_{k\mathbb{N}}) = k\mathbb{N} \quad \text{and} \quad \underline{d}(\overline{G}(F^r, T^r_{k\mathbb{N}})) = \underline{d}(G^*(F^r, T^r_{k\mathbb{N}})) = k^{-1}.$$

Hence

$$\underline{d}(\overline{G}(F^r, T^r_{k\mathbb{N}})) = \frac{\overline{\dim}_M T^r_{k\mathbb{N}} - \log_q(r-1)}{1 - \log_q(r-1)},$$

and

$$\underline{d}(G^*(F^r, T^r_{k\mathbb{N}})) = \frac{\dim T^r_{k\mathbb{N}} - \log_q(r-1)}{1 - \log_q(r-1)}.$$

2.0.2. *Sharpness of Theorem C.* Next, we modify the construction of $T^r_{k\mathbb{N}}$ to obtain for every $\varepsilon > 0$ a tree T_ε with

$$0 < \underline{d}(G^*(F^r, T_\varepsilon)) \leqslant (1+\varepsilon)\frac{\dim T_\varepsilon - \log_q(r-1)}{1-\log_q(r-1)}$$

such that there exists $n \geqslant 1$ with $V_1^{r,k,n}$ not appearing in T_ε.

Let $T_\varepsilon = T^r_E$, where $E = k\mathbb{N} \setminus kN\mathbb{N}$ for some $N \geqslant 1$ satisfying

$$1 - \frac{1}{N} > \frac{1}{1+\varepsilon}.$$

Then

$$\overline{d}(E) = \frac{1}{k}\left(1 - \frac{1}{N}\right)$$

and $V_1^{r,k,kN+1}$ does not appear in T_ε. Since the tree T_ε is self-similar,

$$\dim T_\varepsilon = \overline{\dim}_M T_\varepsilon = \frac{1}{k}\left(1-\frac{1}{N}\right) + \log_q(r-1)\left(1 - \frac{1}{k}\left(1-\frac{1}{N}\right)\right).$$

Since $G^*(F^r, T_\varepsilon) = k\mathbb{N}$, we have

$$\underline{d}(G^*(F^r, T_\varepsilon)) = \frac{1}{k}.$$

Finally, we calculate

$$\begin{aligned}\frac{\dim T_\varepsilon - \log_q(r-1)}{1-\log_q(r-1)} &= \frac{\frac{1}{k}\left(1-\frac{1}{N}\right) + \log_q(r-1)\left(1-\frac{1}{k}\left(1-\frac{1}{N}\right)\right) - \log_q(r-1)}{1-\log_q(r-1)} \\ &= \frac{1}{k}\left(1-\frac{1}{N}\right).\end{aligned}$$

Therefore

$$\underline{d}(G^*(F^r, T_\varepsilon)) = \frac{1}{1-\frac{1}{N}} \cdot \frac{\dim T_\varepsilon - \log_q(r-1)}{1-\log_q(r-1)} < (1+\varepsilon)\frac{\dim T_\varepsilon - \log_q(r-1)}{1-\log_q(r-1)}.$$

Dynamical setup. Given a Markov tree τ and $v \in |\tau|$, define the Markov tree τ^v by $\tau^v(w) = \tau(vw)/\tau(v)$ for every $w \in \Lambda^*$. Using this we define a Markov process on the space $M = \Lambda \times \mathcal{P}(\Lambda^{\mathbb{N}})$, where $a \in \Lambda$ can be interpreted as labelling the root of $\tau \in \mathcal{P}(\Lambda^{\mathbb{N}})$ with information about the past under the dynamics $\tau \mapsto \tau^a$. For simplicity of notation, we omit labels if the expressions considered are labelling independent.

Define $p\colon M \to \mathcal{P}(M)$ by $p(\tau) = p_\tau = \sum_{i\in\Lambda} \tau(i)\delta_{(i,\tau^i)}$. Since p is continuous, it induces a Markov operator P on $C(M)$ (a positive contraction satisfying $P1 = 1$) defined by the formula $Pf(\tau) = \sum_{i\in\Lambda}\tau(i)f(\tau^i)$. The pair (M,p) is an example of a *CP-process.*

By a *distribution* we mean a Borel probability measure. A distribution ν on M is *stationary* for (M,p) if $\int_M Pf\,d\nu = \int_M f\,d\nu$ for all continuous f. Note that if ν is stationary, then the above formula for P extends to a well-defined operator on $L^p(M,\nu)$ for $1 \leqslant p \leqslant \infty$, and by Jensen's inequality this extension is a Markov operator.

For $i \in \Lambda$, define the set $B_i = \{(a,\tau) \in M\colon a = i\}$ of Markov trees labelled by i. The sets B_i are clopen and partition M. Define also for $2 \leqslant r \leqslant q$ the set $A_r = \{\tau \in M\colon |\{i\colon p_\tau(B_i) > 0\}| \geqslant r\}$ of Markov trees τ such that there are at least r vertices in $|\tau|(1)$. Observe that A_r is open and dense in M, and hence is not closed for $r > 1$.

Define on M the *information function*

$$H(\tau) = -\sum_{i\in\Lambda} p_\tau(B_i)\log_q p_\tau(B_i) = -\sum_{i\in\Lambda}\tau(i)\log_q\tau(i),$$

where by convention $0\log_q 0 = 0$. The *entropy* of a stationary distribution ν is then $H(\nu) = \int_M H\,d\nu$. Note that $0 \leqslant H(\tau) \leqslant \log_q n$, where $n = \max\{r\colon \tau \in A_r\}$.

Proposition 2.5. If ν is a stationary distibution for (M,p), then

$$\nu(A_r) \geqslant \frac{H(\nu) - \log_q(r-1)}{1-\log_q(r-1)}.$$

Proof. Using the above bounds on $H(\tau)$ and the definition of A_r,

$$H(\nu) = \int_{A_r} H\,d\nu + \int_{M\setminus A_r} H\,d\nu \leqslant \nu(A_r) + (1-\nu(A_r))\log_q(r-1).$$

Rearranging gives the proposition. □

Endomorphic extension. It will be necessary to work with an extension of the CP-process (M,p), following [FW03].

Let $\widetilde{M} = \{\widetilde{\tau} = (\tau_i)_{i\leqslant 0} \in M^{\mathbb{Z}^{\leqslant 0}} : p_{\tau_i}(\tau_{i+1}) > 0 \text{ for all } i < 0\}$. By abuse of notation we denote by p the natural lift of $p\colon M \to \mathcal{P}(M)$ to a continuous function $\widetilde{M} \to \mathcal{P}(\widetilde{M})$. We also denote by P the corresponding Markov operator on $C(\widetilde{M})$. The pair $(\widetilde{M}, p)$ is said to be an *endomorphic extension* of (M,p).

A stationary distribution ν on M induces a stationary distribution $\widetilde{\nu}$ on $\widetilde{M}$, and by construction $\widetilde{\nu}$ is invariant under the right shift $S\colon (\tau_i)_{i\leqslant 0} \mapsto (\tau_{i-1})_{i\leqslant 0}$. The Koopman operator of S therefore acts on $\mathscr{H} = L^2(\widetilde{M}, \mathscr{B}, \widetilde{\nu})$, where $\mathscr{B}$ is the Borel σ-algebra on $\widetilde{M}$. Since $p_{\widetilde{\tau}}(\{\widetilde{\omega}\}) > 0$ implies $S(\widetilde{\omega}) = \widetilde{\tau}$, a straightforward calculation gives

Lemma 2.6. For any $f, g \in \mathscr{H}$ we have $P(fSg) = gPf$. □

Integrating with respect to $\widetilde{\nu}$ shows that P and S are adjoint operators on $\mathscr{H}$, and taking $f=1$ gives the formula $PS = I$. It follows that S^nP^n is the orthogonal projection from $\mathscr{H}$ onto the closed subspace $S^n\mathscr{H} = L^2(\widetilde{M}, S^{-n}\mathscr{B}, \nu)$.

If $f = Sf' \in S\mathscr{H}$ then $SPf = SPSf' = Sf' = f$, so $SP = I$ on $S\mathscr{H}$. Define $\mathscr{H}_\infty = \cap_{n\geqslant 1} S^n\mathscr{H} = L^2(\widetilde{M}, \mathscr{B}_\infty, \nu)$, where $\mathscr{B}_\infty = \cap_{n\geqslant 1} S^{-n}\mathscr{B}$. For $f \in \mathscr{H}_\infty$ we have $Sf \in \mathscr{H}_\infty$ and $Pf \in \mathscr{H}_\infty$ (using $PS = I$), giving

Lemma 2.7. *P and S restrict to mutually inverse operators $\mathscr{H}_\infty \to \mathscr{H}_\infty$.* □

Denote the orthogonal projection of $f \in \mathscr{H}$ onto $\mathscr{H}_\infty$ by $\overline{f}$.

Proposition 2.8. For $f \in \mathscr{H}$, $\|P^nf - P^n\overline{f}\|_2 \to 0$.

Proof. As $\widetilde{\nu}$ is S-invariant, it follows from Lemma 2.7 that

$$\|P^nf - P^n\overline{f}\|_2 = \|S^nP^nf - S^nP^n\overline{f}\|_2 = \|S^nP^nf - \overline{f}\|_2 \to 0$$

since $\|E(f \mid S^{-n}\mathscr{B}) - E(f \mid \cap_{i\geqslant 1} S^{-i}\mathscr{B})\|_2 \to 0$ [EW11, Theorem 5.8]. □

The information function H lifts naturally to $\widetilde{M}$, and hence the entropy of a stationary distribution for $(\widetilde{M}, p)$ is defined as for (M,p).

3. The Furstenberg–Weiss correspondence principle

In [FW03] Furstenberg and Weiss associated to a tree of positive upper Minkowski dimension a stationary distribution for the CP-process (M,p), and showed that the appearance of $D_m^{2,k}$ could be deduced from the positivity of quantities defined on the dynamical system. In this section we extend their construction to arbitrary configurations, and prove an analogous correspondence principle based on [Fur70] for trees of positive Hausdorff dimension.

