Self Similar Sets, Entropy and Additive Combinatorics

Michael Hochman

Abstract This article is an exposition of the main result of [Hoc12], that self-similar sets whose dimension is smaller than the trivial upper bound have "almost overlaps" between cylinders. We give a heuristic derivation of the theorem using elementary arguments about covering numbers. We also give a short introduction to additive combinatorics, focusing on inverse theorems, which play a pivotal role in the proof. Our elementary approach avoids many of the technicalities in [Hoc12], but also falls short of a complete proof; in the last section we discuss how the heuristic argument is turned into a rigorous one.

1 Introduction

1.1 Self-similar Sets

Self-similar sets in the line are compact sets that are composed of finitely many scaled copies of themselves. These are the simplest fractal sets, the prototypical example being the famous middle-$\frac{1}{3}$ Cantor set $X \subseteq [0, 1]$, which satisfies the "geometric recursion"[1] relation $X = \frac{1}{3}X \cup (\frac{1}{3}X + \frac{2}{3})$, using the obvious notation for scaling and translation of a set. In general, a self-similar set is defined by a finite family $\Phi = \{f_i\}_{i \in \Lambda}$ of maps of the form $f_i(x) = r_i x + a_i$, where $0 < |r_i| < 1$ and $a_i \in \mathbb{R}$. The family Φ is called an *iterated function system* (or *IFS*),[2,3] and the self-similar set they define is unique compact set $X \neq \emptyset$ satisfying

---

[1] The mddle-1/3 Cantor set can also be described in other ways, e.g. by a recursive construction, or symbolically as the points in [0, 1] that can be written in base 3 without the digit 1. General self-similar sets also have representations of this kind, but in this paper we shall not use them.
[2] Iterated function systems consisting of non-affine maps and on other metric spaces than $\mathbb{R}$ are also of interest, but we do not discuss them here.
[3] Supported by ERC grant 306494.

M. Hochman (✉)
Department of Mathematics,The Hebrew University of Jerusalem, Jerusalem, Israel
e-mail: mhochman@math.huji.ac.il

$$X = \bigcup_{i \in \Lambda} f_i X. \tag{1}$$

(existence and uniqueness are due to Hutchinson [Hut81]).

Throughout this paper we make a few simplifying assumptions. To avoid trivialities, we always assume that Φ contains at least two distinct maps, otherwise X is just the common fixed point of the maps. We assume that Φ has *uniform contraction*, i.e. all the contraction ratios r_i are equal to the same value r. Finally, we assume that $r > 0$, so the maps preserve orientation. These assumptions are not necessary but they simplify the statements and arguments considerably.

1.2 Dimension of Self-similar Sets

Despite the apparent simplicity of the definition, and of some of the better known examples, there are still large gaps in our understanding of the geometry of self-similar sets. In general, we do not even know how to compute their dimension. Usually one should be careful to specify the notion of dimension that one means, but it is a classical fact that, for self-similar sets, all the major notions of dimension coincide, and in particular the Hausdorff and box (Minkowski) dimensions agree (e.g. [Fal89, Theorem 4 and Example 2]). Thus we are free to choose either one of these, and we shall choose the latter, whose definition we now recall. For a subset $Y \subseteq \mathbb{R}$ denote its covering number at scale ε by

$$N_\varepsilon(Y) = \min\{k \ : \ Y \text{ can be covered by } k \text{ sets of diameter } \leq \varepsilon\}$$

The *box dimension* of Y, if it exists, is the exponential growth rate of $N_\varepsilon(Y)$:

$$\dim_B Y = \lim_{\varepsilon \to 0} \frac{\log N_\varepsilon(Y)}{\log(1/\varepsilon)}$$

Thus $\dim_B Y = \alpha$ means that $N_\varepsilon(Y) = \varepsilon^{-\alpha+o(1)}$ as $\varepsilon \to 0$. It is again well known that the limit exists when Y is self-similar, we shall see a short proof in Sect. 3.1.

It is easy to give upper bounds for the dimension of a self-similar set. Taking X as in (1) and iterating the relation we get

$$X = \bigcup_{i \in \Lambda} f_i(\bigcup_{j \in \Lambda} f_j X) = \bigcup_{i,j \in \Lambda} f_i \circ f_j(X)$$

Writing $f_{i_1 \ldots i_n} = f_{i_1} \circ \ldots \circ f_{i_n}$ for $\underline{i} = i_1 \ldots i_n \in \Lambda^n$ and iterating n times, we have

$$X = \bigcup_{\underline{i} \in \Lambda^n} f_{\underline{i}} X \tag{2}$$

This union consists of $|\Lambda|^n$ sets of diameter $r^n|X|$, so by definition,

$$N_{r^n \operatorname{diam}(X)}(X) \le |\Lambda|^n$$

Hence

$$\dim_{\mathrm{B}}(X) = \lim_{n \to \infty} \frac{\log N_{r^n \operatorname{diam}(X)}(X)}{\log(1/r^n \operatorname{diam}(X))} \le \frac{\log|\Lambda|}{\log(1/r)} \tag{3}$$

The right hand side of (4) is called the *similarity dimension* of X and is denoted $\operatorname{sdim} X$.[4]

Is this upper bound an equality? Note that the bound is purely combinatorial and does not take into account the parameters a_i at all. Equality is known to hold under some assumptions on the separation of the "pieces" $f_i X$, $i \in \Lambda$, for instance assuming *strong separation* (that the union (1) is disjoint), or the open set condition (that there exists open set $\emptyset \ne U \subseteq \mathbb{R}$ such that $f_i U \subseteq U$ and $f_i U \cap f_j U = \emptyset$ for $i \ne j$).

Without separation conditions, however, the inequality (3) can be strict. There are two trivial ways this can occur. First, there could be too many maps: if $|\Lambda| > 1/r$ then the right hand side of (3) is greater then 1, whereas $\dim_{\mathrm{B}} X \le 1$ due to the trivial bound $N(X, \varepsilon) \le \lceil \operatorname{diam}(X)/\varepsilon \rceil$. Thus we should adjust (3) to read

$$\dim_{\mathrm{B}}(X) \le \min\{1, \frac{\log|\Lambda|}{\log(1/r)}\} \tag{4}$$

Second, the combinatorial bound may be over-counting if some of the sets in the union (2) coincide, that is, for some n we have $f_{\underline{i}} = f_{\underline{j}}$ for some distinct $\underline{i}, \underline{j} \in \Lambda^n$. This situation is known as *exact overlaps*. If such $\underline{i}, \underline{j}$ exist then we can re-write (2) as $X = \bigcup_{\underline{u} \in \Lambda^n \setminus \{\underline{i}\}} f_{\underline{u}} X$, which presents X as the attractor of the IFS $\Phi' = \{f_{\underline{u}}\}_{\underline{u} \in \Lambda'}$ for $\Lambda' = \Lambda^n \setminus \{\underline{i}\}$. This IFS consists of $|\Lambda'| = |\Lambda|^n - 1$ maps that contract by r^n, so, applying the trivial bound (3) to this IFS, we have $\dim_{\mathrm{B}} X \le \log (|\Lambda|^n - 1)/\log(1/r^n)$, which is better than the previous bound of $\log|\Lambda|/\log(1/r)$. To take an extreme example, if all the maps f_i coincide then the attractor X is just the unique fixed point of the map, and its dimension is 0.

Are there other situations where a strict inequality occurs in (4)? A-priori, one does not need *exact* coincidences between sets in (2) to make the combinatorial bound very inefficient. It could happen, for example, that many of the sets $f_{\underline{i}} X$, $\underline{i} \in \Lambda^n$, align almost exactly, in which case one may need significantly fewer than $|\Lambda|^n$ ε-intervals to cover them. Nevertheless, although such a situation can easily be arranged for a fixed n, to get a drop in dimension one would need this to happen at all sufficiently small scales. No such examples are known, and the main subject of this paper is the conjecture that this cannot happen:

---

[4] It would be better to write $\operatorname{sdim} \Phi$, since this quantity depends on the presentation of X and not on X itself, but generally there is only one IFS given and no confusion should arise.

Conjecture 1.1 A strict inequality in (4) can occur only in the presence of exact overlaps.

This conjecture appears in [PS00, Question 2.6], though special cases of it have received attention for decades, in particular Furstenberg's projection problem for the 1-dimensional Sierpinski gasket (see e.g. [Ken97]), the 0,1,3-problem (see e.g. [PS95]) and, for self-similar measures instead of sets, the Bernoulli convolutions problem, (e.g. [PSS00]).

One may also draw an analogy between this conjecture and rigidity statements in ergodic theory. Rigidity is the phenomenon that, for certain group actions of algebraic origin, the orbit of the point is as large as it can be (dense or possibly even equidistributed for the volume measure) unless there is an algebraic obstruction to this happening. To see the connection with the conjecture above, note that X is just the orbit closure of (any) $x \in \mathbb{R}$ under the semigroup $\{f_{\underline{i}} : \underline{i} \in \bigcup_{n=1}^{\infty} \Lambda^N\}$ of affine maps, and that exact overlaps occur if and only if this semigroup is not generated freely by $\{f_i\}_{i \in \Lambda}$. Thus the conjecture predicts that the orbit closure of any point is as large as it can be unless there are algebraic obstructions.

