Direct and inverse results for popular differences in trees of positive dimension

ALEXANDER FISH† and LEO JIANG‡

With an Appendix

Stability in an inverse theorem for return times of ergodic systems

by ALEXANDER FISH, LEO JIANG and ILYA D. SHKREDOV§¶‖

† *School of Mathematics and Statistics F07, University of Sydney, NSW 2006, Australia*
(e-mail: alexander.fish@sydney.edu.au)
‡ *Department of Mathematics, University of Toronto, Toronto, ON, M5S 2E4, Canada*
(e-mail: ljiang@math.toronto.edu)
§ *Steklov Mathematical Institute, ul. Gubkina, 8, Moscow 119991, Russia*
¶ *IITP RAS, Bolshoy Karetny per. 19, Moscow 127994, Russia*
‖ *London Institute for Mathematical Sciences, 21 Albemarle St., London, UK*
(e-mail: ilya.shkredov@gmail.com)

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 formalize 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 $E \subset \mathbb{N}$. Recall that the *upper density* and *upper Banach density* of E are

$$\overline{d}(E) = \limsup_{N \to \infty} \frac{|E \cap \{0, \ldots, N\}|}{N + 1}, \quad d^*(E) = \limsup_{N - M \to \infty} \frac{|E \cap \{M, \ldots, N\}|}{N - M + 1}.$$

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

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

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 integers $k \geqslant 1$ and $0 = n_1, \ldots, n_k \in \mathbb{N}$,

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

Taking $k = 2$, 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

$$\underline{d}(\overline{\Delta}_0(E)) \geqslant \underline{d}(\mathcal{R}) \geqslant \lim_{N \to \infty} \frac{1}{N + 1} \sum_{n=0}^{N} \frac{\nu(A \cap S^{-n} A)}{\nu(A)} \geqslant \nu(A) = \overline{d}(E), \qquad (1)$$

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

$$\underline{d}(E) = \liminf_{N \to \infty} \frac{|E \cap \{0, \ldots, N\}|}{N + 1}.$$

If in the above the upper density is replaced by the upper Banach density, then ν can further be chosen to be ergodic [**Fur81**, Proposition 3.9] (see [**BHK05**, Proposition 3.1] for an explicit proof).

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 analyze 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 [**BFS21**] and a stability-type extension proved jointly with Shkredov in Appendix A, whereas in the general case we prove a slightly weaker statement (Theorem 5.1). 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.* To describe our results, we first summarize the necessary definitions (see §2 for precise formulations). For clarity of exposition, in this introduction we restrict our attention to the case $r = 2$ of our results and make corresponding simplifications to the notation.

Fix an integer $q \geqslant 2$. In this paper, a tree can be visualized 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 non-root vertex has exactly one incoming edge. (Technically, we work with the vertices of the graph with the partial order induced by directed paths.) The 'size' of T can be quantified by its upper Minkowski and Hausdorff dimensions $\overline{\dim}_M T$ and $\dim T$, which are defined by an identification of such trees with closed subsets of $[0, 1]$.

1.1.1. *Tree analogues of popular difference sets.* A k-term arithmetic progression (k-AP) in $E \subset \mathbb{N}$ can be viewed as an affine map $\{0, \ldots, k-1\} \to E$. We consider 'affine' maps satisfying certain branching conditions from configurations C ('finite trees') to trees T. If there exists such a map with 'common difference' n taking the root of the configuration to $v \in T$, we say that $v \in C_n = C_n(T)$. The set C_n corresponds to the set $E \cap (E - n) \cap \cdots \cap (E - (k-1)n)$ for k-APs in $E \subset \mathbb{N}$. Using extensions of upper density and upper Banach density to subsets of trees, we define sets of 'generic parameters'

$$\overline{G}(C) = \{n \in \mathbb{N} : \overline{d}(C_n) > 0\}, \quad G^*(C) = \{n \in \mathbb{N} : d^*(C_n) > 0\}.$$

We also introduce certain configurations F and D which are analogues of 2-APs, and their generic parameters can be interpreted as popular differences for trees. In particular, our first result is a version of (1).

THEOREM A. (= Theorems 4.1 and 4.2 for $r = 2$) *For any tree T we have*

$$\underline{d}(\overline{G}(F)) \geqslant \underline{d}(\overline{G}(D)) \geqslant \overline{\dim}_M T \quad \text{and} \quad \underline{d}(G^*(F)) \geqslant \underline{d}(G^*(D)) \geqslant \dim T.$$

1.1.2. *Inverse theorems for sets of return times.* Given the direct result Theorem A, we are interested in characterizing 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 rigid 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 B. (= Theorem 5.1) *If $\overline{d}(\mathcal{R}) = \nu(A) > 0$, then there exists an integer $m \geqslant 1$ such that up to ν-null sets*

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

Question 1.1. Does the assumption $\underline{d}(\mathcal{R}) = \nu(A)$ suffice to prove the conclusion of Theorem B?

If ν is ergodic, then Question 1.1 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 [**BFS21**].

THEOREM 1.2. (= Theorem 5.4) *If $(X, \mathcal{B}, \nu, S)$ is ergodic and*

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

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

Remark 1.3. Example 1.2 in [**BFS21**] shows that for every $\beta > 1$, there exists a non-ergodic measure-preserving system $(X, \mathcal{B}, \nu, S)$ and $A \in \mathcal{B}$ of arbitrarily small measure such that $\overline{d}(\mathcal{R}) \leqslant \beta \nu(A)$ and there is no $m \geqslant 1$ such that $\mathcal{R} = m\mathbb{N}$.

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

PROPOSITION 1.4. *Assume that $E \subset \mathbb{N}$ satisfies $\overline{d}(\overline{\Delta}_0(E)) = \overline{d}(E) > 0$. Then there exists $m \geqslant 1$ such that $m\mathbb{N} \subset \overline{\Delta}_0(E)$ and $\overline{d}(\overline{\Delta}_0(E)) = \overline{d}(E) = m^{-1}$. Moreover, for every $k \geqslant 2$*

$$\overline{d}(E \cap (E - m) \cap \cdots \cap (E - (k-1)m)) = \overline{d}(E).$$

If we consider $\Delta_0^*(E)$ and $d^*(E)$ in place of $\overline{\Delta}_0(E)$ and $\overline{d}(E)$, we can apply Theorem 1.2 to obtain the following inverse result.

PROPOSITION 1.5. *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 $m \geqslant 1$ such that $m\mathbb{N} \subset \Delta_0^*(E)$. Moreover, for every $k \geqslant 2$ that satisfies $(1 - \beta^{-1})k < 1$, we have

$$d^*(E \cap (E - m) \cap \cdots \cap (E - (k-1)m)) > 0.$$

1.1.4. *Inverse results for $\overline{G}(F)$ and $G^*(F)$.* Propositions 1.4 and 1.5 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 $m\mathbb{N}$. In the tree setting, we prove analogous results.

