Annals of Mathematics **180** (2014), 773–822
http://dx.doi.org/10.4007/annals.2014.180.2.7

On self-similar sets with overlaps and inverse theorems for entropy

By Michael Hochman

Abstract

We study the dimension of self-similar sets and measures on the line. We show that if the dimension is less than the generic bound of $\min\{1, s\}$, where s is the similarity dimension, then there are superexponentially close cylinders at all small enough scales. This is a step towards the conjecture that such a dimension drop implies exact overlaps and confirms it when the generating similarities have algebraic coefficients. As applications we prove Furstenberg's conjecture on projections of the one-dimensional Sierpinski gasket and achieve some progress on the Bernoulli convolutions problem and, more generally, on problems about parametric families of self-similar measures. The key tool is an inverse theorem on the structure of pairs of probability measures whose mean entropy at scale 2^{-n} has only a small amount of growth under convolution.

1. Introduction

The simplest examples of fractal sets and measures are self-similar sets and measures on the line. These are objects that, like the classical middle-third Cantor set, are made up of finitely many scaled copies of themselves. When these scaled copies are sufficiently separated from each other the small-scale structure is relatively easy to understand and, in particular, there is a closed formula for the dimension. If one does not assume this separation, however, the picture becomes significantly more complicated, and it is a longstanding open problem to compute the dimension. This problem has spawned a number of related conjectures, the most general of which is that, unless some of the small-scale copies exactly coincide, the dimension should be equal to the combinatorial upper bound; that is, the dimension one would get if the small-scale copies did not intersect at all. Special cases of this conjecture have received wide attention; e.g., Furstenberg's projection problem and the Bernoulli convolutions problem. The purpose of this paper is to shed some new light on these matters.

Supported by ERC grant 306494.
© 2014 Department of Mathematics, Princeton University.

1.1. *Self-similar sets and measures and their dimension.* In this paper an iterated function system (IFS) will mean a finite family $\Phi = \{\varphi_i\}_{i\in\Lambda}$ of linear contractions of $\mathbb{R}$, written $\varphi_i(x) = r_i x + a_i$ with $|r_i| < 1$ and $a_i \in \mathbb{R}$. To avoid trivialities we assume throughout that there are at least two distinct contractions. A self-similar set is the attractor of such a system, i.e., the unique compact set $\emptyset \neq X \subseteq \mathbb{R}$ satisfying

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

The self-similar measure associated to Φ and a probability vector $(p_i)_{i\in\Lambda}$ is the unique Borel probability measure μ on $\mathbb{R}^d$ satisfying

$$\mu = \sum_{i\in\Lambda} p_i \cdot \varphi_i \mu. \tag{2}$$

Here $\varphi\mu = \mu \circ \varphi^{-1}$ denotes the push-forward of μ by φ.

When the images $\varphi_i X$ are disjoint or satisfy various weaker separation assumptions, the small-scale structure of self-similar sets and measures is quite well understood. In particular, the Hausdorff dimension $\dim X$ of X is equal to the similarity dimension[1] $\text{s-dim}\, X$, i.e., the unique solution $s \geq 0$ of the equation $\sum |r_i|^s = 1$. With the dimension of a measure θ defined by[2]

$$\dim\theta = \inf\{\dim E \,:\, \theta(E) > 0\},$$

and assuming again sufficient separation of the images $\varphi_i X$, the dimension $\dim\mu$ of a self-similar measure μ is equal to the similarity dimension of μ, defined by

$$\text{s-dim}\,\mu = \frac{\sum p_i \log p_i}{\sum p_i \log r_i}.$$

It is when the images $\varphi_i X$ have significant overlap that computing the dimension becomes difficult, and much less is known. One can give trivial bounds: the dimension is never greater than the similarity dimension, and it is never greater than the dimension of the ambient space $\mathbb{R}$, which is 1. Hence

$$\dim X \leq \min\{1, \text{s-dim}\, X\}, \tag{3}$$

$$\dim \mu \leq \min\{1, \text{s-dim}\,\mu\}. \tag{4}$$

However, without special combinatorial assumptions on the IFS, current methods are unable even to decide whether or not equality holds in (3) and (4), let

[1]This notation is imprecise, since the similarity dimension depends on the IFS Φ rather than the attractor X, but the meaning should always be clear from the context. A similar remark holds for the similarity dimension of measures.