3.1. **Construction of configuration-detecting functions.** Given a configuration C and an integer $m \geqslant 1$, we say that a function $f\colon M \to [0,1]$ is *C_m-detecting* if $f(\tau) > 0$ if and only if C appears at the root of $|\tau|$ with parameter m. In preparation for proving correspondence principles we construct recursively several families of configuration-detecting functions.

We first construct a set of configuration-detecting functions φ_{C_m}. For the simplest configuration $\{\emptyset\}$, we can take $\varphi_{\{\emptyset\}_m} = 1$ for all $m \geqslant 1$. Given $I \subset \Lambda$ such that $|I| = |C(1)|$ and a bijection $\beta\colon I \to C(1)$, the positivity of $\prod_{i\in I} P(1_{B_i}P^{m-1}\varphi_{C_m^{\beta(i)}})$ at $\tau \in M$ is equivalent to the appearance of C at the root of $|\tau|$ with parameter m such that $\beta(i) \in C(1)$ is mapped to $iv \in |\tau|$ for some $v \in \Lambda^{m-1}$. Summing over all choices of I and β, we define φ_{C_m} by the recursive formula

$$\varphi_{C_m} = \sum_{\substack{I\subset\Lambda\\ |I|=|C(1)|}} \sum_{\beta\colon I \xrightarrow{\sim} C(1)} \prod_{i\in I} P(1_{B_i}P^{m-1}\varphi_{C_m^{\beta(i)}}).$$

We also have $0 \leqslant \varphi_{C_m} \leqslant 1$. Indeed, since $\varphi_{\{\emptyset\}_m} = 1$ and P is positive

$$0 \leqslant \varphi_{C_m} \leqslant \sum_{\substack{I\subset\Lambda \\ |I|=|C(1)|}} \sum_{\beta\colon I \xrightarrow{\sim} C(1)} \prod_{i\in I} P1_{B_i} \leqslant \left(\sum_{i\in\Lambda} P1_{B_i}\right)^{|C(1)|} = 1.$$

Starting instead with $\phi_{D_m^{r,1}} = 1_{A_r}$ and $\phi_{C_m} = 1$ for nonbranching configurations C, we can adapt the above recursion to construct an alternative family of configuration-detecting functions $\phi_{C_m} \geqslant \varphi_{C_m}$ more suitable for computations. Let $C(1)' = \{v \in C(1)\colon C^v \text{ is branching}\}$, and set $n' = |C(1)'| \leqslant |C(1)| =: n$. We define ϕ_{C_m} recursively by the formula

$$\phi_{C_m} = 1_{A_n} \sum_{\substack{I\subset\Lambda \\ |I|=n'}} \sum_{\beta\colon I \xrightarrow{\sim} C(1)'} \prod_{i\in I} P(1_{B_i} P^{m-1} \phi_{C_m^{\beta(i)}}).$$

Note that $\varphi_{D_m^{r,1}} \leqslant 1_{A_r} = \phi_{D_m^{r,1}}$. Similarly we have $0 \leqslant \phi_{C_m} \leqslant 1$.

As the B_i are clopen and P takes continuous functions to continuous functions, the φ_{C_m} are continuous. However, the ϕ_{C_m} are in general not continuous since A_r is not clopen for $r > 1$.

If C is a configuration such that the configurations C^v are all "isomorphic" for $v \in C(1)$, the above recursion can be simplified by omitting the sum over bijections β. For integers $2 \leqslant r \leqslant q$ and $m \geqslant 1$, define (nonlinear) operators $R_{r,m}$ on $L^\infty(M)$ by

$$R_{r,m} f = \sum_{\substack{I\subset\Lambda \\ |I|=r}} \prod_{i\in I} P(1_{B_i} P^{m-1} f).$$

If f detects C_m^v for (all) $v \in C(1)$ and $|C(1)| = r$, then $R_{r,m} f$ detects C_m. Denote by ϕ'_{C_m} the C_m-detecting function obtained by applying a sequence of the above operators to the appropriate 1_{A_r}, and observe that $\phi_{C_m} = c\phi'_{C_m}$ for some integer $c \geqslant 1$.

Example 3.1. For the configuration F^r, we have $n = r$ and $n' = 1$. There is always a unique bijection $I \to C(1)'$, so linearity of P gives

$$\phi_{F_m^r} = 1_{A_r} \sum_{i\in\Lambda} P(1_{B_i} P^{m-1} 1_{A_r}) = 1_{A_r} P^m 1_{A_r}$$

since $1 = \sum_{i\in\Lambda} 1_{B_i}$.

If $n = n'$, the factor 1_{A_n} is redundant in the definition of ϕ_{C_m} as the function in the sum is already supported on a subset of A_n. For example,

$$\phi_{D_m^{r,2}} = \sum_{\substack{I\subset[q] \\ |I|=r}} \sum_{\beta\colon I \xrightarrow{\sim} \Lambda_r} \prod_{i\in I} P(1_{B_i} P^{m-1} 1_{A_r}) = r! \sum_{\substack{I\subset[q] \\ |I|=r}} \prod_{i\in I} P(1_{B_i} P^{m-1} 1_{A_r}) = r! \phi'_{D_m^{r,2}}.$$

The following lemma is used in the proofs of the correspondence principles to account for the lack of continuity of ϕ_{C_m}.

Lemma 3.2. If $(\nu_k)_{k\geqslant 1}$ is a sequence of distributions on M converging to ν in the weak-* topology, then for every configuration C and integer $m \geqslant 1$

$$\limsup_{k\to\infty} \int_M \phi_{C_m}\, d\nu_k \geqslant \int_M \phi_{C_m}\, d\nu.$$

Proof. Define for $\delta \in [0,1]$ open sets $A_r^\delta = \{\tau \in M\colon |\{i\colon p_\tau(B_i) > \delta\}| \geqslant r\} \subset A_r$, and let $\phi_{C_m}^\delta$ be the function obtained by replacing 1_{A_r} with $1_{A_r^\delta}$ in the recursive construction of ϕ_{C_m}. Observe that $\delta \leqslant \delta'$ implies $\phi_{C_m}^\delta \geqslant \phi_{C_m}^{\delta'}$ by the positivity of P. Then the monotone function $\alpha\colon \delta \mapsto \int_M \phi_{C_m}^\delta\, d\nu$ has countably many discontinuities, so we can choose a sequence $\delta_n \to 0$ such that α is continuous at δ_n for all n.

We claim $\int_M \phi_{C_m}^\delta\, d\nu_k \to \int_M \phi_{C_m}^\delta\, d\nu = \alpha(\delta)$ if α is continuous at δ. If $\delta < \delta'$, the closed sets $(A_r^\delta)^c$ and $\overline{A_r^{\delta'}} = \{\tau \in M\colon |\{i\colon p_\tau(B_i) \geqslant \delta'\}| \geqslant r\}$ are disjoint since $\overline{A_r^{\delta'}} \subset A_r^\delta$. By Urysohn's lemma there are continuous functions h_r such that $1_{A_r^{\delta'}} \leqslant h_r \leqslant 1_{A_r^\delta}$. Defining h_{C_m} to be the function obtained by repeating the construction of ϕ_{C_m} with h_r in place of 1_{A_r}, it follows that $\phi^{\delta'}_{C_m} \leqslant h_{C_m} \leqslant \phi^{\delta}_{C_m}$. Since h_{C_m} is continuous,

$$\liminf_{k\to\infty} \int_M \phi^{\delta}_{C_m}\, d\nu_k \geqslant \liminf_{k\to\infty} \int_M h_{C_m}\, d\nu_k = \int_M h_{C_m}\, d\nu \geqslant \int_M \phi^{\delta'}_{C_m}\, d\nu = \alpha(\delta').$$

Continuity of α at δ implies $\liminf_{k\to\infty} \int_M \phi^{\delta}_{C_m}\, d\nu_k \geqslant \alpha(\delta)$, and a similar argument with $\delta' < \delta$ proves the claim. Hence

$$\limsup_{k\to\infty} \int_M \phi_{C_m}\, d\nu_k \geqslant \lim_{k\to\infty} \int_M \phi^{\delta_n}_{C_m}\, d\nu_k = \int_M \phi^{\delta_n}_{C_m}\, d\nu \xrightarrow{n\to\infty} \int_M \phi_{C_m}\, d\nu$$

by the monotone convergence theorem. □

3.2. **Correspondence principle for upper density.**

Theorem 3.3. For every tree T with $\overline{\dim}_M T > 0$, the CP-process (M, p) has a stationary distribution μ such that $H(\mu) = \overline{\dim}_M T$,

$$\mu(A_r) \geqslant \frac{\overline{\dim}_M T - \log_q(r-1)}{1 - \log_q(r-1)}, \tag{2}$$

and for every configuration C and every integer $m \geqslant 1$

$$\overline{d_T}(C_m) \geqslant \int_M \phi_{C_m}\, d\mu. \tag{3}$$

Proof. Let $(L_k)_{k\geqslant 1}$ be an increasing sequence such that

$$\overline{\dim}_M T = \lim_{k\to\infty} \frac{\log_q |T(L_k+1)|}{L_k+1}.$$

Fix an arbitrary label $a \in \Lambda$, and for each $k \geqslant 1$ let π_k be any Markov tree labelled by a such that $\pi_k(v) = |T(L_k)|^{-1}$ for all $v \in T(L_k)$ (note that this condition determines π_k on vertices of level at most L_k). Then any weak-* limit of the distributions

$$\mu_k = \frac{1}{L_k+1} \sum_{n=0}^{L_k} P^n \delta_{\pi_k} = \frac{1}{L_k+1} \sum_{l(v)\leqslant L_k} \pi_k(v) \delta_{\pi_k^v}$$

is stationary, and we choose μ to be such a limit.