For every $m \geqslant 1$, define $T_{m\mathbb{N}}$ to be the tree such that $v \in T_{m\mathbb{N}}$ has q outgoing edges if the directed path from the root to v has length a multiple of m and one outgoing edge otherwise. The inequalities in Theorem A are equalities for $T_{m\mathbb{N}}$ (see §2.1.1).

For every $k \geqslant 1$, define the configuration $V^{m,k}$ to be the first k levels of $T_{m\mathbb{N}}$. The following two theorems are analogues of Propositions 1.4 and 1.5, respectively.

THEOREM C. (= Theorem 6.1 for $r = 2$) *Let T be a tree. Assume that*

$$\overline{d}(\overline{G}(F)) = \overline{\dim}_M T > 0.$$

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

THEOREM D. (= Theorem 6.2 for $r = 2$) *Let T be a tree. Assume that*

$$\underline{d}(G^*(F)) = \dim T > 0 \quad or \quad \overline{d}(G^*(D)) = \dim T > 0.$$

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

Remark 1.6. We show in §2.1.2 that Theorem D cannot be improved. Indeed, for every $\varepsilon > 0$ there exists a tree T_ε such that

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

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

Our final result is another partial analogue of Proposition 1.5.

THEOREM 1.7. (= Theorem 6.4 for $r = 2$) *Let T be a tree. Assume that there exists $\beta < 3/2$ such that*

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

Then there exists $m \geqslant 1$ with $m\mathbb{N} \subset G^*(F)$.

1.2. *Organization of the paper.* After describing the combinatorial and dynamical background (§2) and establishing Furstenberg–Weiss correspondence principles (§3), in §4 we prove lower bounds for the densities of popular differences for trees. We then use inverse theorems for sets of return times in measure-preserving systems (§5 and Appendix A) to prove inverse theorems for these lower bounds (§6).

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 $\mathbb{N} = \{0, 1, \ldots\}$.

2.1. *Combinatorial setup.* Let $\Lambda^* = \bigcup_{n=0}^\infty \Lambda^n$ be the set of finite words over Λ, where Λ^0 is the singleton comprising the empty word $\emptyset$. 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 non-empty 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 $\emptyset$ (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 realizations 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

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

This motivates several definitions we give in the following.

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

$$\dim T = \inf \left\{ \lambda > 0 : \liminf_{N \to \infty} \inf_{\substack{l(\Pi) = N \\ \Pi \text{ section of } T}} \sum_{v \in \Pi} q^{-\lambda l(v)} < 1 \right\}.$$

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

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

A straightforward calculation shows that

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

If E is a 'periodic' set (such as $m\mathbb{N}$), 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 $\tau : \Lambda^* \to [0, 1]$ 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*, because 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 $\mathcal{P}(\Lambda^{\mathbb{N}})$, because 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*

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

and its *upper Banach density*

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

Remark 2.3. These definitions specialize to their integer counterparts in the degenerate case $q = 1$, justifying the notation. The inequality $d^*(V) \geqslant \overline{d}(V)$ also holds for our more general definition. To see this, observe that it is enough to construct Markov trees π_N supported on T such that

$$\sum_{l(w) \leqslant N} \pi_N(w) 1_V(w) = \sum_{v \in T(N)} \frac{|V \cap \{w \in T : w \leqslant v\}|}{|T(N)|}$$

$$= \sum_{l(w) \leqslant N} \frac{|\{v \in T(N) : w \leqslant v\}|}{|T(N)|} 1_V(w)$$

(the last equality follows from reindexing the sum). However, this formula defines such a Markov tree on vertices w with $l(w) \leqslant N$, and we can choose π_N to be any consistent extension to the remaining vertices (cf. the proof of Theorem 3.4).

Example 2.4. If $V = V(E) = \{v \in T : l(v) \in E\}$ for $E \subset \mathbb{N}$ and T a tree, then $\overline{d}(V) = \overline{d}(E)$ and $d^*(V) = d^*(E)$. Both equalities follow directly from the definitions. For example, for the second equality we observe that for any τ with $|\tau| \subset T$ and any $v \in |\tau|$ we have

$$\frac{1}{N+1} \sum_{l(w) \leqslant N} \frac{\tau(vw)}{\tau(v)} 1_V(vw) = \frac{|E \cap \{l(v), \ldots, l(v) + N\}|}{N+1}.$$

We use the term *configuration* to refer to a non-empty finite subset $C \subset \Lambda^*$ closed under predecessors (a finite tree). Terminology and notation defined previously for trees are used for configurations as appropriate without comment. A configuration C is *non-branching* 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 $n \in \mathbb{N}$ we say $v \in C_n = C_n(T)$ if there exists a map $\iota \colon C \to T$ such that:

- $\iota(\emptyset) = v$;
- $\iota(w_1) \leqslant \iota(w_2)$ if $w_1 \leqslant w_2$ (ι is a map of posets);
- if w is the longest initial subword common to w_1 and w_2, then $\iota(w)$ is the longest initial subword common to $\iota(w_1)$ and $\iota(w_2)$ (ι is infimum-preserving);
- $l(\iota(w)) = l(v) + nl(w)$ for all $w \in C$ (ι is 'affine').

Equivalently, we say the configuration C appears at v (with parameter n). Observe that trivially every configuration appears at every vertex with parameter 0.
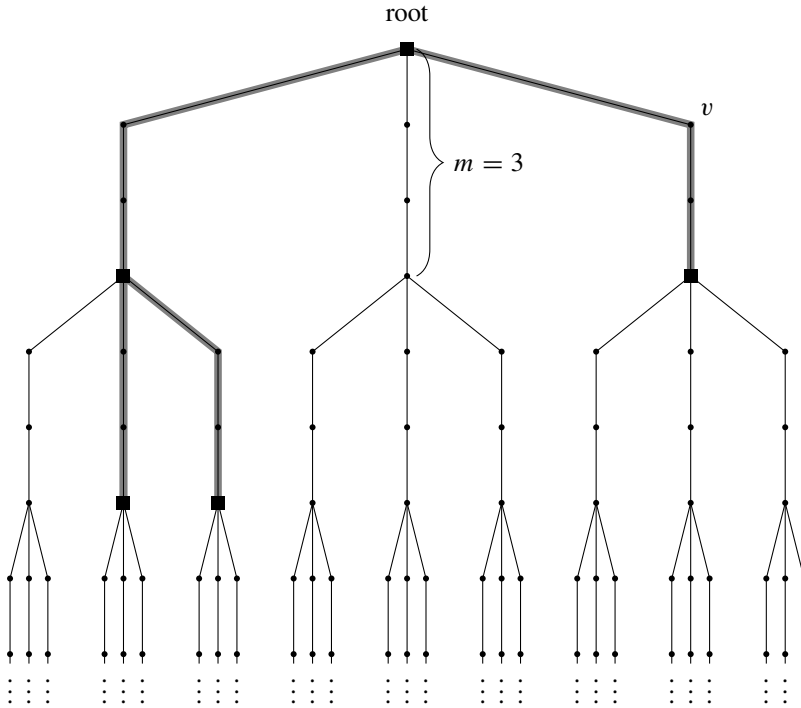
FIGURE 1. The configuration F^2 appears at the root of $T_{3\mathbb{N}}^2$ with parameter $n = 3$, whereas $v \notin F_n^2$ for any $n \geqslant 1$.