[2]This is the lower Hausdorff dimension. There are many other notions of dimension, but for self-similar measures all the major ones coincide since such measures are exact dimensional [9].

alone compute the dimension exactly. The exception is when there are sufficiently many exact overlaps among the "cylinders" of the IFS. More precisely, for $i = i_1 \cdots i_n \in \Lambda^n$, write

$$\varphi_i = \varphi_{i_1} \circ \cdots \circ \varphi_{i_n}.$$

One says that exact overlaps occur if there are an n and distinct $i, j \in \Lambda^n$ such that $\varphi_i = \varphi_j$ (in particular, the images $\varphi_i X$ and $\varphi_j X$ coincide).[3] If this occurs then X and μ can be expressed using an IFS Ψ that is a proper subset of $\{\varphi_i\}_{i\in\Lambda^n}$, and a strict inequality in (3) and (4) sometimes follows from the corresponding bound for Ψ.

1.2. *Main results.* The present work was motivated by the folklore conjecture that *the occurrence of exact overlaps is the only mechanism that can lead to a strict inequality in* (3) *and* (4) (see, e.g., [27, Question 2.6]). Our main result lends some support to the conjecture and proves some special cases of it. All of our results hold, with suitable modifications, in higher dimensions, but this will appear separately.

Fix $\Phi = \{\varphi_i\}_{i\in\Lambda}$ as in the previous section, and for $i \in \Lambda^n$, write $r_i = r_{i_1} \cdot \ldots \cdot r_{i_n}$, which is the contraction ratio of φ_i. Define the distance between the cylinders associated to $i, j \in \Lambda^n$ by

$$d(i,j) = \begin{cases} \infty & r_i \neq r_j, \\ |\varphi_i(0) - \varphi_j(0)| & r_i = r_j. \end{cases}$$

Note that $d(i,j) = 0$ if and only if $\varphi_i = \varphi_j$ and that the definition is unchanged if 0 is replaced by any other point. For $n \in \mathbb{N}$, let

$$\Delta_n = \min\{d(i,j) \,:\, i,j \in \Lambda^n\,,\, i \neq j\}.$$

Let us make a few observations:

- Exact overlaps occur if and only if $\Delta_n = 0$ for some n (equivalently all sufficiently large n).
- $\Delta_n \to 0$ exponentially. Indeed, the points $\varphi_i(0)$, $i \in \Lambda^n$, can be shown to lie in a bounded interval independent of n, and the exponentially many sequences $i \in \Lambda^n$ give rise to only polynomially many contraction ratios r_i. Therefore there are distinct $i, j \in \Lambda^n$ with $r_i = r_j$ and $|\varphi_i(0) - \varphi_j(0)| < |\Lambda|^{-(1-o(1))n}$.
- There can also be an exponential lower bound for Δ_n. This occurs when the images $\varphi_i(X)$, $i \in \Lambda$, are disjoint, or under the open set condition,

[3] If $i \in \Lambda^k$, $j \in \Lambda^m$ and $\varphi_i = \varphi_j$, then i cannot be a proper prefix of j and vice versa, so $ij, ji \in \Lambda^{k+m}$ are distinct and $\varphi_{ij} = \varphi_{ji}$. Thus exact overlaps occurs also if there is exact coincidence of cylinders at "different generations." Stated differently, exact overlaps means that the semigroup generated by the φ_i, $i \in \Lambda$, is not freely generated by them.

but also sometimes without separation as in Garsia's example [12] or the cases discussed in Theorems 1.5 and 1.6 below.

Our main result on self-similar measures is the following.

THEOREM 1.1. *If μ is a self-similar measure on $\mathbb{R}$ and satisfies* $\dim\mu < \min\{1, \text{s-dim}\,\mu\}$, *then* $\Delta_n \to 0$ *super-exponentially*; *i.e.*, $\lim(-\frac{1}{n}\log\Delta_n) = \infty$.

The conclusion is about Δ_n, which is determined by the IFS Φ, not by the measure. Thus, if the conclusion fails, then $\dim\mu = \text{s-dim}\,\mu$ for every self-similar measure of Φ.