1.3 Progress Towards the Conjecture

Our main subject here is a weakened form of the conjecture which proves the full conjecture in some important examples and special classes of IFSs. In order to state it we must first quantify the degree to which the sets $f_{\underline{i}}(X)$ are separated from each other. Since all of the maps in Φ contract by the same ratio, any two of the sets $f_{\underline{i}}(X)$ and $f_{\underline{j}}(X)$ for $\underline{i}.\underline{j} \in \Lambda^n$ are translates of each other. We define the distance between them as the magnitude of this translation, which is given by $f_{\underline{i}}(x) - f_{\underline{j}}(x)$ for any $x \in \mathbb{R}$; we shall choose $x = 0$ for concreteness. Thus a measure of the degree of concentration of cylinders $f_{\underline{i}}(X)$, $\underline{i} \in \Lambda^n$, is provided by

$$\Delta_n = \min\{|f_{\underline{i}}(0) - f_{\underline{j}}(0)| : \underline{i}, \underline{j} \in \Lambda^n , \underline{i} \neq \underline{j}\}$$

Evidently, exact overlaps occur if and only if there exists an n such that $\Delta_n = 0$. Fixing $x \in X$, the points $f_{\underline{i}}(x)$, $\underline{i} \in \Lambda^n$, all lie in X, and so there must be a distinct pair $\underline{i}, \underline{j} \in \Lambda^n$ with $|f_{\underline{i}}(x) - f_{\underline{j}}(x)| \leq \mathrm{diam}(X)/|\Lambda|^n$; hence $\Delta_n \to 0$ at least exponentially. In general there may be an exponential lower bound on Δ_n as well, i.e. $\Delta_n \geq cr^n$ for some $c, r > 0$. This is always the case when the IFS satisfies strong separation or the open set condition, but there are examples where it holds even when these conditions fail (see Garsia [Gar62]). Therefore the following theorem from [Hoc12] gives nontrivial information and should be understood as a weak form of Conjecture 1.1.

Theorem 1.2 *If $X \subseteq \mathbb{R}$ is a self-similar set and* $\dim X < \min\{1, sdimX\}$*, then* $\Delta_n \to 0$ *super-exponentially, that is,* $-\frac{1}{n} \log \Delta_n \to \infty$.

In practice, one applies the theorem after establishing an exponential lower bound on Δ_n to deduce that $\dim X = \min\{1, \operatorname{sdim} X\}$. For example,

Proposition 1.3 *Let $\mathcal{R}$ denote the set of rational IFSs, i.e. such that $r, a_i \in \mathbb{Q}$. Then Conjecture 1.1 holds in $\mathcal{R}$.*

Proof First, a useful identity: For $\underline{i} \in \Lambda^n$, a direct calculation shows that

$$f_{\underline{i}}(x) = r^n x + \sum_{k=1}^{n} a_{i_k} r^{k-1} \tag{5}$$

$$= r^n x + f_{\underline{i}}(0) \tag{6}$$

Now let that $f_i(x) = rx + a_i$ where $r = p/q$ and $a_i = p_i/q_i$ for p, p_i, q, q_i integers and write $Q = \prod_{i \in \Lambda} q_i$. Then $f_{\underline{i}}(0) = \sum_{k=1}^{n} a_{i_k} r^{n-k}$ is a rational number with denominator Qq^n. Suppose that no overlaps occur, so that $\Delta_n > 0$ for all n. Given n, by definition there exist distinct $\underline{i}, \underline{j} \in \Lambda^n$ such that $\Delta_n = f_{\underline{i}}(0) - f_{\underline{j}}(0)$. Therefore Δ_n is a non-zero rational number with denominator Qq^n so we must have $\Delta_n \geq 1/Qq^n$. By Theorem 1.2 we conclude that $\dim X = \min\{1, \operatorname{sdim} X\}$. □

The same argument works in the class of IFSs with algebraic coefficients, using a similar lower bound on polynomial expressions in a given set of algebraic numbers. See [Hoc12, Theorem 1.5]. A simple (but non-trivial) calculation, due to B. Solomyak and P. Shmerkin, also allows one to deal with the case that one of the translation parameters a_i is irrational, resolving Furstenberg's question about linear projections of the one-dimensional Sierpinski gasket [Hoc12, Theorem 1.6]. Theorem 1.2 leads to strong results about parametric families of self-similar sets [Hoc12, Theorem 1.8], and there is a version for measures which has also led to substantial progress on the Bernoulli convolutions problem, see [Hoc12, Theorem 1.9] and the recent advance by Shmerkin [Shm13]. Another interesting application is given in [Orp13].

The rest of this paper is an exposition of the proof of the theorem. Our goal is to present the ideas as transparently as possible, and to this end we frame the argument in terms of covering numbers (rather than entropy as in [Hoc12]). This leads to simpler statements and to an argument that is conceptually correct but, unfortunately, incomplete; some crucial steps of this simplified argument are flawed. In spite of this deficiency we believe that such an exposition will be useful as a guide to the more technical proof in [Hoc12]. To avoid any possible misunderstandings, we have indicated the false statements in quotation marks ("Lemma", "Proof", etc.).

As we shall see, the main idea is to reduce (the negation of) the theorem to a statement about sums of self-similar sets with other sets. Problems about sums of sets fall under the general title of additive combinatorics, and in the next section we give a brief introduction to the parts of this theory that are relevant to us. In Sect. 3 we explain the reduction to a statement about sumsets, and show how an appropriate inverse theorem essentially settles the matter. Finally, in Sect. 4, we discuss how the heuristic argument can be made rigorous.

2 A Birds-Eye View of Additive Combinatorics

2.1 Sumsets and Inverse Theorems

The sum (or sumset) of non-empty sets $A, B \subseteq \mathbb{R}^d$ is

$$A + B = \{a + b : a \in A, b \in B\}$$

Additive combinatorics, or at least an important chapter of it, is devoted to the study of sumsets and the relation between the structure of A, B and $A + B$. We focus here on so-called inverse problems, that is the problem of describing the structure of sets A, B such that $A + B$ is "small" relative to the sizes of the original sets. The general flavor of results of this kind is that, if the sumset is small, there must be an algebraic or geometric reason for it. It will become evident in later sections that this question comes up naturally in the study of self-similar sets.

To better interpret what "small" means, first consider the trivial bounds. Assume that A, B are finite and non-empty. Then $|A + B| \geq \max\{|A|, |B|\}$, with equality if and only if at least one of the sets is a singleton. In the other direction, $|A + B| \leq |A||B|$, and equality can occur (consider $A = \{0, 10, 20, 30, \ldots, 10n\}$ and $B = \{0, 1, \ldots, 9\}$). For "generic" pairs of sets the upper bound is close to the truth. For example, when $A, B \subseteq \{1, \ldots, n\}$ are chosen randomly by including each $1 \leq i \leq n$ in A with probability p and similarly for B, with all choices independent, there is high probability that $|A + B| \geq c|A||B|$. The question becomes, what can be said between these two extremes.

2.2 Minimal Growth

One of the earliest inverse theorems is the Brunn-Minkowski inequality of the late 19th century. The setting is $\mathbb{R}^d$ with the volume measure, and it states that if $A, B \subseteq \mathbb{R}^d$ are convex sets then, given the volumes of A, B, the volume of $A+B$ is minimized when A, B are balls with respect to some common norm. Since the volume of a ball scales like the d-th power of the radius, this means that $\mathrm{vol}(A + B) \geq (\mathrm{vol}(A)^{1/d} + \mathrm{vol}(B)^{1/d})^d$, and equality occurs if and only if, up to a nullset, A, B are dilates of the same convex set. The inequality was later extended to arbitrary Borel sets (note that $A + B$ may not be a Borel set but it is an analytic set and hence Lebesgue measurable). For a survey of this topic see Gardner [Gar02].

Similar tight statements hold in the discrete setting. The analog of a convex body is an *arithmetic progression (AP)*, namely a set of the form $P = \{a, a + d, a + 2d, \ldots, a + (1 - k)d\}$, where d is called the gap (we assume $d \neq 0$) and k is called the length of P. Then for finite sets $A, B \subseteq \mathbb{Z}^d$ with $|A|, |B| \geq 2$ we always have $|A + B| \geq |A| + |B| - 1$, with equality if and only if A, B are APs of the same gap [TV06, Proposition 5.8].

2.3 Linear Growth: Small Doubling and Freiman's Theorem

Now suppose that $A = B \subseteq \mathbb{Z}^d$ but weaken the hypothesis, assuming only that

$$|A + A| \leq C|A| \tag{7}$$

where we think of A as large and C as constant. Such sets are said to have *small doubling*.

The simplest example of small doubling in $\mathbb{Z}^d$ is when $A = \{1, \ldots, n\}^d$, in which case $|A + A| \leq 2^d|A|$. This example can be pushed down to any lower dimension as follows. For $i = 1, \ldots, k$, take intervals of integers $I_i = \{1, 2, \ldots, n_i\}$, and let $T : \mathbb{Z}^k \to \mathbb{Z}^d$ be an affine map given by integer parameters. Suppose that T is injective on $I = I_1 \times \ldots \times I_k$. Then $A = T(I) \subseteq \mathbb{Z}^d$ has the property that

$$|A + A| = |T(I) + T(I)| = |T(I + I)| \leq |I + I| \leq 2^k|I| = 2^k|A|$$

A set A as above is called a *(proper) generalized arithmetic progression (GAP)* of rank k.

GAPs are still extremely algebraic objects but one can get away from this a little using another cheap trick: Begin with a set A satisfying $|A + A| \leq C|A|$ (e.g. a GAP) and choose any $A' \subseteq A$ with cardinality $|A'| \geq D^{-1}|A|$ for some $D > 1$. Then

$$|A' + A'| \leq |A + A| \leq C|A| \leq CD|A'|$$

One of the central results of additive combinatorics is Freiman's theorem, which says that, remarkably, these are the only ways to get small doubling.

Theorem 2.1 (Freiman) *If $A \subseteq \mathbb{Z}^d$ and $|A + A| \leq C|A|$, then $A \subseteq P$ for a GAP P of rank C' and satisfying $|P| \leq C''|A|$. The constants satisfy $C' = O(C(1 + \log C))$ and $C'' = C^{O(1)}$.*

For more information see [TV06, Theorem 5.32 and Theorem 5.33].

Combined with some standard arguments (e.g. the Plünnecke-Rusza inequality), the symmetric version leads to an asymmetric versions: assuming $A, B \subseteq \mathbb{Z}^d$ and $C^{-1} \leq |A|/|B| \leq C$, if $|A + B| \leq C|A|$ then A, B are contained in a GAP P of rank and $\leq C'$ and size $|P| \leq C'|A|$, with similar bounds on the constants.