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

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

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

$$\overline{G}(C) = \overline{G}(C, T) = \{n \in \mathbb{N} \colon \overline{d}(C_n) > 0\},$$

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

Remark 2.5. Note that F^r appears at $v \in T_E^r$ with parameter n if and only if $D^{r,2}$ appears at v with parameter n if and only if $l(v), l(v) + n \in E$. Therefore, $\overline{G}(F^r, T_E^r) = \overline{G}(D^{r,2}, T_E^r) = \overline{\Delta_0}(E)$ and $G^*(F^r, T_E^r) = G^*(D^{r,2}, T_E^r) = \Delta_0^*(E)$ by Example 2.4. This is why the generic parameters of F^r and $D^{r,2}$ are analogues of popular differences for trees.

2.1.1. *Equality in Theorems 4.1 and 4.2.* The tree $T_{m\mathbb{N}}^r$ achieves equality in Theorems 4.1 and 4.2. Indeed, by Example 2.2,

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

The self-similarity of $T^r_{m\mathbb{N}}$ implies that $\dim T^r_{m\mathbb{N}} = \overline{\dim}_M T^r_{m\mathbb{N}}$. In addition, by Remark 2.5 it follows that

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

Hence,

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

and

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

2.1.2. *Sharpness of Theorem 6.2.* Next, we modify the construction of $T^r_{m\mathbb{N}}$ to obtain for every $\varepsilon > 0$ a tree T_ε with

$$0 < \underline{d}(G^*(F^r, T_\varepsilon)) = \overline{d}(G^*(D^{r,2}, T_\varepsilon)) < (1+\varepsilon)\frac{\dim T_\varepsilon - \log_q(r-1)}{1 - \log_q(r-1)}$$

such that there exists $k \geqslant 1$ with $V^{r,m,k}_1$ not appearing in T_ε.

Let $T_\varepsilon = T^r_E$, where $E = m\mathbb{N} \setminus mM\mathbb{N}$ for some positive integer $M > 1 + \varepsilon^{-1}$. Then

$$\overline{d}(E) = \frac{1}{m}\left(1 - \frac{1}{M}\right)$$

and $V^{r,m,mM+1}_1$ does not appear in T_ε. By the self-similarity of T_ε and Example 2.2, we have

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

and, hence,

$$\frac{\dim T_\varepsilon - \log_q(r-1)}{1 - \log_q(r-1)} = \frac{1}{m}\left(1 - \frac{1}{M}\right).$$

As $\Delta^*_0(E) = m\mathbb{N}$, observe that by Remark 2.5 we have $\underline{d}(G^*(F^r, T_\varepsilon)) = \overline{d}(G^*(D^{r,2}, T_\varepsilon)) = m^{-1}$. Therefore,

$$\underline{d}(G^*(F^r, T_\varepsilon)) = \overline{d}(G^*(D^{r,2}, T_\varepsilon)) = \frac{1}{1 - M^{-1}} \cdot \frac{\dim T_\varepsilon - \log_q(r-1)}{1 - \log_q(r-1)}$$

$$< (1+\varepsilon)\frac{\dim T_\varepsilon - \log_q(r-1)}{1 - \log_q(r-1)}.$$

2.2. *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 $p: M \to \mathcal{P}(M)$ on the space $M = \Lambda \times \mathcal{P}(\Lambda^{\mathbb{N}})$ by $p(a, \tau) = \sum_{i \in \Lambda} \tau(i)\delta_{(i,\tau^i)} \in \mathcal{P}(M)$. Here $a \in \Lambda$ can be interpreted as labelling the root of $\tau \in \mathcal{P}(\Lambda^{\mathbb{N}})$ with information about the past under the dynamics $\tau \mapsto \tau^a$. As p is continuous, it induces a Markov

operator P on $C(M)$ (a positive contraction satisfying $P1 = 1$) defined by the formula $Pf(a, \tau) = \sum_{i \in \Lambda} \tau(i) f(i, \tau^i)$. The pair (M, p) is a *CP-process*.

Remark 2.6. For simplicity of notation, frequently we denote a labelled Markov tree by its underlying Markov tree. Similarly, we write $p_\tau = p(a, \tau)$ because the latter is independent of a. Further, a labelled Markov tree denoted by τ^a is always assumed to have label a.

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. The set of stationary distributions for (M, p) is a non-empty compact convex subset of $\mathcal{P}(M)$, and its extremal points are called *ergodic* distributions.

For $i \in \Lambda$, define the set $B_i = \{(a, \tau) \in M : a = i\}$ 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 $A_r = \{\tau \in M : |\{i : p_\tau(B_i) > 0\}| \geqslant r\}$ 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 $0 \log_q 0 = 0$ by convention. The *entropy* of a stationary distribution ν is then $H(\nu) = \int_M H \, d\nu$. Note that $0 \leqslant H(\tau) \leqslant \log_q \||\tau|(1)|$.

PROPOSITION 2.7. *If ν is a stationary distribution 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.   □

2.3. *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 $p : M \to \mathcal{P}(M)$ to a continuous function $\widetilde{M} \to \mathcal{P}(\widetilde{M})$. Explicitly, $p_{\widetilde{\tau}} = \sum_{a \in \Lambda} \tau_0(a) \delta_{\widetilde{\tau}^a}$, where $(\widetilde{\tau}^a)_i = \tau_{i+1}$ for $i < 0$ and $(\widetilde{\tau}^a)_0 = \tau_0^a$. 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 $\widetilde{\nu}$ is invariant under the right shift $S : (\tau_i)_{i \leqslant 0} \mapsto (\tau_{i-1})_{i \leqslant 0}$ by construction [**Ho14**, Definition 6.3, Remark 6.4, and Lemma 6.8]. 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 $\widetilde{M}$. As $p_{\widetilde{\tau}}(\{\widetilde{\omega}\}) > 0$ implies $S(\widetilde{\omega}) = \widetilde{\tau}$, a straightforward calculation gives the following result.

LEMMA 2.8. *For any $f, g \in \mathscr{H}$ we have $P(f Sg) = gPf$.*

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 $S^n P^n$ 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 $\mathscr{H}_\infty = \bigcap_{n \geqslant 1} S^n \mathscr{H} = L^2(\widetilde{M}, \mathscr{B}_\infty, \nu)$, where $\mathscr{B}_\infty = \bigcap_{n \geqslant 1} S^{-n}\mathscr{B}$. For $f \in \mathscr{H}_\infty$ we have $Sf \in \mathscr{H}_\infty$ and $Pf \in \mathscr{H}_\infty$ (using $PS = I$), giving the following result.