COROLLARY 1.2. *If X is the attractor of an* IFS *on* $\mathbb{R}$ *and if* $\dim X < \min\{1, \text{s-dim}\,X\}$, *then* $\lim(-\frac{1}{n}\log\Delta_n) = \infty$.

Proof. The self-similar measure μ associated to the probabilities $p_i = r_i^{\text{s-dim}\,X}$ satisfies $\text{s-dim}\,\mu = \text{s-dim}\,X$. Since $\mu(X) = 1$, we have $\dim\mu \le \dim X$, so by hypothesis, $\dim\mu < \min\{1, \text{s-dim}\,\mu\}$, and by the theorem, $\Delta_n \to 0$ super-exponentially. □

Theorem 1.1 is derived from a more quantitative result about the entropy of finite approximations of μ. Write $H(\mu,\mathcal{E})$ for the Shannon entropy of a measure μ with respect to a partition $\mathcal{E}$, and write $H(\mu,\mathcal{E}|\mathcal{F})$ for the conditional entropy on $\mathcal{F}$; see Section 3.1. For $n \in \mathbb{Z}$ the dyadic partitions of $\mathbb{R}$ into intervals of length 2^{-n} is

$$\mathcal{D}_n = \left\{\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right) : k \in \mathbb{Z}\right\}.$$

For $t \in \mathbb{R}$, we also write $\mathcal{D}_t = \mathcal{D}_{[t]}$. We remark that $\liminf \frac{1}{n}H(\theta,\mathcal{D}_n) \ge \dim\theta$ for any probability measure θ, and the limit exists and is equal to $\dim\theta$ when θ is exact dimensional, which is the case for self-similar measures [9].

We first consider the case that Φ is uniformly contracting; i.e., that all r_i are equal to some fixed r. Fix a self-similar measure μ defined by a probability vector $(p_i)_{i\in\Lambda}$, and for $i \in \Lambda^n$, write $p_i = p_{i_1}\cdot\ldots\cdot p_{i_n}$. Without loss of generality one can assume that 0 belongs to the attractor X. Define the n-th generation approximation of μ by

$$\nu^{(n)} = \sum_{i\in\Lambda^n} p_i \cdot \delta_{\varphi_i(0)}. \tag{5}$$

This is a probability measure on X and $\nu^{(n)} \to \mu$ weakly. Moreover, writing

$$n' = n\log_2(1/r),$$

$\nu^{(n)}$ closely resembles μ up to scale $2^{-n'} = r^n$ in the sense that

$$\lim_{n\to\infty}\frac{1}{n'}H(\nu^{(n)},\mathcal{D}_{n'}) = \dim\mu.$$

The main question we are interested in is the behavior of $\nu^{(n)}$ at smaller scales. Observe that the entropy $H(\nu^{(n)}, \mathcal{D}_{n'})$ of $\nu^{(n)}$ at scale $2^{-n'}$ may not exhaust the entropy $H(\nu^{(n)})$ of $\nu^{(n)}$ as a discrete measure (i.e., with respect to the partition into points). If there is substantial excess entropy, it is natural to ask at what scale and at what rate it appears; it must appear eventually because $\lim_{k\to\infty} H(\nu^{(n)}, \mathcal{D}_k) = H(\nu^{(n)})$. The excess entropy at scale k relative to the entropy at scale n' is just the conditional entropy $H(\nu^{(n)}, \mathcal{D}_k|\mathcal{D}_{n'}) = H(\nu^{(n)}, \mathcal{D}_k) - H(\nu^{(n)}, \mathcal{D}_{n'})$.

THEOREM 1.3. *Let μ be a self-similar measure on $\mathbb{R}$ defined by an* IFS *with uniform contraction ratios. Let $\nu^{(n)}$ be as above. If* $\dim \mu < 1$, *then*

$$\lim_{n\to\infty} \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{qn'}|\mathcal{D}_{n'}) = 0 \quad \textit{for every } q > 1. \tag{6}$$

Note that we assume $\dim \mu < 1$ but not necessarily $\dim \mu < \text{s-dim}\, \mu$. The statement is valid when $\dim \mu = \text{s-dim}\, \mu < 1$, although for rather trivial reasons.