Since $H(x)$ is continuous and $\pi_k(v) = \sum_{a\in\Lambda} \pi_k(va)$,

$$\begin{aligned}
H(\mu) &= \lim_{k\to\infty} \int_M H\, d\mu_k \\
&= -\lim_{k\to\infty} \frac{1}{L_k+1} \sum_{l(v)\leqslant L_k} \pi_k(v) \sum_{a\in\Lambda} \frac{\pi_k(va)}{\pi_k(v)} \log_q \frac{\pi_k(va)}{\pi_k(v)} \\
&= -\lim_{k\to\infty} \frac{1}{L_k+1} \sum_{l(v)\leqslant L_k} \sum_{a\in\Lambda} \pi_k(va) \log_q \pi_k(va) - \pi_k(va) \log_q \pi_k(v) \\
&= -\lim_{k\to\infty} \frac{1}{L_k+1} \sum_{l(v)=L_k} \sum_{a\in\Lambda} \pi_k(va) \log_q \pi_k(va). \qquad (4)
\end{aligned}$$

Recall that for every $v \in |\pi_k|$ we have the bounds

$$-\pi_k(v) \log_q \pi_k(v) \leqslant -\sum_{a\in\Lambda} \pi_k(va) \log_q \pi_k(va) \leqslant -\pi_k(v) \log_q \frac{\pi_k(v)}{q}. \tag{5}$$

Since $-\sum_{l(v)=L_k} \pi_k(v) \log_q \pi_k(v) = \log_q |T(L_k)|$ by definition of π_k, summing the inequalities (5) over $v \in L_k$ and noticing $\sum_{l(v)=L_k} \pi_k(v) = 1$ gives

$$H(\mu) = \int_X H\, d\mu = \lim_{k\to\infty} \frac{\log_q |T(L_k)|}{L_k+1} = \lim_{k\to\infty} \frac{\log_q |T(L_k+1)|}{L_k+1} = \overline{\dim}_M T.$$

where the third equality follows from the bounds

$$q^{-1}|T(L_k+1)| \leqslant |T(L_k)| \leqslant |T(L_k+1)|.$$

Proposition 2.5 immediately gives the inequality (2).

To prove the inequality (3), applying a change of summation variable and using the definitions of π_k and ϕ_{C_m} gives

$$\begin{aligned}\overline{d_T}(C_m) &\geqslant \limsup_{k\to\infty} \frac{1}{|T(L_k)|} \sum_{v\in T(L_k)} \frac{|C_m \cap \{w\in T\colon w\leqslant v\}|}{L_k+1} \\ &= \limsup_{k\to\infty} \frac{1}{L_k+1} \sum_{l(w)\leqslant L_k} \pi_k(w) 1_{C_m}(w) \\ &\geqslant \limsup_{k\to\infty} \frac{1}{L_k+1} \sum_{l(w)\leqslant L_k} \pi_k(w) \phi_{C_m}(\pi_k^w) \\ &= \limsup_{k\to\infty} \int_M \phi_{C_m}\, d\mu_k.\end{aligned}$$

The conclusion follows from Lemma 3.2. □

3.3. Correspondence principle for upper Banach density.

Theorem 3.4. If $\dim T > 0$, for every $\epsilon > 0$ there exists an ergodic stationary distribution $\eta = \eta_\epsilon$ for the CP-process (M, p) such that $H(\eta) \geqslant \dim T - \epsilon$,

$$\eta_\epsilon(A_r) \geqslant \frac{\dim T - \epsilon - \log_q(r-1)}{1-\log_q(r-1)}, \tag{6}$$

and for every configuration C and integer $m \geqslant 1$

$$d^*_T(C_m) \geqslant \int_M \phi_{C_m}\, d\eta. \tag{7}$$

Proof. For any $\epsilon > 0$, by Frostman's lemma there exists $\theta \in M$ such that $|\theta| \subset T$ and $\dim\theta \geqslant \dim T - \epsilon$. Let $(M_k)_{k\geqslant 1}$ be an increasing sequence such that the distributions

$$\eta'_k = \frac{1}{M_k+1} \sum_{n=0}^{M_k} P^n \delta_\theta = \frac{1}{M_k+1} \sum_{l(v)\leqslant M_k} \theta(v) \delta_{\theta^v}.$$

converge to a distribution η' in the weak-* topology.

Lemma 3.5. [Fur70, Lemma 4] $H(\eta') \geqslant \dim\theta$.

Proof. As in the calculation (4) we have

$$\begin{aligned}H(\eta') &= -\lim_{k\to\infty} \frac{1}{M_k+1} \sum_{l(v)=M_k+1} \theta(v)\log_q\theta(v) = \lim_{k\to\infty} \frac{-\sum_{v\in\Pi_k}\theta(v)\log_q\theta(v)}{\sum_{v\in\Pi_k} l(v)\theta(v)} \\ &\geqslant \dim\theta\end{aligned}$$

where Π_k is the section $|\theta|(M_k+1) = \{v \in |\theta|\colon l(v) = M_k+1\}$. □

The support of η' is contained in the compact set $D(\theta) = \overline{\{\theta^v\colon v\in|\theta|\}}$, and the ergodic decomposition of η' gives an ergodic distribution η supported on $D(\theta)$ such that $H(\eta) \geqslant H(\eta') \geqslant \dim\theta \geqslant \dim T - \epsilon$. The inequality (6) immediately follows from Proposition 2.5.

Since η is ergodic, the mean ergodic theorem for contractions [EFHN15, Theorem 8.6] implies

$$\frac{1}{N+1}\sum_{n=0}^{N} P^n f \to \int_M f\, d\eta$$

almost everywhere for $f \in L^2(M,\eta)$. By diagonalisation there exists an increasing sequence $(N_k)_{k\geqslant 1}$ and $\tau \in D(\theta)$ such that

$$\frac{1}{N_k+1}\sum_{n=0}^{N_k} P^n f(\tau) \to \int_M f\, d\eta \tag{8}$$

for all f in a countable set of continuous functions. Taking this set to be dense in $C(M)$ under the uniform norm, the limit (8) holds for all continuous functions. Letting $v_k \in |\theta|$ be a sequence of vertices such that $\theta^{v_k} \to \tau$, it follows that

$$\eta_k = \frac{1}{N_k+1}\sum_{n=0}^{N_k} P^n \delta_{\theta^{v_k}} = \frac{1}{N_k+1}\sum_{l(w)\leqslant N_k} \theta^{v_k}(w)\delta_{\theta^{v_k w}}$$

converges weakly to η. For $\epsilon < \dim T$ it follows that

$$\begin{aligned} d_T^*(C_m) &\geqslant \limsup_{k\to\infty} \frac{1}{N_k+1}\sum_{l(w)\leqslant N_k} \theta^{v_k}(w) 1_{C_m}(vw) \\ &\geqslant \limsup_{k\to\infty} \frac{1}{N_k+1}\sum_{l(w)\leqslant N_k} \theta^{v_k}(w)\phi_{C_m}(\theta^{v_k w}) \\ &= \limsup_{k\to\infty} \int_M \phi_{C_m}\, d\eta_k \end{aligned}$$

and the inequality (7) follows from Lemma 3.2. ☐

Remark 3.6. Composing the projection $(\tau_i)_{i\leqslant 0} \mapsto \tau_0$ with a C_m-detecting function gives a map $\widetilde{M} \to [0,1]$ which is positive at $(\tau_i)_{i\leqslant 0}$ if and only if C appears at the root of $|\tau_0|$ with parameter m. The recursive constructions of configuration-detecting functions in Subsection 3.1 can be used to construct their lifts (which we denote by the same notation) using the abuses of notation $B_i = \{\widetilde{\tau} \in M : \tau_0 \in B_i\}$ and $A_r = \{\widetilde{\tau} \in \widetilde{M} : |\{i : p_{\widetilde{\tau}}(B_i) > 0\}| \geqslant r\}$. Observe that the inequalities (2), (3), (6), and (7) are still valid when the distributions μ and η_ϵ and the configuration detecting functions ϕ_{C_m} are replaced with their lifts on $\widetilde{M}$. In the remainder of the paper we work only with $(\widetilde{M}, p)$ and use Theorems 3.3 and 3.4 for the endomorphic extension without comment.

4. Proof of direct theorems

Using the correspondence principles of Section 3, we bound from below the densities of the sets of generic parameters for F^r and $D^{r,2}$. To illustrate the method we first give a simple proof of Theorem A, which is also an immediate corollary of Theorem 4.1.

Theorem A. Let T be a tree. For $2 \leqslant r \leqslant q$ we have

$$\underline{d}(\overline{G}(F^r)) \geqslant \frac{\overline{\dim}_M T - \log_q(r-1)}{1-\log_q(r-1)} \text{ and } \underline{d}(G^*(F^r)) \geqslant \frac{\dim T - \log_q(r-1)}{1-\log_q(r-1)}.$$

Proof. Since P and S are adjoint, Theorem 3.3 gives

$$\overline{d_T}(F_m^r) \geqslant \int_{\widetilde{M}} \phi_{F_m^r}\, d\widetilde{\mu} = \int_{\widetilde{M}} 1_{A_r} P^m 1_{A_r}\, d\widetilde{\mu} = \int_{\widetilde{M}} 1_{A_r} S^m 1_{A_r}\, d\widetilde{\mu} = \widetilde{\mu}(A_r \cap S^{-m} A_r).$$

Hence $\overline{G}(F^r) \supset \mathcal{R} = \{n \in \mathbb{N} : \widetilde{\mu}(A_r \cap S^{-n}A_r) > 0\}$, so $\underline{d}(\overline{G}(F^r)) \geqslant \underline{d}(\mathcal{R})$. By the mean ergodic theorem

$$\underline{d}(\mathcal{R}) = \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} 1_{\mathcal{R}}(n) \geqslant \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \frac{\widetilde{\mu}(A_r \cap S^{-n}A_r)}{\widetilde{\mu}(A_r)} \geqslant \widetilde{\mu}(A_r),$$

and the theorem follows from inequality (2) of Theorem 3.3.