LEMMA 2.9. *The operators 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 2.10. *For $f \in \mathscr{H}$, $\|P^n f - P^n \overline{f}\|_2 \to 0$.*

Proof. As $\widetilde{\nu}$ is S-invariant, it follows from Lemma 2.9 that

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

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

By composing H with the projection $\widetilde{M} \to M$ onto the zeroth coordinate, the information function H is defined on $\widetilde{M}$ 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 $D_n^{2,k}$ 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 $n \geqslant 1$, we say that a function $f : M \to [0, 1]$ is C_n-*detecting* if $f(\tau) > 0$ if and only if C appears at the root of $|\tau|$ with parameter n. 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_n}. For the simplest configuration $\{\emptyset\}$, we can take $\varphi_{\{\emptyset\}_n} = 1$ for all $n \geqslant 1$. Given $I \subset \Lambda$ such that $|I| = |C(1)|$ and a bijection $\beta : I \to C(1)$, the positivity of $\prod_{i \in I} P(1_{B_i} P^{n-1} \varphi_{C_n^{\beta(i)}})$ at $\tau \in M$ is equivalent to the appearance of C at the root of $|\tau|$ with parameter n such that $\beta(i) \in C(1)$

is mapped to $iv \in |\tau|$ for some $v \in \Lambda^{n-1}$. Summing over all choices of I and β, we define φ_{C_n} by the recursive formula

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

Remark 3.1. Alternatively we could sum over all injections $\gamma : C(1) \to \Lambda$ and define φ_{C_n} by

$$\varphi_{C_n} = \sum_{\gamma} \prod_{i \in C(1)} P(1_{B_{\gamma(i)}} P^{n-1} \varphi_{C_n^i}).$$

We also have $0 \leqslant \varphi_{C_n} \leqslant 1$. Indeed, because $\varphi_{\{\emptyset\}_n} = 1$ and P is positive

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

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

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

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

As the B_i are clopen and P takes continuous functions to continuous functions, the φ_{C_n} are continuous. However, in general the ϕ_{C_n} are not continuous because 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 $n \geqslant 1$, define (nonlinear) operators $R_{r,n}$ on $L^\infty(M)$ by

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

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

Example 3.2. For the configuration F^r, we have $|C(1)| = r$ and $|C(1)'| = 1$. There is always a unique bijection $I \to C(1)'$, so linearity of P gives

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

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

If $C(1) = C(1)'$, the factor $1_{A_{|C(1)|}}$ is redundant in the definition of ϕ_{C_n} as the function in the sum is already supported on a subset of $A_{|C(1)|}$. For example,

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

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

LEMMA 3.3. *If $(v_k)_{k \geqslant 1}$ is a sequence of distributions on M converging to v in the weak-* topology, then for every configuration C and integer $n \geqslant 1$*

$$\limsup_{k \to \infty} \int_M \phi_{C_n} \, dv_k \geqslant \int_M \phi_{C_n} \, dv.$$

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 $\phi_{C_n}^\delta$ be the function obtained by replacing 1_{A_r} with $1_{A_r^\delta}$ in the recursive construction of ϕ_{C_n}. Observe that $\delta \leqslant \delta'$ implies $\phi_{C_n}^\delta \geqslant \phi_{C_n}^{\delta'}$ by the positivity of P. Then the monotone function $\alpha \colon \delta \mapsto \int_M \phi_{C_n}^\delta \, dv$ has countably many discontinuities, so we can choose a sequence $\delta_j \to 0$ such that α is continuous at δ_j for all j.

We claim $\int_M \phi_{C_n}^\delta \, dv_k \to \int_M \phi_{C_n}^\delta \, dv = \alpha(\delta)$ 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_n} to be the function obtained by repeating the construction of ϕ_{C_n} with h_r in place of 1_{A_r}, it follows that $\phi_{C_n}^{\delta'} \leqslant h_{C_n} \leqslant \phi_{C_n}^\delta$. As h_{C_n} is continuous,

$$\liminf_{k \to \infty} \int_M \phi_{C_n}^\delta \, dv_k \geqslant \liminf_{k \to \infty} \int_M h_{C_n} \, dv_k = \int_M h_{C_n} \, dv \geqslant \int_M \phi_{C_n}^{\delta'} \, dv = \alpha(\delta').$$

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

$$\limsup_{k \to \infty} \int_M \phi_{C_n} \, dv_k \geqslant \lim_{k \to \infty} \int_M \phi_{C_n}^{\delta_j} \, dv_k = \int_M \phi_{C_n}^{\delta_j} \, dv \xrightarrow{j \to \infty} \int_M \phi_{C_n} \, dv$$

by the monotone convergence theorem. $\square$

3.2. *Correspondence principle for upper density*

THEOREM 3.4. *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 $n \geqslant 1$,

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

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

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

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

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

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

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

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

As $- \sum_{l(v) = L_k} \pi_k(v) \log_q \pi_k(v) = \log_q |T(L_k)|$ by the definition of π_k, summing the inequalities (5) over $v \in L_k$ and noting $\sum_{l(v) = L_k} \pi_k(v) = 1$ 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 2.7 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_n} gives

$$
\begin{aligned}
\overline{d}(C_n) &\geqslant \limsup_{k \to \infty} \frac{1}{|T(L_k)|} \sum_{v \in T(L_k)} \frac{|C_n \cap \{w \in T : w \leqslant v\}|}{L_k + 1} \\
&= \limsup_{k \to \infty} \frac{1}{L_k + 1} \sum_{l(w) \leqslant L_k} \pi_k(w) 1_{C_n}(w)
\end{aligned}
$$

$$\geqslant \limsup_{k \to \infty} \frac{1}{L_k + 1} \sum_{l(w) \leqslant L_k} \pi_k(w) \phi_{C_n}(\pi_k^w)$$

$$= \limsup_{k \to \infty} \int_M \phi_{C_n} \, d\mu_k.$$

The conclusion follows from Lemma 3.3. □

3.3. *Correspondence principle for upper Banach density*

THEOREM 3.5. *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* $H(\eta) \geqslant \dim T - \epsilon$,

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

and for every configuration C *and integer* $n \geqslant 1$,

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

Proof. For any $\epsilon > 0$, by Frostman's lemma (see [**Ma95**, Theorem 8.8] and [**Ho14**, Theorem 3.12]) 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

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

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

LEMMA 3.6. [**Fur70**, Lemma 4] *We have* $H(\eta') \geqslant \dim \theta$.

Proof. As in the calculation (4) we have

$$H(\eta') = - \lim_{k \to \infty} \frac{1}{M_k + 1} \sum_{l(v) = M_k + 1} \theta(v) \log_q \theta(v)$$

$$= \lim_{k \to \infty} \frac{-\sum_{v \in \Pi_k} \theta(v) \log_q \theta(v)}{\sum_{v \in \Pi_k} l(v) \theta(v)} \geqslant \dim \theta,$$

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

The support of η' is contained in the compact set $D(\theta) = \overline{\{\theta^v : v \in |\theta|\}}$, and by Choquet's theorem [**Ph01**, Ch. 3] there exists 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 2.7.