We now formulate the result in the nonuniformly contracting case. Let

$$r = \prod_{i\in\Lambda} r_i^{p_i}$$

so that $\log r$ is the average logarithmic contraction ratio when φ_i is chosen randomly with probability p_i. Note that, by the law of large numbers, with probability tending to 1, an element $i \in \Lambda^n$ chosen according to the probabilities p_i will satisfy $r_i = r^{n(1+o(1))} = 2^{n'(1+o(1))}$.

With this definition and $\nu^{(n)}$ defined as before, the theorem above holds as stated, but note that now the partitions $\mathcal{D}_k$ are not suitable for detecting exact overlaps, since $\varphi_i(0) = \varphi_j(0)$ may happen for some $i, j \in \Lambda^n$ with $r_i \neq r_j$. To correct this, define the probability measure $\widetilde{\nu}^{(n)}$ on $\mathbb{R} \times \mathbb{R}$ by

$$\widetilde{\nu}^{(n)} = \sum_{i\in\Lambda^n} \delta_{(\varphi_i(0), r_i)}$$

and the partition of $\mathbb{R} \times \mathbb{R}$ given by

$$\widetilde{\mathcal{D}}_n = \mathcal{D}_n \times \mathcal{F},$$

where $\mathcal{F}$ is the partition of $\mathbb{R}$ into points.

THEOREM 1.4. *Let μ be a self-similar measure on $\mathbb{R}$ and $\widetilde{\nu}^{(n)}$ as above. If* $\dim \mu < 1$, *then*

$$\lim_{n\to\infty} \frac{1}{n'} H(\widetilde{\nu}^{(n)}, \widetilde{\mathcal{D}}_{qn'}|\widetilde{\mathcal{D}}_{n'}) = 0 \quad \textit{for every } q > 1. \tag{7}$$

To derive Theorem 1.1, let μ be as in the last theorem with $\dim\mu < \min\{1, \text{s-dim}\,\mu\}$. The conclusion of the last theorem is equivalent to

$$\frac{1}{n'}H(\widetilde{\nu}^{(n)}, \widetilde{\mathcal{D}}_{qn'}) \to \dim\mu$$

for every $q > 1$. Hence for a given q and all sufficiently large n, we will have $\frac{1}{n'}H(\widetilde{\nu}^{(n)}, \widetilde{\mathcal{D}}_{qn'}) < \text{s-dim}\,\mu$. Since $\widetilde{\nu}^{(n)} = \sum_{i\in\Lambda^n} p_i \cdot \delta_{(\varphi_i(0),r_i)}$, if each pair $(\varphi_i(0), r_j)$ belonged to a different atom of $\widetilde{\mathcal{D}}_{qn'}$, then we would have $\frac{1}{n'}H(\widetilde{\nu}^{(n)}, \widetilde{\mathcal{D}}_{qn'}) = -\frac{1}{n\log(1/r)}\sum_{i\in\Lambda^n} p_i \log p_i = \text{s-dim}\,\mu$, a contradiction. Thus there must be distinct $i, j \in \Lambda^n$ for which $(\varphi_i(0), r_i)$, $(\varphi_j(0), r_j)$ lie in the same atom of $\widetilde{\mathcal{D}}_{qn'}$, giving $\Delta_n < 2^{-qn'}$.

1.3. *Outline of the proof.* Let us say a few words about the proofs. For simplicity we discuss Theorem 1.3, where there is a common contraction ratio r to all the maps. For a self-similar measure $\mu = \sum_{i\in\Lambda} p_i \cdot \varphi_i\mu$, iterate this relation n times to get $\mu = \sum_{i\in\Lambda^n} p_i \cdot \varphi_i\mu$. Since each φ_i, $i \in \Lambda^n$, contracts by r^n, all the measures $\varphi_i\mu$, $i \in \Lambda^n$, are translates of each other. The last identity can be rewritten as a convolution

$$\mu = \nu^{(n)} * \tau^{(n)},$$

where as before $\nu^{(n)} = \sum_{i\in\Lambda^n} p_i \cdot \delta_{\varphi_i(0)}$, and $\tau^{(n)}$ is μ scaled down by r^n.