Using Theorem 3.4 in place of Theorem 3.3 in the above argument, we obtain the second inequality after taking $\epsilon \to 0$. ☐

Theorem 4.1. Let T be a tree. For $2 \leqslant r \leqslant q$ we have

$$\underline{d}(\overline{G}(D^{r,2})) \geqslant \frac{\overline{\dim}_M T - \log_q(r-1)}{1-\log_q(r-1)} \text{ and } \underline{d}(G^*(D^{r,2})) \geqslant \frac{\dim T - \log_q(r-1)}{1-\log_q(r-1)}.$$

Proof. We start with the proof of the first inequality. The idea is to show that $\overline{G}(D^{r,2})$ essentially contains the set of return times of A_r, the density of which we can bound from below by the mean ergodic theorem. First observe that by Proposition 2.8

$$\left| \int_{\widetilde{M}} \phi_{D_m^{r,2}} \, d\widetilde{\mu} - r! \int_{\widetilde{M}} \sum_{\substack{I \subset [q] \\ |I|=r}} \prod_{i \in I} P(1_{B_i} P^{m-1} \overline{1_{A_r}}) \, d\widetilde{\mu} \right|$$

$$= \left| r! \int_{\widetilde{M}} \sum_{\substack{I \subset [q] \\ |I|=r}} \prod_{i \in I} P(1_{B_i} P^{m-1} 1_{A_r}) \, d\widetilde{\mu} - r! \int_{\widetilde{M}} \sum_{\substack{I \subset [q] \\ |I|=r}} \prod_{i \in I} P(1_{B_i} P^{m-1} \overline{1_{A_r}}) \, d\widetilde{\mu} \right| \xrightarrow{m \to \infty} 0.$$

Since $\overline{1_{A_r}} \in \mathscr{H}^\infty$, by Lemma 2.7 $P^{m-1}\overline{1_{A_r}} = SP^m\overline{1_{A_r}}$. Then by Lemma 2.6 and orthogonality

$$r! \int_{\widetilde{M}} \sum_{\substack{I \subset [q] \\ |I|=r}} \prod_{i \in I} P(1_{B_i} P^{m-1} \overline{1_{A_r}}) \, d\widetilde{\mu} = \int_{\widetilde{M}} \overline{\varphi_{D_m^{r,1}}} (P^m \overline{1_{A_r}})^r \, d\widetilde{\mu},$$

recalling $\varphi_{D_m^{r,1}} = r! \sum_{\substack{I \subset [q] \\ |I|=r}} \prod_{i \in I} P 1_{B_i}$. Define $Z_\rho = \{\widetilde{\tau} \in \widetilde{M} \colon \overline{\varphi_{D_m^{r,1}}}(\widetilde{\tau}) \geqslant \rho\}$, and observe it is well-defined up to a $\widetilde{\mu}$-null set. Since $0 \leqslant \varphi_{D_m^{r,1}} \leqslant 1_{A_r} \leqslant 1$ and $0 \leqslant \rho 1_{Z_\rho} \leqslant \overline{\varphi_{D_m^{r,1}}} \leqslant 1$, the positivity of both P and conditional expection imply

$$\int_{\widetilde{M}} \overline{\varphi_{D_m^{r,1}}} (P^m \overline{1_{A_r}})^r \, d\widetilde{\mu} \geqslant \int_{\widetilde{M}} \overline{\varphi_{D_m^{r,1}}} (P^m \overline{\varphi_{D_m^{r,1}}})^r \, d\widetilde{\mu} \geqslant \rho^{r+1} \int_{\widetilde{M}} 1_{Z_\rho} (P^m 1_{Z_\rho})^r \, d\widetilde{\mu}.$$

By Jensen's inequality and the adjointness of P and S

$$\int_{\widetilde{M}} 1_{Z_\rho} (P^m 1_{Z_\rho})^r \, d\widetilde{\mu} \geqslant \left(\int_{\widetilde{M}} 1_{Z_\rho} P^m 1_{Z_\rho} \, d\widetilde{\mu} \right)^r = \widetilde{\mu}(Z_\rho \cap S^{-m} Z_\rho)^r.$$

Combining the above with the correspondence principle Theorem 3.3, it follows that $\overline{G}(D^{r,2})$ contains a cofinite subset of

$$\mathcal{R}^\delta(Z_\rho) = \{m \in \mathbb{N} \colon \widetilde{\mu}(Z_\rho \cap S^{-m} Z_\rho) > \delta \widetilde{\mu}(Z_\rho)^2\}$$

for all $\delta, \rho > 0$ (since S is $\widetilde{\mu}$-preserving). Therefore

$$\begin{aligned} \underline{d}(\overline{G}(D^{r,2})) \geqslant \underline{d}(\mathcal{R}^\delta(Z_\rho)) &\geqslant \liminf_{N \to \infty} \frac{1}{N} \sum_{\substack{m \leqslant N \\ m \in \mathcal{R}^\delta(Z_\rho)}} \frac{\widetilde{\mu}(Z_\rho \cap S^{-m} Z_\rho)}{\widetilde{\mu}(Z_\rho)} \\ &\geqslant \left(\liminf_{N \to \infty} \frac{1}{N} \sum_{m=1}^{N} \frac{\widetilde{\mu}(Z_\rho \cap S^{-m} Z_\rho)}{\widetilde{\mu}(Z_\rho)} \right) - \delta \widetilde{\mu}(Z_\rho) \\ &\geqslant (1-\delta)\widetilde{\mu}(Z_\rho) \xrightarrow{\delta \to 0} \widetilde{\mu}(Z_\rho) \xrightarrow{\rho \to 0} \widetilde{\mu}(Z), \end{aligned}$$

where the last inequality follows from the mean ergodic theorem and

$$Z = \{\widetilde{\tau} \in \widetilde{M} \colon \overline{\varphi_{D_m^{r,1}}}(\widetilde{\tau}) > 0\}.$$

Since $1_{Z^c} \in L^\infty(\widetilde{M}, \mathscr{B}_\infty)$, properties of the conditional expectation give

$$0 = \int_{\widetilde{M}} 1_{Z^c} \overline{\varphi_{D_m^{r,1}}} \, d\widetilde{\mu} = \int_{\widetilde{M}} 1_{Z^c} \varphi_{D_m^{r,1}} \, d\widetilde{\mu}.$$

Hence $Z \supset \{\widetilde{\tau} \in \widetilde{M} \colon \varphi_{D_m^{r,1}}(\widetilde{\tau}) > 0\} = A_r$ up to a $\widetilde{\mu}$-null set, so $\underline{d}(\overline{G}(D^{r,2})) \geqslant \widetilde{\mu}(A_r)$. The theorem then follows from inequality (2) of Theorem 3.3.

Using Theorem 3.4 in place of Theorem 3.3 in the above proofs, we obtain the second inequality after taking $\epsilon \to 0$. $\square$

5. Inverse theorems for return times

Let $(X, \mathscr{B}, \nu, S)$ be a measure-preserving system, and let A be a measurable set with $\nu(A) > 0$. If $\mathcal{R} = \{n \in \mathbb{N} : \nu(A \cap S^{-n}A) > 0\}$ is the set of return times of A, then by the mean ergodic theorem $\underline{d}(\mathcal{R}) \geqslant \nu(A)$.

Theorem 1.2. If $\overline{d}(\mathcal{R}) = \nu(A) > 0$, then there exists an integer $k \geqslant 1$ such that up to ν-null sets

$$X = \bigsqcup_{i=0}^{k-1} S^{-i}A.$$

Proof. Define $\mathcal{R}_\gamma = \{n \in \mathbb{N} : \nu(A \cap S^{-n}A) \geqslant (1-\gamma)\nu(A)\}$.

Lemma 5.1. If $m \in \mathcal{R}_\epsilon$ and $n \in \mathcal{R}_\gamma$, then $m + n \in \mathcal{R}_{\epsilon+\gamma}$.

Proof. If $B \subset A$ then $\nu(A \cap S^{-m}B) \geqslant \nu(B) - \epsilon\nu(A)$. For $B = A \cap S^{-n}A$ we have

$$\nu(A \cap S^{-(m+n)}A) \geqslant \nu(A \cap S^{-m}(A \cap S^{-n}A)) \geqslant \nu(A \cap S^{-n}A) - \epsilon\nu(A) \geqslant (1-\gamma-\epsilon)\nu(A),$$

so $m + n \in \mathcal{R}_{\epsilon+\gamma}$. □

Lemma 5.2. If $0 < \gamma < \frac{1}{2}$, then $d(\mathcal{R}_\gamma + \mathcal{R}_\gamma) = d(\mathcal{R}_\gamma) = d(\mathcal{R})$.

Proof. Let $(N_k)_{k\geqslant 1}$ be an increasing sequence such that

$$d_{N_k}(\mathcal{R}_\gamma) = \lim_{k\to\infty} \frac{|\mathcal{R}_\gamma \cap \{1, \ldots, N_k\}|}{N_k}$$

exists. By the mean ergodic theorem

$$\begin{aligned}
\nu(A) &\leqslant \lim_{k\to\infty} \frac{1}{N_k} \sum_{n=1}^{N_k} \frac{\nu(A \cap S^{-n}A)}{\nu(A)} = \lim_{k\to\infty} \frac{1}{N_k} \sum_{n=1}^{N_k} 1_{\mathcal{R}}(n) \frac{\nu(A \cap S^{-n}A)}{\nu(A)} \\
&\leqslant \lim_{k\to\infty} \frac{1}{N_k} \left(\sum_{n \in \mathcal{R}_\gamma, n \leqslant N_k} 1_{\mathcal{R}}(n) + \sum_{n \notin \mathcal{R}_\gamma, n \leqslant N_k} 1_{\mathcal{R}}(n)(1-\gamma) \right) \\
&= d_{N_k}(\mathcal{R}_\gamma) + (1-\gamma)(d(\mathcal{R}) - d_{N_k}(\mathcal{R}_\gamma)).
\end{aligned}$$

Rearranging and using the assumption $d(\mathcal{R}) = \nu(A)$, it follows that

$$\nu(A) \leqslant d_{N_k}(\mathcal{R}_\gamma) \leqslant d(\mathcal{R}) = \nu(A).$$

Hence $d_{N_k}(\mathcal{R}_\gamma) = d(\mathcal{R})$. By Lemma 5.1 $\mathcal{R}_\gamma + \mathcal{R}_\gamma \subset \mathcal{R}_{2\gamma} \subset \mathcal{R}$, so

$$d(\mathcal{R}) = d_{N_k}(\mathcal{R}_\gamma) = \underline{d}_{N_k}(\mathcal{R}_\gamma) \leqslant \underline{d}_{N_k}(\mathcal{R}_\gamma + \mathcal{R}_\gamma) \leqslant \overline{d}_{N_k}(\mathcal{R}_\gamma + \mathcal{R}_\gamma) \leqslant d(\mathcal{R}).$$

Hence $d_{N_k}(\mathcal{R}_\gamma + \mathcal{R}_\gamma)$ exists and equals $d(\mathcal{R})$. Since $(N_k)_{k\geqslant 1}$ was arbitrary, the conclusion follows. □

For $0 < \gamma < \frac{1}{2}$, Lemma 5.2 and Kneser's theorem [Kne53] (see also [Bil97, Theorem 1.1]) therefore imply the existence of $k \geqslant 1$ and $K \subset \{0, \ldots, k-1\}$ such that

- $\mathcal{R}_\gamma \subset K + k\mathbb{N}$,
- $|K + K| = 2|K| - 1$, where the operation on the left hand side is in $\mathbb{Z}/k\mathbb{Z}$, and
- $\mathcal{R}_\gamma + \mathcal{R}_\gamma \subset K + K + k\mathbb{N}$ with $|(K + K + k\mathbb{N}) \setminus (\mathcal{R}_\gamma + \mathcal{R}_\gamma)| < \infty$.