2.4 Power Growth, the "Fractal" Regime

Now relax the growth condition even more and consider finite sets $A \subseteq \mathbb{Z}$ (or $A \subseteq \mathbb{R}$) such that

$$|A + A| \leq |A|^{1+\delta} \tag{8}$$

This is the discrete analog of the condition

$$\dim_B(X + X) \leq (1 + \delta)\dim_B X \tag{9}$$

for $X \subseteq \mathbb{R}$. Indeed, given $X \subseteq \mathbb{R}$ and $n \in \mathbb{N}$ let X_n denote the set obtained by replacing each $x \in X$ with the closest point $k/2^n$, $k \in \mathbb{Z}$. Then $|X_n| \sim 2^{n(\dim_B X + o(1))}$ and $|X_n + X_n| \sim 2^{n(\dim_B(X+X)+o(1))}$ for large n, so (9) is equivalent to $|X_n + X_n| \lesssim |X_n|^{1+o(1)}$. Thus, the difference between (7) and (8) is roughly the difference between using Lebesgue measure or dimension to quantify the size of a set $X \subseteq \mathbb{R}$.

Here is a typical example of a set satisfying (8). Write $P_n = \{0, \ldots, n-1\}$ and let

$$\begin{aligned}
A_n &= \sum_{i=1}^{n} \frac{1}{2^{i^2}} P_{2^i} \\
&= \{\sum_{i=1}^{n} a_i 2^{-i^2} \; : \; 1 \leq a_i \leq 2^i\}
\end{aligned}$$

(again, one can think of this either as a subset of $\mathbb{R}$, or of $\frac{1}{4^{n^2}}\mathbb{Z}$). It is easy to verify that the distance between distinct points $x, x' \in A_n$ is at least $1/4^{n^2}$, and that such x has a unique representation as a sum $\sum_{i=1}^{n} a_i 4^{-i^2} \; : \; 1 \leq a_i \leq 2^i$. Indeed, each term in the sum $\sum_{i=1}^{n} a_i 2^{-i^2}$ determines a distinct block of binary digits. Thus A_n is a GAP, being the image of $P_2 \times P_4 \times \ldots \times P_{2^n}$ by the map $(x_1, \ldots, x_n) \mapsto \sum \frac{1}{2^{i^2}} x_i$. The rank is n, and so, as we saw in the previous section,

$$|A_n + A_n| \leq 2^n |A_n|$$

Since

$$|A_n| = \prod_{i=1}^{n} |P_n| = 2^{\sum_{i=1}^{n} i} = 2^{n(n+1)/2}$$

we have

$$|A_n + A_n| = |A_n|^{1+o(1)} \qquad \text{as } n \to \infty$$

The reader may recognize the example above as the discrete analog of a Cantor set construction, where at stage n we have a collection of intervals $2^{n(n+1)/2}$ of length 2^{-n^2}, and from each of these intervals we keep 2^{n+1} sub-intervals of length $2^{-(n+1)^2}$, separated by gaps of length $2^{-n^2-(n+1)}$. For the resulting Cantor set X it is a standard exercise to see that $\dim X = \dim_B X = 1/2$, and the calculation above shows that $\dim X + X = 1/2$ as well. Such constructions appear in the work of Erdős-Volkmann [EV66], and also in the papers of Schmeling-Shmerkin [SS10] and

Körner [Kör08], who showed that for any sequence $\alpha_1 \leq \alpha_2 \leq \ldots$ there is a set X with dim $\sum_{i=1}^{n} X = \alpha_n$.

Do all examples of (9) look essentially like this one? In principle one can apply Freiman's theorem, since the hypothesis (9) can be written as $|A + A| \leq C|A|$ for $C = |A|^{\delta}$. What one gets, however, is that A is a $|A|^{O(\delta)}$-fraction of a GAP or rank $|A|^{O(\delta)}$, and this gives rather coarse information about A (note that, trivially, every set is a GAP of rank $|A|$).

Instead, it is possible to apply a multi-scale analysis, showing that at some scales the set looks quite "dense" and at others quite "sparse". The best way to explain this is in the language of trees, which we introduce next.

2.5 Trees and Tree-Measures

Denote the length of a finite sequence $\sigma = \sigma_1 \ldots \sigma_n$ by $|\sigma| = n$ and write $\emptyset$ for the empty word, which by definition has $|\emptyset| = 0$. Denote the concatenation of words σ and τ by $\sigma\tau$, in which case we say that σ is a prefix of τ, and that $\sigma\tau$ extends σ.

The *full binary tree* of height h is the set $\{0, 1\}^{\leq h} = \bigcup_{k=0}^{h} \{0, 1\}^k$ of 0, 1-valued sequences of length $\leq h$, where our convention is that $\{0, 1\}^0 = \{\emptyset\}$, so the empty word is included. We define a *tree of height h* is a subset $T \subseteq \bigcup_{i=0}^{h} \{0, 1\}^i$ satisfying

(T1) $\emptyset \in T$.
(T2) If $\sigma \in T$ and η is an initial segment of σ then $\eta \in T$.
(T3) If $\sigma \in T$ then there is an $\eta \in T$ which extends σ and $|\eta| = h$.

One may think of T as a set of vertices and introduce edges between every pair $\sigma_1 \ldots \sigma_i, \sigma_1 \ldots \sigma_i\sigma_{i+1} \in T$. Then T is a tree if $\emptyset \in T$ and in the associated graph there is a path from $\emptyset$ to every node, and all maximal paths are of length h.

The *level (or depth)* of $\sigma \in T$ is its length (the graph-distance from $\emptyset$ to σ). The *leaves* of a tree T of height h are the elements of the lowest (deepest) level, namely h:

$$\partial T = T \cap \{0, 1\}^h$$

The *descendants* of $\sigma \in T$ are the nodes $\eta \in T$ that extend σ. The nodes m generations below σ in T are the nodes of the form $\eta = \sigma\sigma' \in T$ for $\sigma' \in \{0, 1\}^m$.

We also shall need to work with measures "on trees", or, rather, measures on their leaves. For notational purposes it is better to introduce the notion of a *tree-measure*[5] on the full tree $\{0, 1\}^{\leq h}$, namely, a function $\mu : \{0, 1\}^{\leq h} \to [0, 1]$ satisfying

(M1) $\mu(\emptyset) = 1$.
(M2) $\mu(\sigma) = \sum_{i \in \{0,1\}} \mu(\sigma i)$

It is easily to derive from (M1) and (M2) that $\sum_{\sigma \in \{0,1\}^k} \mu(\sigma) = 1$ for every $1 \leq k \leq h$, so a tree-measure induces genuine probability measures on every level

---

[5] This notion is identical to a flow on the tree in the sense of network theory.

of the full tree, and in particular on ∂T. Conversely, if we have a genuine probability measure μ on the set of leaves $\{0, 1\}^h$ of the full tree of height h then it induces a tree-measure by $\mu(\sigma) = \sum_{\eta\,:\,\sigma\eta\in\partial T} \mu(\{\sigma\eta\})$. Given a tree-measure, the set $T = \{\sigma\,:\,\mu(\sigma) > 0\}$ is a tree which might be called the support of μ.

Every tree-measure μ on $\{0, 1\}^{\leq h}$ defines a distribution on the nodes of the tree as follows: first choose a level $0 \leq i \leq h$ uniformly, and then choose a node $\sigma \in \{0, 1\}^i$ in level i with the probability given by μ, i.e. $\mu(\sigma)$ (we have already noted that at each level the masses sum to 1). Thus the probability of $A \subseteq T$ is

$$\mathbb{P}_\mu(A) = \frac{1}{h+1} \sum_{\sigma\in A} \mu(\sigma)$$

and the expectation of $f : \{0, 1\}^{\leq h} \to \mathbb{R}$ is

$$\mathbb{E}_\mu(f) = \frac{1}{h+1} \sum_{k=0}^{n} \sum_{\sigma\in\{0,1\}^k} \mu(\sigma) f(\sigma)$$

Sometimes we write $\mathbb{P}_{\sigma\sim\mu}$ or $\mathbb{E}_{\sigma\sim\mu}$ to define σ as a random node, as in the expression

$$\mathbb{P}_{\sigma\sim\mu}(\sigma \in T \text{ and } \sigma \text{ has two children in } T) = \frac{1}{h+1} \sum_{\sigma\in T} \mu(\sigma) 1_{\{\sigma 0,\sigma 1\in T\}}$$

Given the tree T of height h, it is natural to consider the uniform probability measure $\mu_{\partial T}$ on ∂T and, as described above, extend it to a tree-measure, which we denote μ_T. In this case we abbreviate the probability and expectation operators above by $\mathbb{P}_T$ and $\mathbb{E}_T$, etc. It is important to note that *choosing a node according to μ_T is not the same as choosing a node uniformly from T*. The latter procedure is usually heavily biased towards sampling from the leaves, since these generally constitute a large fraction of the nodes (in the full binary tree, sampling this way gives a leaf with probability $> 1/2$). In contrast, μ_T samples uniformly from the levels, and within each level we sample according to the relative number of leaves descended from each node.

Trees and tree-measures are naturally related to sets and measures on $[0, 1)$ using binary coding. Given a set $X \subseteq [0, 1)$ and $h \in \mathbb{N}$, we lift X to a tree T of height h by taking all the initial sequences of length $\leq h$ of binary expansions of points in X, with the convention that the expansion terminates in 1s if there is an ambiguity. We remark that for $k \leq h$,

$$N_{1/2^k}(X) \leq \left| T \cap \{0, 1\}^k \right| \leq 2N_{1/2^k}(X)$$

Similarly, a probability measure μ on $[0, 1)$ can be lifted to a tree-measure $\widetilde{\mu}$ on $\{0, 1\}^{\leq h}$ by defining $\widetilde{\mu}(\sigma)$ equal to the mass of the interval of numbers whose binary expansion begins with σ.