Fix q, and write $a \approx b$ to indicate that the difference tends to 0 as $n \to \infty$. From the entropy identity $H(\mu, \mathcal{D}_{(q+1)n'}) = H(\mu, \mathcal{D}_{n'}) + H(\mu, \mathcal{D}_{(q+1)n'}|\mathcal{D}_{n'})$ and the fact that $\frac{1}{n'}H(\mu, \mathcal{D}_{n'}) \approx \frac{1}{n'}H(\nu^{(n)}, \mathcal{D}_{n'})$, we find that the mean entropy

$$A = \frac{1}{(q+1)n'}H(\mu, \mathcal{D}_{(q+1)n'})$$

is approximately a convex combination $A \approx \frac{1}{(q+1)}B + \frac{q}{(q+1)}C$ of the mean entropy

$$B = \frac{1}{n'}H(\nu^{(n)}, \mathcal{D}_{n'})$$

and the mean conditional entropy

$$C = \frac{1}{qn'}H(\mu, \mathcal{D}_{(q+1)n'}|\mathcal{D}_{n'}) = \sum_{I\in\mathcal{D}_{n'}} \mu(I) \cdot \frac{1}{qn'}H(\nu_I^{(n)} * \tau^{(n)}, \mathcal{D}_{(q+1)n'}),$$

where $\nu_I^{(n)}$ is the normalized restriction of $\nu^{(n)}$ on I. Since $A \approx \dim\mu$ and $B \approx \dim\mu$, we find that $C \approx \dim\mu$ as well. On the other hand, we also have $\frac{1}{qn'}h(\tau^{(n)}, \mathcal{D}_{(q+1)n'}) \approx \dim\mu$. Thus by the expression above, C is an average of terms each of which is close to the mean, and therefore most of them are equal to the mean. We find that

$$\frac{1}{qn'}H(\nu_I^{(n)} * \tau^{(n)}, \mathcal{D}_{(q+1)n'}) \approx C \approx \dim\mu \approx \frac{1}{qn'}H(\tau^{(n)}, \mathcal{D}_{(q+1)n'}) \tag{8}$$

for large n and "typical" $I \in \mathcal{D}_{n'}$. The argument is then concluded by showing that (8) implies that either $\frac{1}{qn'}H(\tau^{(n)}, \mathcal{D}_{(q+1)n'}) \approx 1$ (leading to $\dim \mu = 1$), or that typical intervals I satisfy $\frac{1}{qn'}H(\nu_I^{(n)}, \mathcal{D}_{(q+1)n'}) \approx 0$ (leading to (6)).

Now, for a general pair of measures ν, τ, the relation $\frac{1}{n}H(\nu * \tau, \mathcal{D}_n) \approx \frac{1}{n}H(\nu, \mathcal{D}_n)$ analogous to (8) does not have such an implication. But, while we know nothing about the structure of $\nu_I^{(n)}$, we do know that $\tau^{(n)}$, being self-similar, is highly uniform at different scales. We will be able to utilize this fact to draw the desired conclusion. Evidently, the main ingredient in the argument is an analysis of the growth of measures under convolution, which will occupy us starting in Section 2.

1.4. *Applications.* Theorem 1.1 and its corollaries settle a number of cases of the aforementioned conjecture. Specifically, in any class of IFSs where one can prove that cylinders are either equal or exponentially separated, the only possible cause of dimension drop is the occurrence of exact overlaps. Thus,

THEOREM 1.5. *For* IFS*s on* $\mathbb{R}$ *defined by algebraic parameters, there is a dichotomy: Either there are exact overlaps, or the attractor X satisfies* $\dim X = \min\{1, \text{s-dim}\, X\}$.