It follows that $K = \{0\}$, so $\mathcal{R}_\gamma \subset k\mathbb{N}$ and $d(\mathcal{R}) = d(\mathcal{R}_\gamma) = d(\mathcal{R}_\gamma + \mathcal{R}_\gamma) = k^{-1}$. Further, for all $m \in \mathcal{R}$ there exists $\gamma > 0$ small enough such that $m + \mathcal{R}_\gamma \subset \mathcal{R}$ by Lemma 5.1. Since in addition $\mathcal{R}_\gamma \subset \mathcal{R}$ and $d(\mathcal{R}) = d(\mathcal{R}_\gamma)$, it follows that $m \in k\mathbb{N}$. Then the k sets $S^{-i}A$, $0 \leqslant i \leqslant k-1$ are disjoint (up to ν-null sets) and each of measure k^{-1}. □

Theorem 1.4. If $(X, \mathscr{B}, \nu, S)$ is ergodic and

$$0 < \underline{d}(\mathcal{R}) < \frac{3}{2}\nu(A),$$

then there exists an integer $k \geqslant 1$ such that $\mathcal{R} = k\mathbb{N}$ and $X = \bigsqcup_{i=0}^{k-1} S^{-i}\left(\bigcup_{j=0}^{\infty} S^{-jk}A\right)$ up to ν-null sets.

Proof. By [BFS19, Section 1.5] all the theorems in [BFS19] hold for ergodic $\mathbb{N}$-actions, so [BFS19, Theorem 1.3] gives the existence of $k \geqslant 1$ such that $\mathcal{R} = k\mathbb{N}$. Therefore, the sets

$$\bigcup_{j=0}^{\infty} S^{-jk}A, S^{-1}\left(\bigcup_{j=0}^{\infty} S^{-jk}A\right), \dots, S^{-(k-1)}\left(\bigcup_{j=0}^{\infty} S^{-jk}A\right)$$

are mutually disjoint up to ν-null sets, and ergodicity implies that they partition X. □

6. Inverse theorems for trees

In this section we prove Theorems B, C, and 1.9.

Theorem B. If T is a tree and $2 \leqslant r \leqslant q$ with

$$\bar{d}(\overline{G}(F^r)) = \frac{\overline{\dim}_M T - \log_q(r-1)}{1 - \log_q(r-1)} > 0,$$

then $\overline{\dim}_M T = k^{-1}(1-\log_q(r-1)) + \log_q(r-1)$ for some positive integer k. Moreover, $\overline{d_T}(V_1^{r,k,kn}) > 0$ for every $n \geqslant 1$.

Proof. Let $\mathcal{R} = \{n \in \mathbb{N} \colon \widetilde{\mu}(A_r \cap S^{-n}A_r) > 0\}$. Combining Theorem 1.2 and the hypothesis we obtain

$$\bar{d}(\overline{G}(F^r)) \geqslant \bar{d}(\mathcal{R}) \geqslant \underline{d}(\mathcal{R}) \geqslant \widetilde{\mu}(A_r) \geqslant \frac{\overline{\dim}_M T - \log_q(r-1)}{1 - \log_q(r-1)} = \bar{d}(\overline{G}(F^r)).$$

Therefore $\widetilde{\mu}(A_r) = \bar{d}(\overline{G}(F^r)) = \bar{d}(\mathcal{R}) = \underline{d}(\mathcal{R})$ is positive; by Theorem 1.2 it equals k^{-1} for some positive integer k, and $\widetilde{M} = \bigsqcup_{i=0}^{k-1} S^{-i}A_r$ up to $\widetilde{\mu}$-null sets.

The above also shows that equality holds in Proposition 2.5 for $\widetilde{\mu}$, whence $\int_{A_r} H \, d\widetilde{\mu} = \widetilde{\mu}(A_r)$ and $\int_{A_r^c} H \, d\widetilde{\mu} = (1 - \widetilde{\mu}(A_r)) \log_q(r-1)$. The bounds on H then imply $\widetilde{\mu}$-almost everywhere equalities

$$\prod_{i \in \Lambda} P 1_{B_i} = c_1 1_{A_r} = c_1 1_{A_q} \tag{9}$$

$$1_{A_{r-1}} \sum_{\substack{I \subset \Lambda \\ |I| = r-1}} \prod_{i \in I} P 1_{B_i} = 1_{A_r^c} \sum_{\substack{I \subset \Lambda \\ |I| = r-1}} \prod_{i \in I} P 1_{B_i} = c_2 1_{A_r^c}, \tag{10}$$

where $c_1 = q^{-q}$ and $c_2 = (r-1)^{1-r}$.

Recall from Section 3.1 the operators $R_{r-1,1}$ and $R_{q,1}$ on $L^\infty(\widetilde{M}, \widetilde{\mu})$, which for simplicity we denote by R_1 and R_2:

$$R_1 f = \sum_{\substack{I \subset \Lambda \\ |I| = r-1}} \prod_{i \in I} P(1_{B_i} f); \qquad R_2 f = \prod_{i \in \Lambda} P(1_{B_i} f).$$

Using the facts determined above we compute $\phi'_{V_1^{r,k,kn}} = (R_2 R_1^{k-1})^n 1_{A_q}$. In the following, equalities are only up to $\widetilde{\mu}$-null sets. We compute first the case $n = 1$. Note that $A_q = S^{-k}A_q$ and $1_{S^{-i}A_q} = S 1_{S^{-i+1}A_q}$ for $i \geqslant 1$. By Lemma 2.6

$$\begin{aligned}
R_1 1_{A_q} = \sum_{\substack{I \subset \Lambda \\ |I| = r-1}} \prod_{i \in I} P(1_{B_i} 1_{A_q}) &= \sum_{\substack{I \subset \Lambda \\ |I| = r-1}} \prod_{i \in I} P(1_{B_i} S 1_{S^{-k+1}A_q}) \\
&= 1_{S^{-k+1}A_q} \sum_{\substack{I \subset \Lambda \\ |I| = r-1}} \prod_{i \in I} P 1_{B_i} \\
&= c_2 1_{S^{-k+1}A_q}
\end{aligned}$$

where the last equality follows from (10) and the fact $S^{-k+1}A_q = S^{-k+1}A_r \subset A_r^c$. Since R_1 is homogeneous of degree $r-1$, repeating this calculation gives

$$R_1^{k-1} 1_{A_q} = c_2^{\sum_{j=0}^{k-2} (r-1)^j} 1_{S^{-1}A_q}$$

and hence

$$\begin{aligned}\phi'_{V_1^{r,k,k}} &= R_2\left(c_2^{\sum_{j=0}^{k-2}(r-1)^j} 1_{S^{-1}A_q}\right)\\ &= c_2^{q\sum_{j=0}^{k-2}(r-1)^j}\prod_{i\in\Lambda} P(1_{B_i}S1_{A_q})\\ &= c_2^{q\sum_{j=0}^{k-2}(r-1)^j} 1_{A_q}\prod_{i\in\Lambda} P1_{B_i}\\ &= c_1 c_2^{q\sum_{j=0}^{k-2}(r-1)^j} 1_{A_q}.\end{aligned}$$

Letting $d_1 = c_1 c_2^{q\sum_{j=0}^{k-2}(r-1)^j}$ and defining inductively $d_n = d_{n-1}^{q(r-1)^{k-1}} d_1$, it follows that $\phi'_{V_1^{r,k,kn}} = d_n 1_{A_q}$ $\widetilde{\mu}$-almost everywhere. Then by the correspondence principle Theorem 3.3

$$\overline{d_T}(V_1^{r,k,kn}) \geqslant \int_{\widetilde{M}} \phi_{V_1^{r,k,kn}}\, d\widetilde{\mu} \geqslant \int_{\widetilde{M}} \phi'_{V_1^{r,k,kn}}\, d\widetilde{\mu} = d_n\widetilde{\mu}(A_q) = \frac{d_n}{k} > 0$$

for all $n \geqslant 1$. □

Theorem C. If T is a tree and $2 \leqslant r \leqslant q$ with

$$\underline{d}(G^*(F^r)) = \frac{\dim T - \log_q(r-1)}{1-\log_q(r-1)} > 0 \tag{11}$$

or

$$\overline{d}(G^*(D^{r,2})) = \frac{\dim T - \log_q(r-1)}{1-\log_q(r-1)} > 0, \tag{12}$$

then $\dim T = k^{-1}(1-\log_q(r-1)) + \log_q(r-1)$ for some positive integer k. Moreover, $d_T^*(V_1^{r,k,kn}) > 0$ for every $n \geqslant 1$.