Fig. 1 A tree with alternating levels having full branching at some levels, full concentration at others, and a few levels omitted. Schematically this is what the tree associated to A_n from Sect. 2.4 looks like, as well as the conclusion of Theorem 2.2 (with W indicated by the small space between levels)
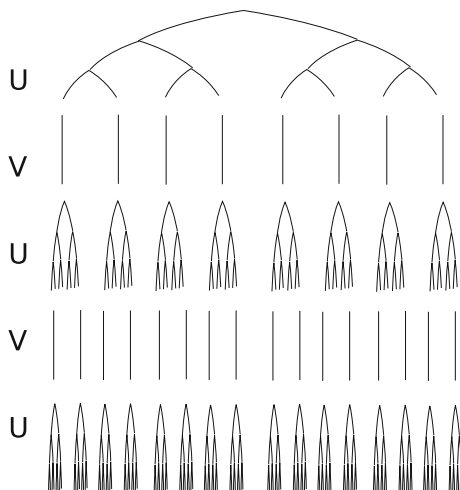
2.6 Inverse Theorems in the Power-Growth Regime

We need some terminology for describing the local structure of trees. We say that T has *full branching for m generations at σ* if σ has all 2^m possible descendants m generations below it, that is, $\sigma\eta \in T$ for all $\eta \in \{0, 1\}^m$. At the other extreme, we say that T is *fully concentrated for m generations at σ* if σ has a single descendant m generations down, that is, there is a unique $\eta \in \{0, 1\}^m$ with $\sigma\eta \in T$.

Let us return to the example A_n from Sect. 2.4 and examine the associated tree T_n of height n^2. For every $i < n$, every node at level i^2 has full branching for i generations; and every node at level $i^2 + i$ is fully concentrated for $i + 1$ generations. Consequently, for every $j \in [i^2, i^2 + i)$ every node of level j has full branching for one generation; for $j \in [i^2 + i, (i+1)^2)$, every node at level j is fully concentrated for one generation. We also have the following statement: For every m we can partition the levels $0, 1, \ldots, n^2$ into three sets U, V, W, such that (a) For every $i \in U$, every level-i node has full branching for m generations; (b) For every $j \in V$, every level-j node is fully concentrated for m generations; and (c) W is a negligible fraction of the levels, specifically $|W|/n^2 = o(1)$ as $n \to \infty$ (with m fixed). Of course, $U = \bigcup_{i>m}[i^2, i^2 + i - m)$, $V = \bigcup_{i>m}[i^2 + i, (i + 1)^2 - m)$, and W is the set of remaining levels. This is pictured schematically in Fig. 1.

Does this picture hold in general when $|A + A| \le |A|^{1+\delta}$? Certainly not exactly, since we can always pass to a subset $A' \subseteq A$ of size $|A'| \ge |A|^{1-\delta}$ and get a set with similar doubling behavior (for a constant loss in δ), but much less structure. One can also perturb it in other ways. However, in a looser sense, the picture above is quite general. One approach is to pass to a subtree of reasonably large relative size. Such an approach was taken by Bourgain in [Bou03, Bou10]. The approach taken in [Hoc12] is more statistical, and in a sense it gives a description of the entire tree, but requires us to weaken the notion of concentration. Given $\delta > 0$, we say that T is

δ-*concentrated* for m generations at $\sigma \in T$ if there exists $\eta \in \{0, 1\}^m$ such that

$$\mu_T(\sigma\eta) \geq (1 - \delta)\mu_T(\sigma)$$

where μ_T is the tree-measure associated to T. In other words, T is δ-concentrated at σ if it is possible to remove an δ-fraction of the leaves descended from σ in such a way that the resulting tree becomes fully concentrated for m generations at σ. Note that this definition is not purely local, since it depends not only on the depth-m subtree of T rooted at σ, but on the entire subtree rooted at σ, since the weights on $S = \{\sigma\eta \ : \ \eta \in \{0, 1\}^m\}$ are determined by the number of leaves of T, not by S itself.

Theorem 2.2 *For every $\varepsilon > 0$ and $m > 1$, there is a $\delta > 0$ such that for all sufficiently small $\rho > 0$ the following holds. Let $X \subseteq [0, 1]$ be a finite set such that*

$$N_\rho(X + X) \leq N_\rho(X)^{1+\delta}$$

and let T be the associated tree of height $h = \lceil\log(1/\rho)\rceil$. Then the levels $0, 1, \ldots, h$ can be partitioned into sets U, V, W such that

1. *For every $i \in U$,*

   $$\mathbb{P}_{\sigma \sim T}(T \text{ has full branching at } \sigma \text{ for } m \text{ generations} \mid \sigma \text{ is in level } i) > 1 - \varepsilon.$$

2. *For every $j \in V$,*

   $$\mathbb{P}_{\sigma \sim T}(T \text{ is } \varepsilon\text{-concentrated at } \sigma \text{ for } m \text{ generations} \mid \sigma \text{ is in level } j) > 1 - \varepsilon.$$

3. *$|W| < \varepsilon h$.*

Note that if X is ρ-separated, the hypothesis is essentially the same as $|X + X| \leq |X|^{1+\varepsilon}$.

Our analysis of self-similar sets requires the following asymmetric variant, which is easily seen to imply the symmetric one above. To motivate it, note that $|A + B| \leq C|A|$ can occur for two trivial reasons: One is that $A = \{1, \ldots, n\}$ for some n and $B \subseteq \{1, \ldots, n\}$ is arbitrary. The second is that $B = \{b\}$, a singleton, and A is arbitrary. The following theorem says that when $|A + B| \leq |A|^{1+\delta}$ then there are essentially two kinds of scales: those where, locally, the sets A, B look like in the first trivial case, and those where, locally, A, B look like the second trivial case. See Fig. 2.

Theorem 2.3 *For every $\varepsilon > 0$ and $m > 1$, there is a $\delta > 0$ such that for all sufficiently small $\rho > 0$ the following holds. Let $X, Y \subseteq [0, 1]$ be finite sets such that*

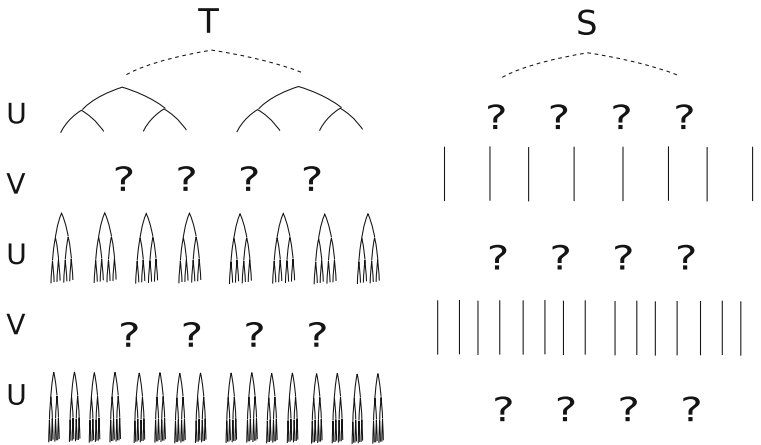$$N_\rho(X + Y) \leq N_\rho(X)^{1+\delta}$$

Fig. 2 Schematic representation of the conclusion of Theorem 2.3 (with W indicated by the small space between levels)

Let T, S be the associated trees of height $h = \lceil \log(1/\rho) \rceil$, respectively. Then the levels $0, 1, \dots, h$ can be partitioned into sets U, V, W such that

1. For every $i \in U$,

$$\mathbb{P}_{\sigma \sim T}(T \text{ has full branching at } \sigma \text{ for } m \text{ generations} \mid \sigma \text{ is in level } i) > 1 - \varepsilon$$

   (but we know nothing about S at level i).
2. For every $j \in V$,

$$\mathbb{P}_{\sigma \sim S}(S \text{ is } \varepsilon\text{-concentrated at } \sigma \text{ for } m \text{ generations} \mid \sigma \text{ is in level } j) > 1 - \varepsilon$$

   (but we know nothing about T at level j).
3. $|W| < \varepsilon h$.

The theorems above follow from [Hoc12, Theorems 2.7 and 2.9], using the fact that high enough entropy at a given scale implies full branching, and small enough entropy at a given scale implies δ-concentration.

3 A Conceptual Proof of Theorem 1.2

In this section we give a heuristic proof of Theorem 1.2. We begin with some general observations about self-similar sets. Then we explain how the theorem is reduced to a statement about sumset growth. Finally, we demonstrate how the inverse theorems of the previous section are applied.

From now on let $\Phi = \{f_i\}_{i \in \Lambda}$ be an IFS with attractor X, as in the introduction. We assume that $0 \in X \subseteq [0, 1)$; this can always be achieved by a change of coordinates, which does not affect the statement of Theorem 1.2.

3.1 Sumset Structure of Self-similar Sets

Our analysis will focus on finite approximations of X. Define the n-th approximations by

$$X_n = \{f_{\underline{i}}(0) : \underline{i} \in \Lambda^n\}$$

Clearly $X_n \subseteq X$. Also note that $|X_n| \leq |\Lambda|^n$, with a strict inequality for some n if and only if exact overlaps occur. Self similarity enters our argument via the following lemma.

Lemma 3.1 *For any $m, n \in \mathbb{N}$,*

$$X = X_m + r^m X \tag{10}$$

$$X_{m+n} = X_m + r^m X_n \tag{11}$$

Proof By (2) and (6),

$$X = \bigcup_{\underline{i} \in \Lambda^n} f_{\underline{i}}(X)$$

$$= \bigcup_{\underline{i} \in \Lambda^n} \{f_{\underline{i}}(0) + r^m x : x \in X\}$$

$$= X_m + r^m X$$

which is the first identity. To prove the second, for $\underline{i} \in \Lambda^m$ and $\underline{j} \in \Lambda^n$ denote by $\underline{i}\,\underline{j}$ their concatenation. By (5),

$$f_{\underline{i}\,\underline{j}}(0) = \sum_{k=1}^{m} a_{i_k} r^{k-1} + r^m \sum_{k=1}^{n} a_{j_k} r^{k-1}$$

$$= f_{\underline{i}}(0) + r^m f_{\underline{j}}(0)$$

hence

$$X_{m+n} = \{f_{\underline{i}\,\underline{j}}(0) : \underline{i}\,\underline{j} \in \Lambda^{m+n}\}$$

$$= \{f_{\underline{i}}(0) + r^m f_{\underline{j}}(0) : \underline{i} \in \Lambda^m, \underline{j} \in \Lambda^n\}$$

$$= X_m + r^m X_n \qquad \qquad \square$$

Let us demonstrate the usefulness of this lemma by showing that $\dim_B(X)$ exists. First, since $r^m X$ is of diameter $\leq r^m$, it is easy to deduce from (10) that $N_{r^n}(X_n)$, $N_{r^n}(X)$ differ by at most a factor of 2. Thus the existence of $\dim_B X$ is equivalent to existence of the limit $\frac{1}{m} \log N_{r^m}(X_m)$ as $n \to \infty$. Next, we have a combinatorial lemma.