Proof. Let $\varphi_i(x) = r_i x + a_i$, and suppose r_i, a_i are algebraic. For distinct $i, j \in \Lambda^n$, the distance $|\varphi_i(0) - \varphi_j(0)|$ is a polynomial of degree n in r_i, a_i, and hence is either equal to 0, or is $\geq s^n$ for some constant $s > 0$ depending only on the numbers r_i, a_i (see Lemma 5.10). Thus $\Delta_n \geq s^n$, and the conclusion follows from Corollary 1.2. $\square$

There are a handful of cases where a similar argument can handle non-algebraic parameters. Among these is a conjecture by Furstenberg from the 1970's, asserting that if the "one dimensional Sierpinski gasket"

$$F = \left\{ \sum (i_n, j_n) 3^{-n} \,:\, (i_n, j_n) \in \{(0,0), (1,0), (0,1)\} \right\}$$

is projected orthogonally to a line of irrational slope, then the dimension of the image is 1 (see, e.g., [27, Question 2.5]).[4] It is more convenient to replace orthogonal projections with the parametrized linear maps $\pi_t : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\pi_t(x, y) = tx + y.$$

[4]This was motivated by a dual conjecture asserting that any line ℓ of irrational slope meets F in a zero dimensional set, and this, in turn, is an analog of similar conjectures arising in metric number theory and layed out in [10]. The intersections and projections conjectures are related by the heuristic that for a map $F \to \mathbb{R}$, a large image corresponds to small fibers, but there is only an implication in one direction. (The statement about intersections implies the one about projections using [11].)

(Up to a linear change of coordinates in the range, this represents the orthogonal projection to the line with slope $-1/t$.) One may verify that the image $F_t = \pi_t F$ is the self-similar defined by the contractions

$$x \mapsto \frac{1}{3}x, \quad x \mapsto \frac{1}{3}(x+1), \quad x \mapsto \frac{1}{3}(x+t). \tag{9}$$

Therefore s-dim $F_t = 1$ for all t, and it is not hard to show that exact overlaps occur only for certain rational values of t. Thus, Furstenberg's conjecture is a special case of the motivating conjecture of this paper.

From general considerations such as Marstrand's theorem, we know that $\dim F_t = 1$ for almost every t, and Kenyon showed that this holds also for a dense G_δ set of t [19]. In the same paper Kenyon also classified those rational t for which $\dim F_t = 1$ and showed that F_t has Lebesgue measure 0 for all irrational t (strengthening the conclusion of a general theorem of Besicovitch that gives this for almost every t). For some other partial results, see [34].

THEOREM 1.6. *If* $t \notin \mathbb{Q}$, *then* $\dim F_t = 1$.

Proof. Fix t, and suppose that $\dim F_t < 1$. Let $\Lambda = \{0, 1, t\}$ and $\varphi_i(x) = \frac{1}{3}(x+i)$, so F_t is the attractor of $\{\varphi_i\}_{i\in\Lambda}$. For $i \in \Lambda^n$, one may check that $\varphi_i(0) = \sum_{k=1}^n i_k 3^{-k}$. Inserting this into the difference $\varphi_i(0) - \varphi_j(0)$ we can separate the terms that are multiplied by t from those that are not, and we find that $|\varphi_i(0) - \varphi_j(0)| = p_{i,j} - t \cdot q_{i,j}$ for rational numbers $p_{i,j}, q_{i,j}$ belonging to the set

$$X_n = \left\{ \sum_{i=1}^n a_i 3^{-i} \;:\; a_i \in \{\pm 1, 0\} \right\}.$$

Therefore there are $p_n, q_n \in X_n$ such that $\Delta_n = |p_n - tq_n|$, so by Corollary 1.2,

$$|p_n - t \cdot q_n| < 30^{-n} \qquad \text{for large enough } n. \tag{10}$$

If $q_n = 0$ for n satisfying (10), then $|p_n| < 30^{-n}$, but, since p_n is rational with denominator 3^n, this can only happen if $p_n = 0$. This in turn implies that $\Delta_n = 0$; i.e., there are exact overlaps, so $t \in \mathbb{Q}$.

On the other hand, suppose $q_n \neq 0$ for all large n. Since q_n is a nonzero rational with denominator 3^n, we have $q_n \geq 3^{-n}$. Dividing (10) by q_n we get $|t - p_n/q_n| < 10^{-n}$. Subtracting successive terms, by the triangle inequality we have

$$|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}| < 2 \cdot 10^{-n} \qquad \text{for large enough } n.$$

But $p_n, q_n, p_{n+1}, q_{n+1} \in X_{n+1}$, so $p_{n+1}/q_{n+1} - p_n/q_n$ is rational with denominator $\leq 9^{n+1}$, giving

$$\left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| \neq 0 \quad \Longrightarrow \quad \left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| \geq 9^{-(n+1)}.$$