Proof. Fix $\epsilon > 0$ small enough, and let $\mathcal{R} = \{n \in \mathbb{N} \colon \widetilde{\eta}_\epsilon(A_r \cap S^{-n}A_r) > 0\}$. In the case of (11), from the proof of Theorem A we have

$$\frac{\dim T - \log_q(r-1)}{1-\log_q(r-1)} = \underline{d}(G^*(F^r)) \geqslant \underline{d}(\mathcal{R}) \geqslant \widetilde{\eta}_\epsilon(A_r) \geqslant \frac{\dim T - \epsilon - \log_q(r-1)}{1-\log_q(r-1)},$$

so $\underline{d}(\mathcal{R}) \leqslant \frac{3}{2}\widetilde{\eta}_\epsilon(A_r)$ for small enough ϵ. By Theorem 1.4 there is a positive integer k such that $\mathcal{R} = k\mathbb{N}$ and $\widetilde{M} = \bigsqcup_{i=0}^{k-1} S^{-i}\left(\bigcup_{j=0}^{\infty} S^{-kj}A_r\right)$ up to $\widetilde{\eta}_\epsilon$-null sets.

In the case of (12), we invoke the proof of Theorem 4.1. Recall that there exist a measurable set Z such that $A_r \subset Z$ modulo $\widetilde{\eta}_\epsilon$-null sets and an increasing chain of measurable sets $(Z_\rho)_{\rho>0}$ with $Z_\rho \subset Z$ and $\bigcup_{\rho>0} Z_\rho = Z$ such that for every $\delta > 0$ we have

$$\begin{aligned}\frac{\dim T - \log_q(r-1)}{1-\log_q(r-1)} &= \overline{d}(G^*(D^{r,2})) \geqslant \overline{d}(\mathcal{R}^\delta(Z_\rho))\\ &\geqslant \left(\liminf_{N\to\infty}\frac{1}{N}\sum_{m=1}^{N}\frac{\widetilde{\eta}_\epsilon(Z_\rho\cap S^{-m}Z_\rho)}{\widetilde{\eta}_\epsilon(Z_\rho)}\right) - \delta\widetilde{\eta}_\epsilon(Z_\rho)\\ &\geqslant (1-\delta)\widetilde{\eta}_\epsilon(Z_\rho) \xrightarrow{\delta\to0} \widetilde{\eta}_\epsilon(Z_\rho) \xrightarrow{\rho\to0} \widetilde{\eta}_\epsilon(Z)\\ &\geqslant \widetilde{\eta}_\epsilon(A_r) \geqslant \frac{\dim T - \epsilon - \log_q(r-1)}{1-\log_q(r-1)},\end{aligned}$$

where $\mathcal{R}^\delta(Z_\rho) = \{n \in \mathbb{N} \colon \widetilde{\eta}_\epsilon(Z_\rho \cap S^{-n}Z_\rho) > \delta\widetilde{\eta}_\epsilon(Z_\rho)^2\}$. Hence for small enough ϵ and ρ the assumptions of Theorem A.3 are satisfied, so there exists $k \geqslant 1$ such that $\mathcal{R}(Z_\rho) = \mathcal{R}^\delta(Z_\rho) = k\mathbb{N}$, where $\mathcal{R}(Z_\rho) = \{n \in \mathbb{N} \colon \widetilde{\eta}_\epsilon(Z_\rho \cap S^{-n}Z_\rho) > 0\}$. Since this is true for all $\rho > 0$ small enough and $\mathcal{R} \subset \bigcup_{\rho>0}\mathcal{R}(Z_\rho)$, we conclude that for ϵ small enough there exists $k \geqslant 1$ such that $\mathcal{R} \subset k\mathbb{N}$. This immediately implies that $\widetilde{M} = \bigsqcup_{i=0}^{k-1} S^{-i}\left(\bigcup_{j=0}^{\infty} S^{-kj}A_r\right)$ up to $\widetilde{\eta}_\epsilon$-null sets.

In both cases, for small ϵ the above inequalities force

$$\frac{\dim T - \log_q(r-1)}{1-\log_q(r-1)} = k^{-1},$$

and hence

$$\widetilde{\eta}_\epsilon(A_r) \geqslant (1-\epsilon')\frac{1}{k} = (1-\epsilon')\widetilde{\eta}_\epsilon\left(\bigcup_{j=0}^{\infty} S^{-kj}A_r\right), \tag{13}$$

where $\epsilon' \to 0$ as $\epsilon \to 0$. We also have

$$\dim T \geqslant \widetilde{\eta}_\epsilon(A_r) + (1-\widetilde{\eta}_\epsilon(A_r))\log_q(r-1) \geqslant H(\widetilde{\eta}_\epsilon) \geqslant \dim T - \epsilon$$

and hence the pair of inequalities

$$\int_{A_r} H\, d\widetilde{\eta}_\epsilon \geqslant \widetilde{\eta}_\epsilon(A_r) - \epsilon \tag{14}$$

$$\int_{A_r^c} H\, d\widetilde{\eta}_\epsilon \geqslant (1-\widetilde{\eta}_\epsilon(A_r))\log_q(r-1) - \epsilon. \tag{15}$$

We denote by $\mathbf{A_r} = \bigcup_{j=0}^{\infty} S^{-kj}A_r$. Then we have $\widetilde{M} = \bigsqcup_{i=0}^{k-1} S^{-i}\mathbf{A_r}$. Given $\widetilde{\tau} \in \widetilde{M}$ and $E \subset \widetilde{M}$, observe that $S^{-i}(\{\widetilde{\tau}\}) \subset E$ if and only if $P^i 1_E(\widetilde{\tau}) = 1$. For $i \geqslant 0$, define E_i to be $\mathbf{A_r}$ if k divides i and $\mathbf{A_r^c}$ otherwise. Then the k-periodicity of $\mathbf{A_r}$ and $\mathbf{A_r^c} = \bigsqcup_{i=1}^{k-1} S^{-i}\mathbf{A_r}$ under S^{-1} gives $\widetilde{\eta}_\epsilon$-almost everywhere equalities

$$P^i 1_{E_i} = P^i S^i 1_{S^{-i}E_i} = 1_{S^{-i}E_i},$$

so the set $\mathbf{A'_r} = \bigcap_{i\geqslant 0}\{\widetilde{\tau} \in \widetilde{M} \colon P^i 1_{E_i}(\widetilde{\tau}) = 1\}$ is a $\widetilde{\eta}_\epsilon$-conull subset of $\mathbf{A_r}$.

Define for $\delta > 0$ the set

$$A_\delta = \bigcap_{i=0}^{kn}\{\widetilde{\tau} \in \mathbf{A'_r} \colon P^i H(\widetilde{\tau}) \geqslant c_i - \delta\}, \qquad c_i = \begin{cases} 1 & k \mid i \\ \log_q(r-1) & \text{otherwise.}\end{cases}$$

It follows from (13) and inequalities (14) and (15) that by choosing ϵ small enough we can guarantee that $\widetilde{\eta}_\epsilon(A_\delta) > 0$. We will show the existence of δ such that the configuration $V_1^{r,k,kn}$ appears at the root of $|\tau_0|$ for every $\widetilde{\tau} = (\tau_i)_{i\leqslant 0} \in A_\delta$. First notice that by construction of $\mathbf{A'_r}$, if $\widetilde{\tau} \in A_\delta$ and $v \in |\tau_0|$ with $0 \leqslant l(v) \leqslant kn$ then $H(\widetilde{\tau}^v) \leqslant c_{l(v)}$. Hence for $0 \leqslant i \leqslant kn$

$$c_i - \delta \leqslant P^i H(\widetilde{\tau}) = \sum_{l(v)=i} \tau_0(v) H(\widetilde{\tau}^v) \leqslant c_i. \tag{16}$$

If $H(\widetilde{\tau}) > \log_q(q-1)$ then $\widetilde{\tau} \in A_q$, and if $H(\widetilde{\tau}) > \log_q(r-2)$ then $\widetilde{\tau} \in A_{r-1}$. To prove the appearance of $V_1^{r,k,kn}$ at the root of $|\tau_0|$ it therefore suffices to give sufficiently large lower bounds for $H(\widetilde{\tau}^v)$ for $l(v) \leqslant kn$.

Lemma 6.1. For every $\delta_1, \delta_2 > 0$ there exists $\delta > 0$ such that for $1 \leqslant j \leqslant kn+1$ (a) the set $\{\tau_0(v) \colon \widetilde{\tau} \in A_\delta, v \in |\tau_0|(j)\} \subset [0,1]$ is contained in an interval of length $< \delta_1$, and (b) for all $\widetilde{\tau} \in A_\delta$ and $v \in |\tau_0|$ with $l(v) \leqslant j-1$ we have $H(\widetilde{\tau}^v) \geqslant c_{l(v)} - \delta_2$.

Proof. We prove both statements simultaneously by induction on j. For $j = 1$ we have $H(\widetilde{\tau}) \geqslant 1-\delta$ for all $\widetilde{\tau} \in A_\delta$ by (16), so any $\delta < \delta_2$ suffices. Further, observe that H is a continuous function attaining its maximum at $\widetilde{\tau}$ such that $p_{\widetilde{\tau}}(B_i) = q^{-1}$ for all $i \in \Lambda$. Hence given $\delta_1 > 0$ the set $\{\tau_0(v) \colon \widetilde{\tau} \in A_\delta, v \in |\tau_0|(1)\}$ is contained in an interval of length $< \delta_1$ (containing q^{-1}) for δ small enough.

Assuming the lemma is true for $j \leqslant m < kn+1$, we prove it for $j = m+1$. We first consider (b). For $w \in |\tau_0|(m)$ with $\widetilde{\tau} \in A_\delta$ the inequality (16) gives

$$c_m - \delta \leqslant P^m H(\widetilde{\tau}) = \sum_{l(v)=m} \tau_0(v) H(\widetilde{\tau}^v) \leqslant \tau_0(w) H(\widetilde{\tau}^w) + (1-\tau_0(w))c_m,$$

and rearranging gives

$$c_m - \frac{\delta}{\tau_0(w)} \leqslant H(\widetilde{\tau}^w).$$

By statement (a) of the induction hypothesis

$$\sup_{\widetilde{\tau}\in A_\delta, w\in|\tau_0|(m)} \frac{\delta}{\tau_0(w)} \to 0$$

as $\delta \to 0$, so by taking δ small enough statement (b) is satisfied for $j = m+1$. Combining statement (a) for $j = m$ and statement (b) for $j = m+1$ with the same argument as in the base case proves statement (a), noting that if k does not divide j then we consider maxima of H on A_r^c. □

It follows that any $V_1^{r,k,kn}$-detecting function is positive on A_δ. By the correspondence principle Theorem 3.4 we have for all $\epsilon > 0$

$$d_T^*(V_1^{r,k,kn}) \geqslant \int_{\widetilde{M}} \phi_{V_1^{r,k,kn}}\, d\widetilde{\eta}_\epsilon \geqslant \int_{A_\delta} \phi_{V_1^{r,k,kn}}\, d\widetilde{\eta}_\epsilon > 0,$$

since $\widetilde{\eta}_\epsilon(A_\delta) > 0$. □

Theorem 1.9. Let $\beta < 3/2$ and assume that $0 < \underline{d}(G^*(F^r)) < \beta \cdot \frac{\dim T - \log_q(r-1)}{1-\log_q(r-1)}$. Then there exists an integer $k \geqslant 1$ such that $k\mathbb{N} \subset G^*(F^r)$.