Lemma 3.2 *Let $A, B \subseteq \mathbb{R}$ with $B \subseteq [0, \varepsilon)$. Then for any $\gamma < \varepsilon$,*

$$N_\gamma(A + B) \geq \frac{1}{3} \cdot N_\varepsilon(A) \cdot N_\gamma(B)$$

Proof Let $\mathcal{I} = \{I_i\}_{i=1}^{N_\varepsilon(A)}$ be an optimal cover of A by disjoint intervals of length ε. Let $\mathcal{J} = \{J_j\}_{j=1}^{N_\gamma(A+B)}$ be an optimal cover of $A + B$ by intervals of length γ. For each $I_i \in \mathcal{I}$ fix a point $a_i \in A \cap I_i$ and note that $a_i + B \subseteq A + B$ is covered by $\mathcal{J}$, so $a_i + B$ intersects at least $N_\gamma(B)$ intervals in $\mathcal{J}$. If each interval J_j intersects a unique translate $a_i + B$, we would conclude that $N_\gamma(A+B) \geq N_\varepsilon(A) N_\gamma(B)$. While a_i may not be unique, we can argue as follows: Since $B \subseteq [0, \varepsilon)$, if $J_j = [u, u+\varepsilon]$ and intersects $a + B$ for some $a \in A$, then $a \in [u - \varepsilon, u + 2\varepsilon)$. Since the intervals I_i are disjoint and of length ε, there are most 3 intervals $I_i \in \mathcal{I}$ that a could belong to. The claim follows. $\square$

Since $X \subseteq [0, 1)$ we have $r^m X_n \subseteq [0, r^m)$, so by the lemma,

$$\begin{aligned}
N_{r^{m+n}}(X_{m+n}) &= N_{r^{m+n}}(X_m + r^m X_n) \\
&\geq \frac{1}{3} \cdot N_{r^m}(X_m) \cdot N_{r^{m+n}}(r^m X_n) \\
&= \frac{1}{3} \cdot N_{r^m}(X_m) \cdot N_{r^n}(X_n)
\end{aligned}$$

where in the last equality we used the identity $N_{t\varepsilon}(tZ) = N_\varepsilon(Z)$. Taking logarithms, this shows that the sequence $s_n = \log N_{r^m}(X_m)$ is approximately super-additive in the sense that $s_{m+n} \geq s_m + s_n - C$ for a constant C. The existence of the limit of $\frac{1}{n} s_n$ as $n \to \infty$ is then well known (perhaps it is better known when $C = 0$ and s_n is (really) super-additive. The proof for $C = 0$ works also in the $C > 0$ case; alternatively, note that $s_n' = s_n - \log n$ becomes super-additive after excluding finitely many terms, so $\lim \frac{1}{n} s_n'$ exists, and $\frac{1}{n} s_n' - \frac{1}{n} s_n \to 0$).

3.2 From Theorem 1.2 to Additive Combinatorics

Let us return to our main objective, Theorem 1.2. Continuing with the previous notation, write

$$\begin{aligned}
\alpha &= \dim_B X \\
\beta &= \min\{1, \operatorname{sdim} X\}
\end{aligned}$$

and suppose, by way of contradiction, that $\alpha < \beta$ and that for some $k \in \mathbb{N}$ we have $\Delta_n \geq 2^{-kn}$ for all n (in particular, there are no exact overlaps). We make a number of observations. The first is rather trivial: that "too small" dimension means that there are intervals of length r^m containing exponentially many points from X_m. Precisely,

Proposition 3.3 *Let $\sigma = \frac{1}{2}(\beta - \alpha) > 0$. Then for every large enough m, there is an interval I_m of length r^m such that $|X_m \cap I_m| > r^{-\sigma m}$.*

Proof As we have already noted, $\frac{1}{m \log(1/r)} \log N_{r^m}(X_m) \to \alpha$ as $m \to \infty$. Thus for large enough m,

$$N_{r^m}(X_m) < r^{-(\beta-\sigma)m}$$

On the other hand, since there are no exact overlaps,

$$|X_m| = |\Lambda|^m = r^{-m \dim(X)} \geq r^{-m\beta}$$

Thus in an optimal cover of X_m by r^m-intervals, at least one must contain $|X_m|/N_{r^m}(X_m) \geq (1/r)^{\sigma m}$ points. □

We now wish to extract more information from the sumset identity $X_{m+n} = X_m + r^m X_n$. In itself it provides limited information about the covering number $N_{m+n}(X_n)$, since the summands live at different scales. This is what was used earlier in proving super-additivity of $s_n = \log N_{r^m}(X_m)$. The next step is to localize the sumset relation.

Proposition 3.4 *For all $\delta > 0$, for all large m there exists an interval J_m of length r^m such that $X_m \cap J_m \neq \emptyset$ and, writing $n = km$,*

$$N_{r^{m+n}}((X_m \cap J_m) + r^m X_n) < r^{-(1+\delta)\alpha n} \tag{12}$$

Proof Fix m, set $n = km$, and let $\mathcal{J}$ denote the partition of $\mathbb{R}$ into intervals $[ur^m, (u+1)r^m)$, $u \in \mathbb{Z}$, whose lengths are r^m. Since $X_m = \bigcup_{J \in \mathcal{J}}(X_m \cap J)$, we can re-write (11) as

$$
\begin{aligned}
X_{m+n} &= X_m + r^m X_n \\
&= \bigcup_{J \in \mathcal{J}}((X_m \cap J) + r^m X_n)
\end{aligned} \tag{13}
$$

Since $X_m \cap J \subseteq [ur^m, (u+1)r^m)$ for some u and $r^m X_n \subseteq [0, r^m)$, we have $(X_m \cap J) + r^m X_n \subseteq [ur^m, (u+2)r^m)$ and in particular each set in the union (13) is of diameter $\leq 2r^m$. On the other hand, no interval of length r^{m+n} intersects more than three of the sets $[ur^m, (u+2)r^m)$. Therefore, arguing as in the proof of Lemma (3.2),

$$N_{r^{m+n}}(X_{m+n}) \geq \frac{1}{3} \cdot N_{r^m}(X_m) \cdot \min_{J \in \mathcal{J}\,:\,X_m \cap J \neq \emptyset} N_{r^{m+n}}((X_{m+n} \cap J) + r^m X_n)$$

so

$$\min_{J \in \mathcal{J} \,:\, X_m \cap J \neq \emptyset} N_{r^{m+n}}((X_{m+n} \cap J) + r^m X_n) \leq 3 \cdot \frac{N_{r^{m+n}}(X_{m+n})}{N_{r^m}(X_m)}$$

$$\leq 3 \cdot \frac{r^{-(\alpha + o(1))(m+n)}}{r^{-(\alpha + o(1))m}}$$

$$= r^{-(\alpha + o(1))n} \qquad \text{as } m \to \infty$$

The proposition follows.   □

Now suppose that it so happens that, for large m, the propositions above produce the *same* interval: $I_m = J_m$. Then we would have the following:

Proposition 3.5 *Suppose that* $\dim X < \min\{1, sdim X\}$ *and* $\Delta_n \geq 2^{-kn}$ *for all n. Then there is a constant* $\tau > 0$ *such that, for every* $\delta > 0$ *and all suitably large n, there is a subset* $Y_n \subseteq [0, 1]$ *with*

$$N_{r^n}(Y_n) \geq 2^{\tau n} \tag{14}$$

$$N_{r^n}(X_n + Y_n) \leq N_{r^n}(X_n)^{1+\delta} \tag{15}$$

Proof Let σ be as in Proposition 3.3 and take $\tau = \sigma/(k \log(1/r))$. As before write $n = (k+1)m$, and assume that the intervals I_m, J_m provided by the two previous propositions coincide for arbitrarily large m: $I_m = J_m = [a_m, b_m)$. Let

$$Y_m = r^{-m}(X_m \cap I_m - a_m)$$

and note that $Y_m \subseteq [0, 1)$. Now, by choice of I_m we know that $|X_m \cap I_m| \geq r^{-\sigma m}$, and since $\Delta_m \geq 2^{-km} = 2^{-n}$, we know that every two points in $X_m \cap I_m$ are separated by at least 2^{-n}. Therefore,

$$N_{r^n}(Y_m) = N_{r^{m+n}}(X_m \cap I_m)$$

$$\geq r^{-\sigma m}$$

Using the identity $N_{t\varepsilon}(tZ) = N_\varepsilon(Z)$ with $t = r^m$ and $Z = Y_m$, we conclude that

$$N_{r^{-n}}(Y_m) \geq r^{-\sigma m} = r^{\tau n}$$

Similarly, since $X_n + Y_m = r^{-m}((X_m \cap I_m) + r^m X_n)$, from the definition of J_m and the identity $N_{t\varepsilon}(tZ) = N_\varepsilon(Z)$ again, we find that for large enough n (equivalently, m),

$$N_{r^n}(X_n + Y_n) = N_{r^{m+n}}((X_m \cap I_m) + r^m X_n)$$

$$\leq r^{-(1+\delta)\alpha n}$$

$$\leq N_{r^n}(X_n)^{(1+2\delta)}$$

where in the last inequality we again used the fact that $N_{r^m}(X_m) \sim r^{-n\alpha}$.   □

The task of showing that the conclusion of the "Proposition" is impossible falls within the scope of additive combinatorics. Heuristically, it cannot happen because, being a fractal, X_n has very little "additive structure". This intuition is correct, as we discuss in the next section.

But can one really ensure that I_m, J_m coincide? A natural attempt would be to show that, for a fixed optimal r^m-cover of X_m, "most" intervals of length r^m can play each of the roles, and hence a positive fraction can play both. In fact, for every $\eta > 0$, for large m at least a $(1 - \eta)$-proportion of these intervals will be a good choice for J_m. Unfortunately, although the number of candidates for I_m can be shown to be exponential in m, it could still be exponentially small compared to $N_{r^m}(X_m)$, and so we cannot conclude that the two families of "good" intervals have members in common. It is possible that more sophisticated counting can make this work, but the approach that is currently simplest is to replace covering numbers by the entropy, at an appropriate scale, of the uniform measure on X_m. We return to this in Sect. 4.