Since $9^{-(n+1)} \le 2 \cdot 10^{-n}$ is impossible for large n, the last two equations imply that $p_n/q_n = p_{n+1}/q_{n+1}$ for all large n. Therefore there is an n_0 such that $|t - p_{n_0}/q_{n_0}| < 10^{-n}$ for $n > n_0$, which gives $t = p_0/q_0$. $\square$

The argument above is due to B. Solomyak and P. Shmerkin, and we thank them for permission to include it here. Similar considerations work in a few other cases, but one already runs into difficulties if in the example above we replace the contraction ratio $1/3$ with any nonalgebraic $0 < r < 1$. (See also the discussion following Theorem 1.9 below.)

In the absence of a resolution of the general conjecture, we turn to parametric families of self-similar sets and measures. The study of parametric families of general sets and measures is classical; examples include the projection theorems of Besicovitch and Marstrand and more recent results like those of Peres-Schlag [24] and Bourgain [3]. When the sets and measures in question are self-similar we shall see that the general results can be strengthened considerably.

Let I be a set of parameters, and let $r_i : I \to (-1,1) \setminus \{0\}$ and $a_i : I \to \mathbb{R}$, $i \in \Lambda$. For each $t \in I$, define $\varphi_{i,t} : \mathbb{R} \to \mathbb{R}$ by $\varphi_{i,t}(x) = r_i(t)(x - a_i(t))$. For a sequence $i \in \Lambda^n$, let $\varphi_{i,t} = \varphi_{i_1,t} \circ \cdots \circ \varphi_{i_n,t}$, and define

$$\Delta_{i,j}(t) = \varphi_{i,t}(0) - \varphi_{j,t}(0). \tag{11}$$

The quantity $\Delta_n = \Delta_n(t)$ associated as in the previous section to the IFS $\{\varphi_{i,t}\}_{i\in\Lambda}$ is not smaller than the minimum of $|\Delta_{i,j}(t)|$ over distinct $i, j \in \Lambda^n$ (since it is the minimum over pairs i, j with $r_i = r_j$). Thus, $\Delta_n \to 0$ super-exponentially implies that $\min\{|\Delta_{i,j}(t)|, \, i, j \in \Lambda^n\} \to 0$ super-exponentially as well, so Theorem 1.1 has the following formal implication.

Theorem 1.7. *Let $\Phi_t = \{\varphi_{i,t}\}$ be a parametrized* IFS *as above. For every $\varepsilon > 0$, let*

$$E_\varepsilon = \bigcup_{N=1}^{\infty} \bigcap_{n>N} \left(\bigcup_{i,j\in\Lambda^n} (\Delta_{i,j})^{-1}(-\varepsilon^n, \varepsilon^n) \right) \tag{12}$$

and

$$E = \bigcap_{\varepsilon>0} E_\varepsilon. \tag{13}$$

Then for $t \in I \setminus E$, for every probability vector $p = (p_i)$, the associated self-similar measure μ_t of Φ_t satisfies $\dim \mu_t = \min\{1, \text{s-dim}\, \mu_t\}$, *and the attractor X_t of Φ_t satisfies* $\dim X_t = \min\{1, \text{s-dim}\, X_t\}$.

Our goal is to show that the set E defined in the theorem above is small. We restrict ourselves to the case that $I \subseteq \mathbb{R}$ is a compact interval; a multi-parameter version will appear in [14]. Extend the definition of $\Delta_{i,j}$ to infinite sequences $i, j \in \Lambda^{\mathbb{N}}$ by

$$\Delta_{i,j}(t) = \lim_{n\to\infty} \Delta_{i_1\cdots i_n, j_1\cdots j_n}(t). \tag{14}$$

Convergence is uniform over I and i, j, and if $a_i(\cdot)$ and $r_i(\cdot)$ are real analytic in a neighborhood of I, then so are the functions $\Delta_{i,j}(\cdot)$.