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

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

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

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

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 $\theta^{v_k} \to \tau$, and passing to a subsequence of (v_k) if necessary, it follows that the sequence of measures η_k defined by

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

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

$$d^*(C_n) \geqslant \limsup_{k \to \infty} \frac{1}{N_k + 1} \sum_{l(w) \leqslant N_k} \theta^{v_k}(w) 1_{C_n}(v_k w)$$

$$\geqslant \limsup_{k \to \infty} \frac{1}{N_k + 1} \sum_{l(w) \leqslant N_k} \theta^{v_k}(w) \phi_{C_n}(\theta^{v_k w})$$

$$= \limsup_{k \to \infty} \int_M \phi_{C_n} \, d\eta_k$$

and the inequality (7) follows from Lemma 3.3. $\square$

Remark 3.7. Composing the projection $(\tau_i)_{i \leqslant 0} \mapsto \tau_0$ with a C_n-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 n. The recursive constructions of configuration-detecting functions in §3.1 can be used to construct their lifts 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_n} are replaced with their lifts on $\widetilde{M}$. In the remainder of the paper, we work only with $(\widetilde{M}, p)$ and use Theorems 3.4 and 3.5 for the endomorphic extension without comment.

4. *Proof of direct theorems*

Using the correspondence principles of §3, we bound from below the densities of the sets of popular differences for trees. We first prove such a result for the generic parameters of the configuration F^r, because the proof is relatively simple but contains the main ideas.

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

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

Proof. As P and S are adjoint, Theorem 3.4 gives

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

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

$$\underline{d}(\mathcal{R}) = \liminf_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} 1_{\mathcal{R}}(n) \geqslant \liminf_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} \frac{\widetilde{\mu}(A_r \cap S^{-n}A_r)}{\widetilde{\mu}(A_r)} \geqslant \widetilde{\mu}(A_r),$$

and the theorem follows from inequality (2) of Theorem 3.4.

Using Theorem 3.5 in place of Theorem 3.4 in the above argument, we obtain the second inequality after taking $\epsilon \to 0$. $\qquad \square$

Theorem 4.1 is also immediate from the corresponding result for $D^{r,2}$, which we now prove.

THEOREM 4.2. *Let T be a tree. For $2 \leqslant r \leqslant q$ we have*

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

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 2.10,

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

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

As $\overline{1_{A_r}} \in \mathcal{H}^\infty$, by Lemma 2.9 $P^{n-1}\overline{1_{A_r}} = SP^n\overline{1_{A_r}}$. Then, by Lemma 2.8 and orthogonality,

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

recalling $\varphi_{D_n^{r,1}} = r! \sum_{\substack{I \subset \Lambda \\ |I|=r}} \prod_{i \in I} P1_{B_i}$. Define $Z_\rho = \{\widetilde{\tau} \in \widetilde{M}: \overline{\varphi_{D_n^{r,1}}}(\widetilde{\tau}) \geqslant \rho\}$, and observe it is well-defined up to a $\widetilde{\mu}$-null set. As $0 \leqslant \varphi_{D_n^{r,1}} \leqslant 1_{A_r} \leqslant 1$ and $0 \leqslant \rho 1_{Z_\rho} \leqslant \overline{\varphi_{D_n^{r,1}}} \leqslant 1$, the positivity of both P and conditional expectation imply

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

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

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

Combining the above with the correspondence principle Theorem 3.4, it follows that $\overline{G}(D^{r,2})$ contains a cofinite subset of

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

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

$$\underline{d}(\overline{G}(D^{r,2})) \geqslant \underline{d}(\mathcal{R}^{\delta}(Z_{\rho})) \geqslant \liminf_{N \to \infty} \frac{1}{N+1} \sum_{\substack{n \leqslant N \\ n \in \mathcal{R}^{\delta}(Z_{\rho})}} \frac{\widetilde{\mu}(Z_{\rho} \cap S^{-n} Z_{\rho})}{\widetilde{\mu}(Z_{\rho})}$$

$$\geqslant \left(\liminf_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} \frac{\widetilde{\mu}(Z_{\rho} \cap S^{-n} Z_{\rho})}{\widetilde{\mu}(Z_{\rho})} \right) - \delta \widetilde{\mu}(Z_{\rho})$$

$$\geqslant (1 - \delta) \widetilde{\mu}(Z_{\rho}) \xrightarrow{\delta \to 0} \widetilde{\mu}(Z_{\rho}) \xrightarrow{\rho \to 0} \widetilde{\mu}(Z),$$

where the last inequality follows from the mean ergodic theorem and

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

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

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

Hence, $Z \supset \{\widetilde{\tau} \in \widetilde{M} : \varphi_{D_n^{r,1}}(\widetilde{\tau}) > 0\} = A_r$ 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 3.4.

Using Theorem 3.5 in place of Theorem 3.4 in the above proofs, we obtain the second inequality after taking $\epsilon \to 0$. □

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

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

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

LEMMA 5.2. *If $n \in \mathcal{R}_{\gamma}$ and $n' \in \mathcal{R}_{\gamma'}$, then $n + n' \in \mathcal{R}_{\gamma+\gamma'}$.*

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

$$\nu(A \cap S^{-(n+n')} A) \geqslant \nu(A \cap S^{-n}(A \cap S^{-n'} A)) \geqslant \nu(A \cap S^{-n'} A) - \gamma \nu(A)$$

$$\geqslant (1 - \gamma - \gamma')\nu(A),$$

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

LEMMA 5.3. *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

$$d_{(N_k)}(\mathcal{R}_\gamma) = \lim_{k \to \infty} \frac{|\mathcal{R}_\gamma \cap \{0, \dots, N_k\}|}{N_k + 1}$$

exists. By the mean ergodic theorem

$$\nu(A) \leqslant \lim_{k \to \infty} \frac{1}{N_k + 1} \sum_{n=0}^{N_k} \frac{\nu(A \cap S^{-n}A)}{\nu(A)} = \lim_{k \to \infty} \frac{1}{N_k + 1} \sum_{n=0}^{N_k} 1_{\mathcal{R}}(n) \frac{\nu(A \cap S^{-n}A)}{\nu(A)}$$

$$\leqslant \lim_{k \to \infty} \frac{1}{N_k + 1} \left(\sum_{n \in \mathcal{R}_\gamma, n \leqslant N_k} 1_{\mathcal{R}}(n) + \sum_{n \notin \mathcal{R}_\gamma, n \leqslant N_k} 1_{\mathcal{R}}(n)(1 - \gamma) \right)$$

$$= d_{(N_k)}(\mathcal{R}_\gamma) + (1 - \gamma)(d(\mathcal{R}) - d_{(N_k)}(\mathcal{R}_\gamma)).$$