3.3 Getting a Contradiction

Our goal now is to demonstrate that the conclusion of "Proposition" 3.5 is impossible. The argument we give again falls short of this goal, but it gives the essential ideas of the proof. Thus, we ask the reader to suspend his disbelief a little longer.

Let $\tau > 0$ be as given in "Proposition" 3.5. Choose a very small parameter $\varepsilon > 0$ which we shall later assume is small compared to τ. Choose m large enough that

$$N_{r^m}(X_m) \geq r^{-m(1-\varepsilon)\alpha}$$

Apply the inverse Theorem 2.3 with parameters ε, m and obtain the promised $\delta > 0$. From "Proposition" 3.5 obtain the corresponding $Y_n \subseteq [0, 1)$ satisfying (14) and (15).

Write T^n for the tree of height $h_n = [1/r^n]$ associated to X_n and S^n for the tree of the same height associated to Y_n. From our choice of δ and (15), by the inverse theorem there is a partition $U_n \cup V_n \cup W_n$ of $\{1, \ldots, h_n\}$ such that

(I) At scales $i \in U_n$, a $1 - \varepsilon$ fraction of nodes of T^n at level i have full branching for m-generations.
(II) At scales $j \in V_n$, a $1 - \varepsilon$ fraction of nodes of S^n at level j are ε-concentrated for m generations.
(III) $|W_n| \leq \varepsilon h_n$.

Our first task is to show that V_n is not too large. It is quite clear (or at least believable) that if a tree has few nodes with more than one child, then it can have only an exponentially small number of leaves. The same is true if we only assume, for a small $\lambda > 0$, that most nodes are λ-concentrated. More precisely,

Lemma 3.6 *Let S be a tree of height h, let $\lambda > 0$ and $\ell \geq 1$. Suppose that*

$$\mathbb{P}_S(\sigma \in S : S \text{ is } \lambda\text{-concentrated at } \sigma \text{ for } \ell \text{ generations}\} > 1 - \lambda$$

Then $|\partial S| \leq 2^{\lambda' \cdot h}$ where $\lambda' \to 0$ as $\lambda \to 0$ and $h/\ell \to \infty$.

We leave the proof to the motivated reader. We note that this lemma is superseded by Proposition 4.5, which gives a stronger statement and has a simpler proof.

We apply the lemma to $S = S^n$ with $\ell = m$. Choose λ small enough that $\lambda' < \tau$ for large n (hence large h_n). Thus λ depends only on τ and we may assume that at the start we chose $\varepsilon < \frac{1}{2}\lambda$. Suppose that we had $|V_n| > (1 - \lambda/2)h_n$. Since in each level $j \in V_n$ a $(1 - \varepsilon)$-fraction of the nodes (with respect to the tree measure μ_{S^n}) is ε-concentrated, at least the same fraction is λ-concentrated, and we would conclude

$$\mathbb{P}_{S^n}(\sigma \in S^n : S^n \text{ is } \lambda\text{-concentrated at } \sigma \text{ for } m \text{ generations}) \geq \frac{1}{h_n}|V_n| \cdot (1 - \varepsilon)$$
$$> (1 - \frac{\lambda}{2})(1 - \varepsilon)$$
$$> 1 - \lambda$$

From the lemma we would have $N_{r^n}(Y_n) \leq |\partial S^n| < 2^{\lambda' h_n} < 2^{\tau h_n}$, contradicting (14). Thus, we conclude that

$$|V_n| < (1 - \frac{\lambda}{2})h_n$$

Consequently, assuming as we may that $\varepsilon < \lambda/6$,

$$|U_n| = h_n - |V_n| - |W_n| \geq (\frac{\lambda}{2} - \varepsilon)h_n > \frac{\lambda}{3}h_n \tag{16}$$

So far we have seen that U_n consists of a positive fraction of the levels of T^n, and hence a positive fraction of nodes in T^n have full branching for m generations. Our next task will be to show that most of the remaining nodes have roughly $r^{-\alpha m}$ descendants m generations down. This is where we use self-similarity again in an essential way.

Proposition 3.7 *If m is large enough, then for all large enough n,*

$$\mathbb{P}_{\sigma \sim \mu_{T^n}}\left(\begin{array}{c} \sigma \text{ has } \geq 2^{(1-\varepsilon)\alpha m} \text{ descendants} \\ m \text{ generations down in } T \end{array}\right) > 1 - \varepsilon$$

Proof (sketch.) A node $\sigma \in T^n$ of level ℓ corresponds to an interval $I = [\frac{u}{2^\ell}, \frac{u+1}{2^\ell})$. We call such intervals level-ℓ intervals, and recall that the probability induced from μ_{T^n} on level-ℓ intervals is just proportional to $|I \cap X_n|$. The claim is then that if we choose $0 \leq \ell \leq h_n$ uniformly and then choose a level-ℓ interval I at random,

then with probability at least $1 - \varepsilon$, we will have $N_{2^{-\ell-m}}(I \cap X_n) \geq 2^{-(1-\varepsilon)\alpha m}$. In order to prove this, it is enough to show that for all levels $0 \leq \ell \leq (1 - \frac{\varepsilon}{2})h_n$, if we choose a level-$\ell$ interval I at random, then with probability at least $1 - \frac{\varepsilon}{2}$ we have $N_{2^{-\ell-m}}(I \cap X_n) \geq 2^{-(1-\varepsilon)\alpha m}$.

Fix a parameter m_0 depending on ε and assume m, n large with respect to it. Observe that X_n decomposes into a union of copies of $X_{n'}$ scaled by approximately $2^{-\ell-m_0}$. More precisely, choosing $u \in \mathbb{N}$ such that $r^u \approx 2^{-\ell-m_0}$, by (11) we have

$$X = X_u + r^u X_{n-u} = \bigcup_{x \in X_u} (x + r^u X_{n-u})$$

The idea is now the following. The translates $x + r^u X_{n-u}$ in the union are of diameter $r^u \approx 2^{-\ell}/2^{m_0}$, which is much smaller than $2^{-\ell}$, and hence with probability at least $1 - \frac{\varepsilon}{2}$ a level-ℓ interval I will contain an entire translate $x + r^u X_{n-u}$ from the union above. The details of the proof are somewhat tedious and we omit them. The point is that, if $x + r^u X_{n-u} \subseteq I$, and assuming that m is large enough relative to ε, m_0, we have

$$\begin{aligned}
N_{2^{-\ell-m}}(I \cap X_n) &\geq N_{2^{-\ell-m}}(x + r^u X_{n-u}) \\
&= N_{2^{-\ell-m}r^{-u}}(X_{n-u}) \\
&\approx N_{2^{-(m-m_0)}}(X_{n-u}) \\
&> 2^{(1-\varepsilon)\alpha m}
\end{aligned}$$

which is what we wanted to prove.   □

Now that we know that most nodes in T^n have many descendants, and a positive fraction have the maximal possible number of descendants, m generations down, the last ingredient we need is a way to use this information to get a lower bound on the number of leaves in T^n. Heuristically, this is the analog of the upper bound we had in Lemma 3.6.

Proposition 3.8 *Let T be a tree of height h, let $m \geq 0$, and suppose that the nodes of T can be partitioned into disjoint sets $A_1, \ldots, A_\ell$ such that each node $\sigma \in A_i$ has $2^{c_i m}$ descendants m generations down. Write $p_i = \mathbb{P}_{\mu_T}(A_i)$. Then*

$$|\partial T| \geq \prod_{i=1}^{\ell} 2^{c_i \cdot p_i h}$$

This "Proposition" is, unfortunately, incorrect, and the reader may find it instructive to look for a counterexample. The statement could be fixed if we made stronger assumptions than just bounding the branching in each of the sets A_i, but the resulting argument would almost certainly be more complicated than the proof in [Hoc12], and we do not pursue it. The correct statement is given in Proposition 4.5 below.

We can now put the pieces together. By the defining property (I) of U_n and equation (16), the set $A_1^n \subseteq T^n$ of nodes with full branching for m-generations satisfies

$$\mathbb{P}_{T^n}(A_1^n) \geq \frac{1}{h_n}|U_n| \cdot (1 - \varepsilon)$$
$$\geq \frac{\lambda}{3}(1 - \varepsilon)$$
$$\geq \frac{\lambda}{4}$$

assuming again ε small compared to λ (equivalently τ). Let A_2^n denote the set of nodes of $T^n \setminus A_1^n$ which do *not* have at least $2^{m(1-2\varepsilon)\dim X}$ descendants m generations down; by Proposition 3.7,

$$\mathbb{P}_{T^n}(A_2^n) < \varepsilon$$

Therefore if we define $A_3^n = T^n \setminus \{A_1^n \cup A_2^n\}$ then all nodes in A_3^n have at least $2^{m(1-2\varepsilon)\dim X}$ descendants m generations down and

$$\mathbb{P}_{T^n}(A_3^n) = 1 - \mathbb{P}_{T^n}(A_1^n) - \mathbb{P}_{T^n}(A_2^n)$$

In the terminology of the "Proposition", we have $p_1 \geq \lambda/4$ and $p_2 < \varepsilon$, hence $p_3 \geq 1 - p_1 - \varepsilon$. Also $c_1 = 1$, $c_3 = (1 - 2\varepsilon)\dim X$ and by default $c_2 \geq 0$. From the "Proposition" we find that

$$|\partial T^n| \geq 2^{p_1 h_n} \cdot 2^{p_2 \cdot c_2 h_n} \cdot 2^{p_3(1-2\varepsilon)\dim X \cdot h_n}$$
$$\geq 2^{p_1 h_n + (1-2\varepsilon)\dim X \cdot (1-p_1-\varepsilon)h_n}$$
$$\geq 2^{(\dim X + \varepsilon)h_n}$$

where in the last inequality we assumed that ε is small compared to p_1 and $\dim X$. Since $N_{r^n}(X) = |\partial T^n|^{1+o(1)}$ as $n \to \infty$, this contradicts the definition of $\dim X$.