Proof. For small enough $\epsilon > 0$, by the correspondence principle Theorem 3.4 and the proof of Theorem A

$$\beta\widetilde{\eta}_\epsilon(A_r) \geqslant \beta\frac{\dim T - \epsilon - \log_q(r-1)}{1-\log_q(r-1)} > \underline{d}(G^*(F^r)) \geqslant \underline{d}(\mathcal{R}),$$

where $\mathcal{R} = \{n \in \mathbb{N} \colon \widetilde{\eta}_\epsilon(A_r \cap S^{-n}A_r) > 0\}$. Theorem 1.4 then implies that there exists $k \geqslant 1$ such that $k\mathbb{N} = \mathcal{R} \subset G^*(F^r)$. □

Appendix A. Stability in inverse theorem 1.4

In the proof of Theorem C for the configuration $D^{r,2}$, we are unable to apply Theorem 1.4 since we have no upper bound for $\underline{d}(\mathcal{R})$. However, we have bounds on the densities of the sets of δ-return times. Here we prove a stability result (Theorem A.3) giving the same conclusion as Theorem 1.4 under assumptions involving δ-return times instead of return times.

Given an ergodic measure-preserving system $(X, \mathscr{B}, \nu, S)$ and $A \in \mathscr{B}$ with $\nu(A) > 0$, define for $\delta > 0$ the set of δ-return times of A

$$\mathcal{R}^\delta = \{n \in \mathbb{N} \colon \nu(A \cap S^{-n}A) > \delta\nu(A)^2\}.$$

Define also for $0 < \gamma < 1$ the set

$$\mathcal{R}_\gamma = \{n \in \mathbb{N} \colon \nu(A \cap S^{-n}A) \geqslant (1-\gamma)\nu(A)\}.$$

Lemma A.1. If $\overline{d}(\mathcal{R}^\delta) \leqslant (1+\eta)\nu(A)$ for all $\delta > 0$, then for any $\gamma > 0$

$$\underline{d}(\mathcal{R}_\gamma) \geqslant \left(\frac{\gamma - \eta + \gamma\eta}{\gamma}\right)\nu(A).$$

Proof. Given γ, choose δ such that $0 < \delta < \frac{1-\gamma}{\nu(A)}$ (so $\mathcal{R}_\gamma \subset \mathcal{R}^\delta$). First observe that by the mean ergodic theorem

$$\underline{d}(\mathcal{R}^\delta) = \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} 1_{\mathcal{R}^\delta}(n) \geqslant \liminf_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \frac{\nu(A\cap S^{-n}A)}{\nu(A)} - \delta\nu(A) = (1-\delta)\nu(A). \tag{17}$$

Let $(N_k)_{k\geqslant 1}$ be an increasing sequence such that $\underline{d}(\mathcal{R}_\gamma) = \lim_{k\to\infty} \frac{1}{N_k}\sum_{n=1}^{N_k} 1_{\mathcal{R}_\gamma}(n)$. By the mean ergodic theorem

$$\begin{aligned}
\nu(A) &= \lim_{k\to\infty} \frac{1}{N_k}\sum_{n=1}^{N_k} \frac{\nu(A\cap S^{-n}A)}{\nu(A)} \\
&\leqslant \limsup_{k\to\infty} \frac{1}{N_k}\sum_{\substack{n\leqslant N_k\\ n\in\mathcal{R}_\gamma}} \frac{\nu(A\cap S^{-n}A)}{\nu(A)} + \limsup_{k\to\infty}\frac{1}{N_k}\sum_{\substack{n\leqslant N_k\\ n\in\mathcal{R}^\delta\setminus\mathcal{R}_\gamma}}(1-\gamma) + \limsup_{k\to\infty}\frac{1}{N_k}\sum_{\substack{n\leqslant N_k\\ n\in(\mathcal{R}^\delta)^c}}\delta\nu(A) \\
&\leqslant \underline{d}(\mathcal{R}_\gamma) + (1-\gamma)\left(\overline{d}(\mathcal{R}^\delta) - \underline{d}(\mathcal{R}_\gamma)\right) + \delta\nu(A)(1-\underline{d}(\mathcal{R}^\delta)) \\
&\leqslant \gamma\underline{d}(\mathcal{R}_\gamma) + (1-\gamma)(1+\eta)\nu(A) + \delta\nu(A)(1-(1-\delta)\nu(A))
\end{aligned}$$

where in the last inequality we used the assumption $\overline{d}(\mathcal{R}^\delta)\leqslant(1+\eta)\nu(A)$ and (17). Rearranging, we obtain

$$\underline{d}(\mathcal{R}_\gamma) \geqslant \left(\frac{\gamma-\eta+\gamma\eta-\delta+\delta\nu(A)-\delta^2\nu(A)}{\gamma}\right)\nu(A).$$

Taking $\delta\to 0$ gives the required inequality. □

For $m\in\mathbb{N}$ and $\delta>0$, define the set

$$\mathcal{R}^\delta_m = \{n\in\mathbb{N}\colon \nu(A\cap S^{-n}A\cap S^{-(m+n)}A) > \delta\nu(A)^3\}.$$

Lemma A.2. If $m\in\mathcal{R}^\delta$, then $\underline{d}(\mathcal{R}^{\delta\varepsilon}_m)\geqslant(1-\varepsilon)\nu(A)$ for all $\varepsilon>0$.

Proof. Given $\varepsilon>0$ and $A,B\in\mathscr{B}$ of positive measure, the set of ε-transfer times from A to B is $\mathcal{R}^\varepsilon_{A,B}=\{n\in\mathbb{N}\colon \nu(A\cap S^{-n}B)>\varepsilon\nu(A)\nu(B)\}$. Observe that

$$\mathcal{R}^\varepsilon_{A,A\cap S^{-m}A} = \{n\in\mathbb{N}\colon \nu(A\cap S^{-n}(A\cap S^{-m}A)) > \varepsilon\nu(A)\nu(A\cap S^{-m}A)\}\subset\mathcal{R}^{\delta\varepsilon}_m.$$

By the mean ergodic theorem,

$$\begin{aligned}
\underline{d}(\mathcal{R}^\varepsilon_{A,B}) &= \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^N 1_{\mathcal{R}^\varepsilon_{A,B}}(n) \\
&\geqslant \liminf_{N\to\infty}\frac{1}{N}\sum_{n=1}^N \frac{\nu(A\cap S^{-n}B)}{\min(\nu(A),\nu(B))} - \varepsilon\max(\nu(A),\nu(B)) \\
&\geqslant (1-\varepsilon)\max(\nu(A),\nu(B)),
\end{aligned}$$

so we have

$$\underline{d}(\mathcal{R}^{\delta\varepsilon}_m)\geqslant\underline{d}(\mathcal{R}^\varepsilon_{A,A\cap S^{-m}A})\geqslant(1-\varepsilon)\nu(A)$$

as required. □

Theorem A.3. If for $\eta<\frac{1}{5}$ we have $\overline{d}(\mathcal{R}^\delta)\leqslant(1+\eta)\nu(A)$ for every $\delta>0$, then there exists $k\geqslant 1$ such that $\mathcal{R}^\delta = k\mathbb{N}$ for all sufficiently small δ.

Proof. Fix $0<\eta<\frac{1}{5}$ such that $\overline{d}(\mathcal{R}^\delta)\leqslant(1+\eta)\nu(A)$, and choose γ satisfying

$$\frac{3\eta}{1+\eta}<\gamma<\frac{1}{2}. \tag{18}$$

Observe that $\mathcal{R}_\gamma+\mathcal{R}_\gamma\subset\mathcal{R}_{2\gamma}\subset\mathcal{R}^\delta$ for $0<\delta<\frac{1-2\gamma}{\nu(A)}$ by Lemma 5.1. Noting that (18) implies $\gamma-\eta+\gamma\eta>0$ and $\frac{(1+\eta)\gamma}{\gamma-\eta+\gamma\eta}<2$, Lemma A.1 gives

$$\underline{d}(\mathcal{R}_\gamma+\mathcal{R}_\gamma)\leqslant\underline{d}(\mathcal{R}^\delta)\leqslant\overline{d}(\mathcal{R}^\delta)\leqslant(1+\eta)\nu(A)\leqslant\left(\frac{(1+\eta)\gamma}{\gamma-\eta+\gamma\eta}\right)\underline{d}(\mathcal{R}_\gamma)<2\underline{d}(\mathcal{R}_\gamma).$$

Kneser's theorem then gives the existence of an integer $k\geqslant 1$ and $K\subset\{0,1,\dots,k-1\}$ such that

- $\mathcal{R}_\gamma\subset K+k\mathbb{N}$,
- $|K+K|=2|K|-1$, where the operation on the left hand side is in $\mathbb{Z}/k\mathbb{Z}$, and
- $\mathcal{R}_\gamma+\mathcal{R}_\gamma\subset K+K+k\mathbb{N}$ with $|(K+K+k\mathbb{N})\setminus(\mathcal{R}_\gamma+\mathcal{R}_\gamma)|<\infty$.