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 5.2 $\mathcal{R}_\gamma + \mathcal{R}_\gamma \subset \mathcal{R}_{2\gamma} \subset \mathcal{R}$, 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 $d(\mathcal{R})$. As $(N_k)_{k \geqslant 1}$ was arbitrary, the conclusion follows. □

For $0 < \gamma < \frac{1}{2}$, Lemma 5.3 and Kneser's theorem [**Kne53**] (see also [**Bil97**, Theorem 1.1]) therefore imply the existence of $m \geqslant 1$ and $K \subset \{0, \dots, m - 1\}$ such that:

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

It follows that $K = \{0\}$, so $\mathcal{R}_\gamma \subset m\mathbb{N}$ and $d(\mathcal{R}) = d(\mathcal{R}_\gamma) = d(\mathcal{R}_\gamma + \mathcal{R}_\gamma) = m^{-1}$. Further, for all $n \in \mathcal{R}$ there exists $\gamma > 0$ small enough such that $n + \mathcal{R}_\gamma \subset \mathcal{R}$ by Lemma 5.2. As, in addition, $\mathcal{R}_\gamma \subset \mathcal{R}$ and $d(\mathcal{R}) = d(\mathcal{R}_\gamma)$, it follows that $n \in m\mathbb{N}$. Then the m sets $S^{-i}A, 0 \leqslant i \leqslant m - 1$ are disjoint (up to ν-null sets) and each of measure m^{-1}. □

THEOREM 5.4. *If* $(X, \mathscr{B}, \nu, S)$ *is ergodic and*

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

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

Proof. By [**BFS21**, §1.5] all the theorems in [**BFS21**] hold for ergodic $\mathbb{N}$-actions, so [**BFS21**, Theorem 1.3] gives the existence of $m \geqslant 1$ such that $\mathcal{R} = m\mathbb{N}$. Therefore, the sets

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

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 inverse results for Theorems 4.1 and 4.2 (Theorems 6.1, 6.2 and 6.4) using the results of the previous section.

THEOREM 6.1. *If T is a tree and $2 \leqslant r \leqslant q$ with*

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

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

Proof. Let $\mathcal{R} = \{n \in \mathbb{N} : \widetilde{\mu}(A_r \cap S^{-n} A_r) > 0\}$. Combining the proof of Theorem 4.1 and the hypothesis, we obtain

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

Therefore, $\widetilde{\mu}(A_r) = \overline{d}(\overline{G}(F^r)) = \overline{d}(\mathcal{R}) = \underline{d}(\mathcal{R})$ is positive; by Theorem 5.1 it equals m^{-1} for some positive integer m, and $\widetilde{M} = \bigsqcup_{i=0}^{m-1} S^{-i} A_r$ up to $\widetilde{\mu}$-null sets.

The above also shows that equality holds in Proposition 2.7 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 §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:

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

Using the facts determined previously, we compute $\phi'_{V_1^{r,m,mk}} = (R_2 R_1^{m-1})^k 1_{A_q}$. In the following, equalities are only up to $\widetilde{\mu}$-null sets. We compute first the case $k = 1$. Note that $A_q = S^{-m} A_q$ and $1_{S^{-i} A_q} = S 1_{S^{-i+1} A_q}$ for $i \geqslant 1$. By Lemma 2.8

$$R_1 1_{A_q} = \sum_{\substack{I \subset \Lambda \\ |I| = r-1}} \prod_{i \in I} P(1_{B_i} 1_{A_q}) = \sum_{\substack{I \subset \Lambda \\ |I| = r-1}} \prod_{i \in I} P(1_{B_i} S 1_{S^{-m+1} A_q})$$

$$= 1_{S^{-m+1} A_q} \sum_{\substack{I \subset \Lambda \\ |I| = r-1}} \prod_{i \in I} P 1_{B_i}$$

$$= c_2 1_{S^{-m+1} A_q},$$

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

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

and, hence,

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

Letting $d_1 = c_1 c_2^{q \sum_{j=0}^{m-2}(r-1)^j}$ and defining inductively $d_k = d_{k-1}^{q(r-1)^{m-1}} d_1$, it follows that $\phi'_{V_1^{r,m,mk}} = d_k 1_{A_q}$ $\widetilde{\mu}$-almost everywhere. Then

$$\overline{d}(V_1^{r,m,mk}) \geqslant \int_{\widetilde{M}} \phi_{V_1^{r,m,mk}} \, d\widetilde{\mu} \geqslant \int_{\widetilde{M}} \phi'_{V_1^{r,m,mk}} \, d\widetilde{\mu} = d_k \widetilde{\mu}(A_q) = \frac{d_k}{m} > 0$$

for all $k \geqslant 1$ by the correspondence principle Theorem 3.4. $\qquad \square$

THEOREM 6.2. *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 $\dim T = m^{-1}(1 - \log_q(r-1)) + \log_q(r-1)$ for some positive integer m. Moreover, $d^*(V_1^{r,m,mk}) > 0$ for every $k \geqslant 1$.

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 4.1 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 5.4 there is a positive integer m such that $\mathcal{R} = m\mathbb{N}$ and $\widetilde{M} = \bigsqcup_{i=0}^{m-1} S^{-i}(\bigcup_{j=0}^{\infty} S^{-mj} A_r)$ up to $\widetilde{\eta}_\epsilon$-null sets.

In the case of (12), we invoke the proof of Theorem 4.2. 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 $\bigcup_{\rho>0} Z_\rho = Z$ such that for every $\delta > 0$ we have

$$\frac{\dim T - \log_q(r-1)}{1 - \log_q(r-1)} = \overline{d}(G^*(D^{r,2})) \geqslant \overline{d}(\mathcal{R}^\delta(Z_\rho))$$

$$\geqslant \left(\liminf_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} \frac{\widetilde{\eta_\epsilon}(Z_\rho \cap S^{-n} Z_\rho)}{\widetilde{\eta_\epsilon}(Z_\rho)} \right) - \delta \widetilde{\eta_\epsilon}(Z_\rho)$$

$$\geqslant (1-\delta)\widetilde{\eta_\epsilon}(Z_\rho) \xrightarrow{\delta \to 0} \widetilde{\eta_\epsilon}(Z_\rho) \xrightarrow{\rho \to 0} \widetilde{\eta_\epsilon}(Z)$$

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

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 $m \geqslant 1$ such that $\mathcal{R}(Z_\rho) = \mathcal{R}^\delta(Z_\rho) = m\mathbb{N}$, where $\mathcal{R}(Z_\rho) = \{n \in \mathbb{N} \colon \widetilde{\eta_\epsilon}(Z_\rho \cap S^{-n} Z_\rho) > 0\}$. As 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 $m \geqslant 1$ such that $\mathcal{R} \subset m\mathbb{N}$. This immediately implies that $\widetilde{M} = \bigsqcup_{i=0}^{m-1} S^{-i}(\bigcup_{j=0}^{\infty} S^{-mj} A_r)$ up to $\widetilde{\eta_\epsilon}$-null sets.