3.4 Sums with Self-similar Sets

What we "proved" above is the following statement which is of independent interest, and is, moreover, true (a proof follows easily from the methods of [Hoc12]).

Theorem 3.9 *For every any self-similar set X with $\dim X < 1$ and every $\tau > 0$ there is a $\delta > 0$ such that for all small enough $\rho > 0$ and any set $Y \subseteq \mathbb{R}$,*

$$N_\rho(Y) > (1/\rho)^\tau \qquad \Longrightarrow \qquad N_\rho(X + Y) > N_\rho(X)^{1+\delta}$$

There is also a fractal version for Hausdorff dimension:

Theorem 3.10 *For every any self-similar set X with $\dim X < 1$ and every $\tau > 0$ there is a $\delta > 0$ such that for any set $Y \subseteq \mathbb{R}$,*

$$\dim Y > \tau \quad \implies \quad \dim(X + Y) > \dim X + \delta$$

For box dimension (lower or upper) the analogous statement follows directly from the previous theorem. The version for Hausdorff dimension requires slightly more effort and will appear in [Hoc13] along with the analog for measures.

4 Entropy

In this final section we discuss how to turn the outline above into a valid proof. The main change is to replace sets by measures and covering numbers by entropy. Each of the three parts of the argument (inverse theorem, reduction to a statement about sumsets, and the analysis of the sums) has an entropy analog which we indicate below, along with a reference to the relevant part of [Hoc12].

The reader should note that the outline given below is designed to match as closely as possible the argument from the previous section, rather than the proof from [Hoc12]. Although the ideas and many of the details are the same, the original proof is direct, whereas the one here is by contradiction. For this reason not all of the statements below have exact analogs in [Hoc12].

4.1 Entropy

We assume that the reader is familiar with the basic properties of Shannon entropy, see for example [CT06]. Let $\mathcal{I}_\varepsilon = \{[k\varepsilon, (k+1)\varepsilon)\}_{k \in \mathbb{Z}}$, which is a partition of $\mathbb{R}$ into intervals of length ε. The entropy $H(\mu, \mathcal{I}_\varepsilon)$ of μ at scale ε is the natural measure-analog of the covering number $N_\varepsilon(X)$, albeit in a logarithmic scale. For a measure ν supported on a set X, the two quantities are related by the basic inequality

$$0 \le H(\nu, \mathcal{I}_\varepsilon) \le \log \#\{I \in \mathcal{I}_\varepsilon \,:\, X \cap I \ne \emptyset\} \le \log N_\varepsilon(X) + O(1)$$

(the $O(1)$ error is because we are choosing a sub-cover of X from a fixed cover of $\mathbb{R}$ rather than allowing arbitrary ε-intervals). We introduce the normalized ε-scale entropy:

$$H_\varepsilon(\nu) = \frac{1}{\log(1/\varepsilon)} H(\nu, \mathcal{I}_\varepsilon)$$

Thus, for ν supported on a set X with well-defined box dimension, the previous inequality implies

$$\limsup_{\varepsilon \to 0} H_\varepsilon(\nu) \le \dim_B X \tag{17}$$

4.2 Inverse Theorems for Entropy

The measure-analog of the sumset operation is convolution, which for discrete probability measures $\mu = \sum p_i \delta_{x_i}$ and $\nu = \sum q_j \delta_{y_j}$ is[6]

$$\mu * \nu = \sum_{i,j} p_i q_j \delta_{x_i + y_j}$$

The entropy-analog of the small doubling condition $|A + A| \le C|A|$ is the inequality $H(\mu * \mu) \le H(\mu) + C'$, where $H(\mu)$ is the entropy of a measure with respect to the partition into points (remember that entropy is like cardinality, but in logarithmic scale). Alternatively we could discretize at scale ε, giving $H_\rho(\mu*\mu) \le H_\rho(\mu) + O(1/\log(1/\varepsilon))$. Tao [Tao10] has shown that such inequalities have implications similar to Freiman's theorem. Related results were also obtained by Madiman [Mad08], see also [MMT12].

The regime that interests us is, as before, the analog of $|A + B| \le |A|^{1+\delta}$, which by formal analogy takes the form $H_\rho(\mu * \nu) \le (1 + \delta)H_\rho(\mu)$. When μ is supported on $[0, 1]$ we have $H_\rho(\mu) \le 1 + o(1)$ as $\rho \to 0$, and this inequality is implied (and in the cases that interest us essentially equivalent to)

$$H_\rho(\mu * \nu) \le H_\rho(\mu) + \delta \tag{18}$$

Before stating the inverse theorem for entropy we need a few more definitions. Consider the lift of μ to a tree-measure $\widetilde{\mu}$ on the full binary tree of height h (see Sect. 2.5). Given a node $\sigma = \sigma_1 \dots \sigma_k$ and $m \in \mathbb{N}$, write $\sigma\{0, 1\}^m$ for the set of descendants of σ m-generations down. Let $\widetilde{\mu}_{\sigma,m}$ denote the probability measure on $\sigma\{0, 1\}^m$ that assigns to each node its normalized weight according to $\widetilde{\mu}$. Since $\sum_{\eta \in \sigma\{0,1\}^m} \widetilde{\mu}(\eta) = \widetilde{\mu}(\sigma)$, this measure is given by $\widetilde{\mu}_{\sigma,m}(\eta) = \widetilde{\mu}(\eta)/\widetilde{\mu}(\sigma)$.

We say that $\widetilde{\mu}$ is δ-concentrated at σ for m generations if $H(\widetilde{\mu}_{\sigma,m}) < \delta$, that is, if

$$-\frac{1}{m} \sum_{\eta \in \sigma\{0,1\}^m} \frac{\widetilde{\mu}(\eta)}{\widetilde{\mu}(\sigma)} \log \frac{\widetilde{\mu}(\eta)}{\widetilde{\mu}(\sigma)} < \delta.$$

---

[6] In general there is a similar formula: $\mu*\nu = \int\int \delta_{x+y} d\mu(x)d\nu(y)$, where the integral is interpreted as a measure by integrating against Borel functions.

For a tree measure μ_T associated to a tree T and for fixed m, this is equivalent to T being δ'-concentrated for m generations at σ for an appropriate δ' which tends to 0 together with δ. We say that $\widetilde{\mu}$ is δ-uniform at σ for m generations if $H(\widetilde{\mu}_{\sigma,m}) > \log m - \delta$. Note that for m fixed, when δ is small enough this implies that $\widetilde{\mu}(\eta) > 0$ for all $\eta \in \sigma\{0, 1\}^m$, so this indeed generalizes full branching. We can now state the inverse theorem:

Theorem 4.1 (Theorem 2.7 of [Hoc12]) *For every $\varepsilon > 0$ and $m \geq 1$, there is a $\delta > 0$ such that for sufficiently small $\rho > 0$ the following holds. Let μ, ν be probability measures on $[0, 1]$ and suppose that*

$$H_\rho(\mu * \nu) \leq H_\rho(\mu) + \delta$$

Let $\widetilde{\mu}, \widetilde{\nu}$ denote the lifts of μ, ν to the full binary trees of height $h = \lceil \log_2(1/\rho) \rceil$. Then there is a partition of the levels $\{0, \dots, h\}$ into three sets $U \cup V \cup W$ such that

1. *For $i \in U$,*

    $$\mathbb{P}_{\sigma \sim \widetilde{\mu}}(\widetilde{\mu} \text{ is } \varepsilon\text{-uniform at } \sigma \text{ for } m \text{ generations} \mid \sigma \text{ is in level } i) > 1 - \varepsilon$$

2. *For $j \in V$,*

    $$\mathbb{P}_{\sigma \sim \widetilde{\mu}}(\widetilde{\nu} \text{ is } \varepsilon\text{-concentrated at } \sigma \text{ for } m \text{ generations} \mid \sigma \text{ is in level } i) > 1 - \varepsilon$$

3. $|W| < \delta n.$

4.3 Reduction of Theorem 1.2 to a Convolution Inequality

We return to our IFS Φ with attractor $0 \in X \subseteq [0, 1]$, as in Sect. 3. Define measures $\mu^{(n)}$ analogous to X_n by

$$\mu^{(n)} = \frac{1}{|\Lambda|^n} \sum_{\underline{i} \in \Lambda^n} \delta_{f_{\underline{i}}(0)}$$

Write $S_t \mu(A) = \mu(t^{-1} A)$ (this is the usual push-forward of μ by S_t). Then the analog of the sumset relation $X_{m+n} = X_m + r^m X_n$ is

$$\mu^{(m+n)} = \mu^{(m)} * S_{r^m} \mu^{(n)}$$

The derivation is elementary, using the definition of convolution, equation (6) and the identity $S_t \delta_y = \delta_{ty}$. Next, as in Sect. 3.1, if we define $s_m = H(\mu^{(m)}, \mathcal{I}_{r^m})$ then the sequence s_n is almost super-additive in the sense that $s_{m+n} \geq s_m + s_n - O(1)$. This is proved by a similar argument to the covering number case but in the language of entropy. It follows that the limit

$$\alpha = \lim_{m \to \infty} H_{r^m}(\mu^{(m)})$$

exists. Since $\mu^{(m)}$ is supported on X, by (17) we have $\alpha \leq \dim X$.

Turning to Theorem 1.2, write

$$\beta = \min\{1, \operatorname{sdim} X\}$$

and assume for the sake of contradiction that $\dim X < \beta$ and $\Delta_n \geq 2^{-kn}$ for some k. Since $\alpha \leq \dim X$, we can choose $\varepsilon > 0$ so that $\alpha < \beta - \varepsilon$. Arguing analogously to Proposition 3.3 one obtains the analogous result:

Proposition 4.2 *There is a constant c (depending on β, ε) such that for large enough m,*

$$\mu^{(m)} \left(\bigcup \left\{ I \in \mathcal{I}_{r^m} \ : \ H_{r^{m+n}}(\mu_I^{(m)}) > cm \right\} \right) > c$$

This lemma does not appear explicitly in [Hoc12], since that is a direct proof. Ours is a proof by contradiction, and the contradiction can be interpreted as showing that the lemma above is false. This falsehood is demonstrated directly in the last displayed equation of Sect. 5.3 of [Hoc12].