Combining this with Lemma A.1 gives

$$\frac{2|K|-1}{k} = \frac{|K+K|}{k} \leqslant \overline{d}(\mathcal{R}_\gamma + \mathcal{R}_\gamma) \leqslant \overline{d}(\mathcal{R}^\delta) \leqslant (1+\eta)\nu(A)$$
$$\leqslant \left(\frac{(1+\eta)\gamma}{\gamma - \eta + \gamma\eta}\right) \underline{d}(\mathcal{R}_\gamma) \leqslant \left(\frac{(1+\eta)\gamma}{\gamma - \eta + \gamma\eta}\right) \frac{|K|}{k},$$

and rearranging gives

$$|K| \leqslant 1 + \frac{\eta}{\gamma + \gamma\eta - 2\eta} < 2,$$

where the last inequality follows from (18). Hence $|K| = 1$. Furthermore $K = \{0\}$, since otherwise $\mathcal{R}_\gamma$ and $\mathcal{R}_\gamma + \mathcal{R}_\gamma$ would be disjoint subsets of $\mathcal{R}^\delta$ giving the contradiction

$$\overline{d}(\mathcal{R}^\delta) \geqslant \underline{d}(\mathcal{R}_\gamma + \mathcal{R}_\gamma) + \underline{d}(\mathcal{R}_\gamma) \geqslant 2\underline{d}(\mathcal{R}_\gamma) > \overline{d}(\mathcal{R}^\delta).$$

We first prove that $\mathcal{R}^\delta \subset k\mathbb{N}$ for small enough $\delta > 0$. For $m \in \mathbb{N}$ and $\delta > 0$, recall

$$\mathcal{R}_m^{\delta^2} = \{n \in \mathbb{N} \colon \nu(A \cap S^{-n}A \cap S^{-(m+n)}A) > \delta^2\nu(A)^3\}.$$

Since $\nu(A \cap S^{-(m+n)}A) \geqslant \nu(A \cap S^{-n}A \cap S^{-(m+n)}A)$, if $n \in \mathcal{R}_m^{\delta^2}$ then $m + n \in \mathcal{R}^{\delta^2\nu(A)}$. Assuming $m \in \mathcal{R}^\delta \setminus k\mathbb{N}$, we derive a contradiction. Observe that $\mathcal{R}_\gamma + \mathcal{R}_\gamma \subset \mathcal{R}^\delta \subset \mathcal{R}^{\delta^2\nu(A)}$, so

$$\begin{aligned}\overline{d}(\mathcal{R}^{\delta^2\nu(A)}) &\geqslant \underline{d}(\mathcal{R}_\gamma + \mathcal{R}_\gamma) + \underline{d}((m + \mathcal{R}_m^{\delta^2}) \setminus k\mathbb{N}) \\ &\geqslant k^{-1} + \underline{d}(m + (\mathcal{R}_m^{\delta^2} \cap k\mathbb{N})) \\ &= k^{-1} + \underline{d}(\mathcal{R}_m^{\delta^2} \cap k\mathbb{N}),\end{aligned}$$

where the second inequality uses the assumption on m. Since $\mathcal{R}_m^{\delta^2}, k\mathbb{N} \subset \mathcal{R}^{\delta^2\nu(A)}$ (up to a finite set), by Lemma A.2

$$\begin{aligned}\underline{d}(\mathcal{R}_m^{\delta^2} \cap k\mathbb{N}) &\geqslant \underline{d}(\mathcal{R}_m^{\delta^2}) + \underline{d}(k\mathbb{N}) - \overline{d}(\mathcal{R}_m^{\delta^2} \cup k\mathbb{N}) \\ &\geqslant (1-\delta)\nu(A) + k^{-1} - \overline{d}(\mathcal{R}^{\delta^2\nu(A)}).\end{aligned}$$

Using the hypothesis $\overline{d}(\mathcal{R}^{\delta^2\nu(A)}) \leqslant (1+\eta)\nu(A)$ we obtain

$$2(1+\eta)\nu(A) \geqslant 2\overline{d}(\mathcal{R}^{\delta^2\nu(A)}) \geqslant (1-\delta)\nu(A) + 2k^{-1}. \tag{19}$$

Since $|K| = 1$, Kneser's theorem and Lemma A.1 imply

$$k^{-1} \geqslant \underline{d}(\mathcal{R}_\gamma) \geqslant \left(\frac{\gamma - \eta + \gamma\eta}{\gamma}\right)\nu(A),$$

and combining with (19) gives $\gamma \leqslant \frac{2\eta}{1-\delta}$ after rearranging. This is compatible with (18) only if $\eta > \frac{1-3\delta}{2}$. Since $\eta < \frac{1}{5}$, it follows that the above requires $\delta > \frac{1}{5}$. Hence $m \in \mathcal{R}^\delta \setminus k\mathbb{N}$ gives a contradiction and $\mathcal{R}^\delta \subset k\mathbb{N}$ for small enough $\delta > 0$.

Finally we show $\mathcal{R}^\delta = k\mathbb{N}$ for small δ. Indeed, since $\underline{d}(\mathcal{R}_\gamma) > \frac{1}{2k}$ by Lemma A.1 and (18), for every $m \in \mathbb{N}$ there exists $n \in \mathbb{N}$ such that $kn, k(n+m) \in \mathcal{R}_\gamma$. Therefore

$$\begin{aligned}\nu(A \cap S^{-km}A) &= \nu(S^{-kn}A \cap S^{-k(n+m)}A) \\ &\geqslant \nu((A \cap S^{-kn}A) \cap (A \cap S^{-k(n+m)}A)) \\ &\geqslant \nu(A \cap S^{-kn}A) + \nu(A \cap S^{-k(n+m)}A) - \nu(A) \\ &\geqslant (1-2\gamma)\nu(A) > \delta\nu(A)^2\end{aligned}$$

for $\delta < \frac{1-2\gamma}{\nu(A)}$, so for sufficiently small $\delta > 0$ we have $km \in \mathcal{R}^\delta$ for all $m \in \mathbb{N}$. $\square$

Discussion. The set of transfer times $\mathcal{R}_{A,B}$ has strong parallels with the difference set $\mathcal{A}-\mathcal{B}=\{a-b\colon a\in\mathcal{A},\, b\in\mathcal{B}\}$, $\mathcal{A},\mathcal{B}\subset\mathbb{Z}/r\mathbb{Z}$, which is one of the main objects of Additive Combinatorics. For example, the lower bound for $\underline{d}(\mathcal{R}^{\varepsilon}_{A,B})$ in Lemma A.2 corresponds to the simple fact that $|\mathcal{A}-\mathcal{B}|\geqslant\max\{|\mathcal{A}|,|\mathcal{B}|\}$. It is easy to see that the bound is tight and is attained when $\mathcal{B}-\mathcal{B}$ belongs to the centraliser of $\mathcal{A}$ (or vice versa). It implies that $\mathcal{A}$ and $\mathcal{B}$ have some periodic structure and it is analogous to our conclusions in Theorems 1.4 and A.3 on the structure of our dynamical system. On the other hand, if $\mathcal{A}=\{0,1\}\subseteq\mathbb{Z}/r\mathbb{Z}$ for large r, then $\mathcal{A}-\mathcal{A}=\{0,1,-1\}$ and hence η in Theorem A.3 must be less than $1/2$. Moreover, the sets $\mathcal{R}^{\delta}_{m}$ from Lemma A.2 which are used in the proof of Theorem A.3 can be thought as a dynamical version of the well–known combinatorial e–transform, see, e.g., [TV06, Section 5.1]. Although it is non–obvious how to define the higher sumsets in the dynamical context, an analogue of the Plünnecke–Rusza triangle inequality for dynamical systems would be a first step towards such a theory.

Question A.4. Assume that $(X,\mathscr{B},\nu,S)$ is an invertible ergodic system and $d(\mathcal{R}_{A,B})$, $d(\mathcal{R}_{A,C})$, $d(\mathcal{R}_{B,C})$ exist for $A,B,C\in\mathscr{B}$. Is it true that

$$\mu(C)d(\mathcal{R}_{A,B})\leqslant d(\mathcal{R}_{A,C})d(\mathcal{R}_{B,C})\,?$$

References

[BFS19] M. Björklund, A. Fish, and I. D. Shkredov, *Sets of transfer times with small densities*, Preprint, 2019, arXiv:1912.08999.

[Bil97] Y. Bilu, *Addition of sets of integers of positive density*, J. Number Theory **64** (1997), no. 2, 233–275.

[EFHN15] T. Eisner, B. Farkas, M. Haase, and R. Nagel, *Operator theoretic aspects of ergodic theory*, Graduate Texts in Mathematics, vol. 272, Springer, Cham, 2015.

[EW11] M. Einsiedler and T. Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London, Ltd., London, 2011.

[Fur70] H. Furstenberg, *Intersections of Cantor sets and transversality of semigroups*, Problems in analysis (Sympos. Salomon Bochner, Princeton Univ., Princeton, N.J., 1969), 1970, pp. 41–59.

[Fur77] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.

[Fur81] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, N.J., 1981, M. B. Porter Lectures.

[FW03] H. Furstenberg and B. Weiss, *Markov processes and Ramsey theory for trees*, Combin. Probab. Comput. **12** (2003), no. 5-6, 547–563.

[Kne53] M. Kneser, *Abschätzung der asymptotischen Dichte von Summenmengen*, Math. Z. **58** (1953), 459–484.

[PST12] J. Pach, J. Solymosi, and G. Tardos, *Remarks on a Ramsey theory for trees*, Combinatorica **32** (2012), no. 4, 473–482.

[TV06] T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006.

(Alexander Fish) School of Mathematics and Statistics F07, University of Sydney, NSW 2006, Australia
E-mail address: alexander.fish@sydney.edu.au

(Leo Jiang) Department of Mathematics, University of Toronto, Toronto, ON M5S 2E4, Canada
E-mail address: ljiang@math.toronto.edu

(Ilya D. Shkredov) Steklov Mathematical Institute, ul. Gubkina, 8, Moscow, Russia, 119991, and IITP RAS, Bolshoy Karetny per. 19, Moscow, Russia, 127994, and MIPT, Institutskii per. 9, Dolgoprudnii, Russia, 141701
E-mail address: ilya.shkredov@gmail.com