In both cases, for small ϵ the above inequalities force

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

and, hence,

$$\widetilde{\eta_\epsilon}(A_r) \geqslant (1-\epsilon')m^{-1} = (1-\epsilon')\widetilde{\eta_\epsilon}\left(\bigcup_{j=0}^{\infty} S^{-mj} A_r \right), \tag{13}$$

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 $\mathbf{A_r} = \bigcup_{j=0}^{\infty} S^{-mj} A_r$. Then we have $\widetilde{M} = \bigsqcup_{i=0}^{m-1} S^{-i} \mathbf{A_r}$. 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 m divides i and $\mathbf{A_r^c}$ otherwise. Then the m-periodicity of $\mathbf{A_r}$ and $\mathbf{A_r^c} = \bigsqcup_{i=1}^{m-1} S^{-i} \mathbf{A_r}$ 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

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

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 show the existence of δ such that the configuration $V_1^{r,m,mk}$ appears at the root of $|\tau_0|$ for every $\widetilde{\tau} = (\tau_i)_{i \leqslant 0} \in A_\delta$. First note that, by the construction of $\mathbf{A_{r'}}$, if $\widetilde{\tau} \in A_\delta$ and $v \in |\tau_0|$ with $0 \leqslant l(v) \leqslant mk$, then $H(\widetilde{\tau}^v) \leqslant c_{l(v)}$. Hence, for $0 \leqslant i \leqslant mk$

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

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 $V_1^{r,m,mk}$ at the root of $|\tau_0|$ it therefore suffices to give sufficiently large lower bounds for $H(\widetilde{\tau}^v)$ for $l(v) \leqslant mk$.

LEMMA 6.3. *For every $\delta_1, \delta_2 > 0$ there exists $\delta > 0$ such that for $1 \leqslant j \leqslant mk + 1$: (a) the set $\{\tau_0(v) : \widetilde{\tau} \in A_\delta, v \in |\tau_0|(j)\} \subset [0, 1]$ is contained in an interval of length less than δ_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 $\{\tau_0(v) : \widetilde{\tau} \in A_\delta, v \in |\tau_0|(1)\}$ is contained in an interval of length $< \delta_1$ (containing q^{-1}) for δ small enough.

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

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

and rearranging gives

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

By statement (a) of the induction hypothesis,

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

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

It follows that any $V_1^{r,m,mk}$-detecting function is positive on A_δ. By the correspondence principle Theorem 3.5 we have for all $\epsilon > 0$

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

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

THEOREM 6.4. *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 $m \geqslant 1$ such that $m\mathbb{N} \subset G^*(F^r)$.

Proof. For small enough $\epsilon > 0$, by the correspondence principle Theorem 3.5 and the proof of Theorem 4.1

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

where $\mathcal{R} = \{n \in \mathbb{N} \colon \widetilde{\eta}_\epsilon(A_r \cap S^{-n} A_r) > 0\}$. Theorem 5.4 then implies that there exists $m \geqslant 1$ such that $m\mathbb{N} = \mathcal{R} \subset G^*(F^r)$. $\square$

Question 6.5. It follows from the work of Furstenberg and Weiss in [**FW03**] that for every k there exists n such that $d^*(D_n^{2,k}) > 0$ provided that $\dim T > 0$. On the other hand, under the assumptions of Theorem 6.4, there exists $m \geqslant 1$ such that $m\mathbb{N} \subset G^*(F)$. In analogy to Proposition 1.5, is it true that the stronger claim $d^*(D_m^{2,k}) > 0$ holds true for every k satisfying $(1 - \beta^{-1})k < 1$?

A. *Appendix. Stability in an inverse theorem for return times of ergodic systems*
In the proof of Theorem 6.2 for the configuration $D^{r,2}$, we are unable to apply Theorem 5.4 because 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 5.4 under assumptions involving δ-return times instead of return times.

Given an ergodic measure-preserving system $(X, \mathcal{B}, \nu, S)$ and $A \in \mathcal{B}$ 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 $0 < \delta < (1 - \gamma)/\nu(A)$ (so $\mathcal{R}_{\gamma} \subset \mathcal{R}^{\delta}$). First, observe that by the mean ergodic theorem,

$$\underline{d}(\mathcal{R}^{\delta}) = \liminf_{N \to \infty} \frac{1}{N + 1} \sum_{n=0}^{N} 1_{\mathcal{R}^{\delta}}(n)$$

$$\geqslant \liminf_{N \to \infty} \frac{1}{N + 1} \sum_{n=0}^{N} \frac{\nu(A \cap S^{-n}A)}{\nu(A)} - \delta\nu(A) = (1 - \delta)\nu(A). \qquad \text{(A.1)}$$

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

$$\underline{d}(\mathcal{R}_{\gamma}) = \lim_{k \to \infty} (N_k + 1)^{-1} \sum_{n=0}^{N_k} 1_{\mathcal{R}_{\gamma}}(n).$$

By the mean ergodic theorem

$$\nu(A) = \lim_{k \to \infty} \frac{1}{N_k + 1} \sum_{n=0}^{N_k} \frac{\nu(A \cap S^{-n}A)}{\nu(A)}$$

$$\leqslant \limsup_{k \to \infty} \frac{1}{N_k + 1} \sum_{\substack{n \leqslant N_k \\ n \in \mathcal{R}_{\gamma}}} \frac{\nu(A \cap S^{-n}A)}{\nu(A)} + \limsup_{k \to \infty} \frac{1}{N_k + 1} \sum_{\substack{n \leqslant N_k \\ n \in \mathcal{R}^{\delta} \setminus \mathcal{R}_{\gamma}}} (1 - \gamma)$$

$$+ \limsup_{k \to \infty} \frac{1}{N_k + 1} \sum_{\substack{n \leqslant N_k \\ n \in (\mathcal{R}^{\delta})^c}} \delta\nu(A)$$

$$\leqslant \underline{d}(\mathcal{R}_{\gamma}) + (1 - \gamma)(\overline{d}(\mathcal{R}^{\delta}) - \underline{d}(\mathcal{R}_{\gamma})) + \delta\nu(A)(1 - \underline{d}(\mathcal{R}^{\delta}))$$

$$\leqslant \gamma\underline{d}(\mathcal{R}_{\gamma}) + (1 - \gamma)(1 + \eta)\nu(A) + \delta\nu(A)(1 - (1 - \delta)\nu(A)),$$

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

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

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

LEMMA A.2. *If $l \in \mathcal{R}^{\delta}$, then $\underline{d}(\mathcal{R}_{l}^{\delta\varepsilon}) \geqslant (1 - \varepsilon)\nu(A)$ 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 $\mathcal{R}_{A,B}^\varepsilon = \{n \in \mathbb{N} : \nu(A \cap S^{-n}B) > \varepsilon \nu(A)\nu(B)\}$. Observe that

$$\mathcal{R}_{A,A\cap S^{-l}A}^\varepsilon = \{n \in \mathbb{N} : \nu(A \cap S^{-n}(A \cap S^{-l}A)) > \varepsilon \nu(A)\nu(A \cap S^{-l}A)\} \subset \mathcal{R}_l^{\delta\varepsilon}.$$