Next, for a probability measure ν and set E with $\nu(E) > 0$, write ν_E for the conditional measure on E, that is, $\nu_E(A) = \frac{1}{\nu(E)} \nu(E \cap A)$. The analog of Proposition 3.4 then holds, again with an analogous proof:

Proposition 4.3 (See Equation (40) of [Hoc12]) *For every $\delta > 0$, as $m \to \infty$*

$$\mu^{(m)} \left(\bigcup \left\{ I \in \mathcal{I}_{r^m} \ : \ \frac{1}{n} H_{r^{m+n}}(\mu_I^{(m)} * S_{r^m} \mu^{(n)}) \leq \alpha + \delta \right\} \right) \geq 1 - o(1)$$

From the last two propositions one sees that for given $\delta > 0$ and large enough m, there are intervals $I = I_m \in \mathcal{I}_{r^m}$ that appear in the unions in the conclusions of both propositions. Taking ν_m to be the re-scaling of $\mu_I^{(m)}$ by r^{-m} (translated back to $[0, 1)$), we have the rigorous analog of "Proposition" 3.5:

Proposition 4.4 *There is a $\tau > 0$ such that for every $\delta > 0$, for all sufficiently large m, there is a measure ν_m supported on $[0, 1)$ with*

$$\frac{1}{m} H_{r^m}(\nu_m) > \tau \tag{19}$$

$$\frac{1}{m} H_{r^m}(\mu^{(m)} * \nu_m) < \frac{1}{m} H_{r^m}(\mu^{(m)}) + \delta \tag{20}$$

4.4 Getting a Contradiction

The missing ingredient in Sect. 3.3 was the ability to estimate the number of leaves of a tree from the average amount of branching of its nodes. This is where entropy really comes in handy, because of the following (easy!) lemma. Recall that given a tree-measure θ, we write $\theta_{\sigma,m}$ for the normalized weights on the nodes m generations down from σ.

Lemma 4.5 (Lemma 3.4 of [Hoc12]) *Let $\widetilde{\theta}$ be a tree-measure on the full binary tree T of height h. Write $\partial\widetilde{\theta}$ for the measure induced by $\widetilde{\theta}$ on the leaves of the tree. Then for any m,*

$$\frac{1}{h}H(\partial\widetilde{\theta}) = \mathbb{E}_{\sigma\sim\widetilde{\theta}}(\frac{1}{m}H(\widetilde{\theta}_{\sigma,m})) + O(\frac{m}{h})$$

From here the argument proceeds exactly as in Sect. 3.3. Let $\tau > 0$ be the constant provided by Proposition 4.4. Choose a small parameter $\varepsilon > 0$. Choose m large enough that

$$H_{r^m}(\mu^{(m)}) \geq (1-\varepsilon)\alpha$$

Apply the inverse theorem 4.1 with parameters ε, m and let $\delta > 0$ be the resulting number. Applying Proposition 4.4 with this δ, there exist probability measures ν_n on $[0, 1]$ satisfying (19) and (20). Write $\widetilde{\mu}^{(n)}$, $\widetilde{\nu}_n$ for the lift of $\mu^{(n)}$, ν_n, respectively, to the binary tree of height $h_n = \lceil 1/\log(r^n)\rceil$. By the inverse theorem there is a partition $U_n \cup V_n \cup W_n$ of the levels $\{1, \ldots, h_n\}$ such that

(I) At scales $i \in U_n$, the $\widetilde{\mu}^{(n)}$-mass of nodes at level i that are ε-uniform for m generations is at least $1 - \varepsilon$.

(II) At scales $j \in V_n$, the $\widetilde{\nu}_n$-mass of nodes at level i that are ε-concentrated for m generations is at least $1 - \varepsilon$

(III) $|W_n| \leq \varepsilon h_n$.

If $|V_n| > (1-\tau/2)h_n$ and ε is small enough compared to τ, then sufficiently many nodes (with respect to $\widetilde{\nu}_n$) would have $H(\widetilde{\nu}^n_{\sigma,m}) < \varepsilon$ that we could invoke Lemma 4.5 and conclude that the entropy $H_{r^n}(\nu_n) \approx \frac{1}{n\log(1/r)}H(\widetilde{\nu}^n) < \tau$, contradicting (19). Therefore $|V_n| \leq (1-\tau/2)h_n$. In particular, assuming ε is small enough compared to τ,

$$|U_n| \geq h_n - |V_n| - |W_n| \geq \frac{\tau}{3}h_n$$

Next, suppose that m is large enough so that $H_{r^n}(\mu^{(n)}) > (1 - \varepsilon)\alpha$. Using self-similarity of X and an argument analogous to the one outlined in Proposition 3.7, we get the analogous result:

Lemma 4.6 (Lemma 5.4 of [Hoc12]) *For all large enough n,*

$$\frac{1}{h_n + 1} \sum_{\sigma} \left\{ \widetilde{\mu}^{(n)}(\sigma) \; : \; \frac{1}{m} H(\widetilde{\mu}^{(n)}_{\sigma,m}) > (1 - 2\varepsilon)\alpha \right\} > 1 - \varepsilon$$

Now, from the definition of U_n and our bound $|U_n| \geq \frac{\tau}{3} h_n$, we know that at least a $(1 - \varepsilon)\tau/3$-fraction of the nodes of $\widetilde{\mu}^{(n)}$ satisfy $\frac{1}{m} H(\widetilde{\mu}^{(n)}_{\sigma,m}) > (1 - \varepsilon)m$. Of the remaining nodes, by the last lemma all but a ε-fraction satisfy $\frac{1}{m} H(\widetilde{\mu}^{(n)}_{\sigma,m}) \geq (1 - 2\varepsilon)\alpha$. Therefore by Lemma 4.5 again, for all large enough n,

$$H_{r^n}(\mu^{(n)}) \approx \frac{1}{h_n} H(\widetilde{\mu}^{(n)}) > (1 - \varepsilon)^2 \frac{\tau}{3} + (1 - (1 - \varepsilon)\frac{\tau}{3} - \varepsilon)\alpha > \alpha + \varepsilon$$

assuming ε is small compared to τ. This contradicts the definition of α.

References

[Bou03]   Bourgain, J.: On the Erdős-Volkmann and Katz-Tao ring conjectures. Geom. Funct. Anal. **13**(2), 334–365 (2003)

[Bou10]   Bourgain, J.: The discretized sum-product and projection theorems. J. Anal. Math. **112**, 193–236 (2010)

[CT06]   Thomas, M.C., Joy A.T.: Elements of Information Theory. 2nd edn. Wiley, Hoboken (2006)

[EV66]   Erdős, P., Volkmann, B.: Additive Gruppen mit vorgegebener Hausdorffscher Dimension. J. Reine Angew. Math. **221**, 203–208 (1966)

[Fal89]   Falconer, K.J.: Dimensions and measures of quasi self-similar sets. Proc. Amer. Math. Soc. **106**(2), 543–554 (1989)

[Gar02]   Gardner, R.J.: The Brunn-Minkowski inequality. Bull. Amer. Math. Soc. (N.S.), **39**(3), 355–405 (2002)

[Gar62]   Garsia, A.M.: Arithmetic properties of Bernoulli convolutions. Trans. Amer. Math. Soc. **102**, 409–432 (1962)

[Hoc12]   Michael H.: Self-similar sets with overlaps and sumset phenomena for entropy. preprint (2012)

[Hoc13]   Michael H.: Hausdorff dimension of sums and convolutions. in preparation (2013)

[Hut81]   Hutchinson, J.E.: Fractals and self-similarity. Indiana Univ. Math. J. **30**(5), 713–747 (1981)

[Ken97]   Kenyon, R.: Projecting the one-dimensional Sierpinski gasket. Israel J. Math. **97**, 221–238 (1997)

[Kör08]   Körner, T.W.: Hausdorff dimension of sums of sets with themselves. Studia Math. **188**(3), 287–295 (2008)

[Mad08]   Madiman, M.: On the entropy of sums. In: Information Theory Workshop, 2008. ITW '08. IEEE, pp. 303–307 (2008)

[MMT12]   Madiman, M., Marcus, A.W., Tetali, P.: Entropy and set cardinality inequalities for partition-determined functions. Random Struct. Algorithms **40**(4), 399–424 (2012)

[Orp13]   Tuomas O.: On the packing measure of self-similar sets. preprint (2013). http://arxiv.org/abs/1212.4813

[PSS00]   Yuval, P., Wilhelm, S., Boris, S.: Sixty years of Bernoulli convolutions. In: Fractal
          Geometry and Stochastics, II (Greifswald/Koserow, 1998), Progr. Probab., vol. 46,
          pp. 39–65. Birkhäuser, Basel (2000)
[PS00]   Yuval, P., Boris, S.: Problems on self-similar sets and self-affine sets: an update. In:
          Fractal Geometry and stochastics, II (Greifswald/Koserow, 1998), Progr. Probab., vol.
          46, pp. 95–106. Birkhäuser, Basel (2000)
[PS95]   Pollicott, M., Simon, K.: The Hausdorff dimension of λ-expansions with deleted digits.
          Trans. Amer. Math. Soc. **347**(3), 967–983 (1995)
[SS10]   Jörg, S., Pablo, S.: On the dimension of iterated sumsets. In: Recent Developments
          in Fractals and Related Fields, Appl. Numer. Harmon. Anal., pp. 55–72. Birkhäuser
          Boston Inc., Boston (2010)
[Shm13]   Pablo, S.: On the exceptional set for absolute continuity of bernoulli convolutions.
          preprint (2013). http://arxiv.org/abs/1303.3992
[Tao10]   Tao, T.: Sumset and inverse sumset theory for Shannon entropy. Combin. Probab. Com-
          put. **19**(4), 603–639 (2010)
[TV06]   Terence, T., Van V.: Additive Combinatorics, Cambridge Studies in Advanced Mathe-
          matics. vol. 105, Cambridge University Press, Cambridge (2006)