By the mean ergodic theorem,

$$\underline{d}(\mathcal{R}_{A,B}^\varepsilon) = \liminf_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^N 1_{\mathcal{R}_{A,B}^\varepsilon}(n)$$

$$\geqslant \liminf_{N\to\infty} \frac{1}{N+1} \sum_{n=0}^N \frac{\nu(A \cap S^{-n}B)}{\min(\nu(A), \nu(B))} - \varepsilon \max(\nu(A), \nu(B))$$

$$\geqslant (1 - \varepsilon) \max(\nu(A), \nu(B)),$$

so we have

$$\underline{d}(\mathcal{R}_l^{\delta\varepsilon}) \geqslant \underline{d}(\mathcal{R}_{A,A\cap S^{-l}A}^\varepsilon) \geqslant (1-\varepsilon)\nu(A)$$

as required. $\qquad\square$

THEOREM A.3. *If there exists $\eta < 1/5$ such that $\overline{d}(R^\delta) \leq (1+\eta)\nu(A)$ for every $\delta > 0$, then there exists $m \geqslant 1$ such that $\mathcal{R}^\delta = m\mathbb{N}$ 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

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

Observe that $\mathcal{R}_\gamma + \mathcal{R}_\gamma \subset \mathcal{R}_{2\gamma} \subset \mathcal{R}^\delta$ for $0 < \delta < (1-2\gamma)/\nu(A)$ by Lemma 5.2. Noting that (A.2) implies $\gamma - \eta + \gamma\eta > 0$ and $(1+\eta)\gamma/(\gamma - \eta + \gamma\eta) < 2$, Lemma A.1 gives

$$\underline{d}(\mathcal{R}_\gamma + \mathcal{R}_\gamma) \leqslant \underline{d}(\mathcal{R}^\delta) \leqslant \overline{d}(\mathcal{R}^\delta) \leqslant (1+\eta)\nu(A) \leqslant \left(\frac{(1+\eta)\gamma}{\gamma - \eta + \gamma\eta}\right)\underline{d}(\mathcal{R}_\gamma) < 2\underline{d}(\mathcal{R}_\gamma). \tag{A.3}$$

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

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

Combining this with Lemma A.1 gives

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

$$\leqslant \left(\frac{(1+\eta)\gamma}{\gamma - \eta + \gamma\eta}\right)\underline{d}(\mathcal{R}_\gamma) \leqslant \left(\frac{(1+\eta)\gamma}{\gamma - \eta + \gamma\eta}\right)\frac{|K|}{m},$$

and rearranging gives

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

where the last inequality follows from (A.2). Hence, $|K| = 1$. Furthermore, $K = \{0\}$, because 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 $\mathcal{R}^\delta \subset m\mathbb{N}$ for small enough $\delta > 0$. For $l \in \mathbb{N}$ and $\delta > 0$, recall

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

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

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

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

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

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

$$2(1 + \eta)\nu(A) \geqslant 2\overline{d}(\mathcal{R}^{\delta^2\nu(A)}) \geqslant (1 - \delta)\nu(A) + 2m^{-1}. \tag{A.4}$$

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

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

and combining with (A.4) gives $\gamma \leqslant 2\eta/(1 - \delta)$ after rearranging. This is compatible with (A.2) only if $\eta > (1 - 3\delta)/2$. As $\eta < \frac{1}{5}$, it follows that the above requires $\delta > \frac{1}{5}$. Hence, $l \in \mathcal{R}^\delta \setminus m\mathbb{N}$ gives a contradiction and $\mathcal{R}^\delta \subset m\mathbb{N}$ for small enough $\delta > 0$.

Finally, we show $\mathcal{R}^\delta = m\mathbb{N}$ for small δ. Indeed, because $\underline{d}(\mathcal{R}_\gamma) > (2m)^{-1}$ by combining equation (A.3) with the third implication of Kneser's theorem $R_\gamma + R_\gamma \subset K + K + m\mathbb{N}$ and the fact that $|K| = 1$, for every $l \in \mathbb{N}$ there exists $n \in \mathbb{N}$ such that $mn, m(n + l) \in \mathcal{R}_\gamma$. Therefore,

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

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

A.1. *Discussion.* The set of transfer times $\mathcal{R}_{A,B}$ has strong parallels with the difference set $\mathcal{A} - \mathcal{B} = \{a - b : a \in \mathcal{A}, b \in \mathcal{B}\}$, $\mathcal{A}, \mathcal{B} \subset \mathbb{Z}/r\mathbb{Z}$, 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 centralizer 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 5.4 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, §5.1]. Although it is not 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

[BFS21] M. Björklund, A. Fish and I. D. Shkredov. Sets of transfer times with small densities. *J. Éc. polytech. Math.* **8** (2021), 311–329.

[BHK05] V. Bergelson, B. Host and B. Kra. Multiple recurrence and nilsequences. *Invent. Math.* **160**(2) (2005), 261–303. With an appendix by I. Ruzsa.

[Bil97] Y. Bilu. Addition of sets of integers of positive density. *J. Number Theory* **64**(2) (1997), 233–275.

[EFHN15] T. Eisner, B. Farkas, M. Haase and R. Nagel. *Operator Theoretic Aspects of Ergodic Theory (Graduate Texts in Mathematics, 272).* Springer, Cham, 2015.

[EW11] M. Einsiedler and T. Ward. *Ergodic Theory with a View Towards Number Theory (Graduate Texts in Mathematics, 259).* Springer-Verlag London, Ltd., London, 2011.

[Fur70] H. Furstenberg. *Intersections of Cantor sets and transversality of semigroups. Problems in Analysis: A Symposium in Honor of Salomon Bochner (PMS-31).* Ed. R. C. Gunning. Princeton University Press, Princeton, NJ, 1970, pp. 41–59.

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

[Fur81] H. Furstenberg. *Recurrence in Ergodic Theory and Combinatorial Number Theory (M. B. Porter Lectures).* Princeton University Press, Princeton, NJ, 1981.

[FW03] H. Furstenberg and B. Weiss. Markov processes and Ramsey theory for trees. *Combin. Probab. Comput.* **12**(5–6) (2003), 547–563.

[Ho14] M. Hochman. Lectures on dynamics, fractal geometry, and metric number theory. *J. Mod. Dyn.* **8**(3–4) (2014), 437–497.

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

[Ma95] P. Mattila. *Geometry of Sets and Measures in Euclidean Spaces (Cambridge Studies in Advanced Mathematics, 44).* Cambridge University Press, Cambridge, 1995.

[Ph01] R. Phelps. *Lectures on Choquet's Theorem (Lecture Notes in Mathematics, 1757)*, 2nd edn. Springer-Verlag, Berlin, 2001.

[PST12] J. Pach, J. Solymosi and G. Tardos. Remarks on a Ramsey theory for trees. *Combinatorica* **32**(4) (2012), 473–482.

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