THEOREM 1.8. *Let $I \subseteq \mathbb{R}$ be a compact interval, let $r : I \to (-1, 1) \setminus \{0\}$ and $a_i : I \to \mathbb{R}$ be real analytic, and let $\Phi_t = \{\varphi_{i,t}\}_{i\in\Lambda}$ be the associated parametric family of* IFS*s, as above. Suppose that*

$$\forall i, j \in \Lambda^{\mathbb{N}} \quad (\ \Delta_{i,j} \equiv 0 \text{ on } I \quad \iff \quad i = j\ ).$$

Then the set E of "exceptional" parameters in Theorem 1.7 *has Hausdorff and packing dimension* 0.

The condition in the theorem is extremely mild. Essentially it means that the family does not have overlaps "built in." For an example where the hypothesis fails, consider the case that there are $i \neq j$ with $\varphi_{i,t} = \varphi_{j,t}$ for all t. In this case the conclusion sometimes fails as well.

Most existing results on parametric families of IFSs are based on the so-called transversality method, introduced by Pollicott and Simon [28] and developed, among others, by Solomyak [33] and Peres-Schlag [24]. Theorem 1.8 is based on a similar but much weaker "higher order" transversality condition, which is automatically satisfied under the stated hypothesis. We give the details in Section 5.4. See [32] for an effective derivation of higher-order transversality in certain contexts.

As a demonstration we apply this to the Bernoulli convolutions problem. For $0 < \lambda < 1$, let ν_λ denote the distribution of the real random variable $\sum_{n=0}^{\infty} \pm\lambda^n$, where the signs are chosen i.i.d. with equal probabilities. The name derives from the fact that ν_λ is the infinite convolution of the measures $\frac{1}{2}(\delta_{-\lambda^n} + \delta_{\lambda^n})$, $n = 0, 1, 2, \ldots$, but the pertinent fact for us is that ν_λ is a self-similar measure, given by assigning equal probabilities to the contractions

$$\varphi_\pm(x) = \lambda x \pm 1. \tag{15}$$

For $\lambda < \frac{1}{2}$, the measure is supported on a self-similar Cantor set of dimension < 1, but for $\lambda \in [\frac{1}{2}, 1)$, the support is an interval, and it is a well-known open problem to determine whether it is absolute continuous. Exact overlaps can occur only for certain algebraic λ, and Erdős showed that when λ^{-1} is a Pisot number, ν_λ is in fact singular [5]. No other parameters $\lambda \in [\frac{1}{2}, 1)$ are known for which ν_λ is singular. In the positive direction, it is known that ν_λ is absolutely continuous for almost every $\lambda \in [1/2, 1)$ (Solomyak [33]) and the set of exceptional $\lambda \in [a, 1)$ has dimension $< 1 - C(a - 1/2)$ for some $C > 0$ (Peres-Schlag [24]) and its dimension tends to 0 as $a \to 1$ (Erdős [6]).

We shall consider the question of when $\dim \nu_\lambda = 1$. This is weaker than absolute continuity, but little more seems to be known about this question except the relatively soft fact that the set of parameters with $\dim \nu_\lambda = 1$ is also topologically large (contains a dense G_δ set); see [25]. In particular, the only parameters $\lambda \in [1/2, 1)$ for which $\dim \nu_\lambda < 1$ is known are inverses of Pisot numbers (Alexander-Yorke [1]). We also note that in many of the problems related to Bernoulli convolutions it is the dimension of ν_λ, rather than its absolute continuity, that are relevant. For discussion of some applications, see [25, §8] and [29].

THEOREM 1.9. $\dim \nu_\lambda = 1$ *outside a set of* λ *of dimension* 0. *Furthermore, the exceptional parameters for which* $\dim \nu_\lambda < 1$ *are "nearly algebraic" in the sense that for every* $0 < \theta < 1$ *and all large enough* n, *there is a polynomial* $p_n(t)$ *of degree* n *and coefficients* $0, \pm 1$, *such that* $|p_n(\lambda)| < \theta^n$.

Proof. Take the parametrization $r(t) = t$, $a_\pm(t) = \pm 1$ for $t \in [1/2, 1-\varepsilon]$. Then $\Delta_{i,j}(t) = \sum (i_n - j_n) \cdot t^n$, and this vanishes identically if and only if $i = j$, confirming the hypothesis of Theorem 1.8. Since $\Delta_n(t)$ is a polynomial of degree n with coefficients $0, \pm 1$, so the second statement follows the description of the set E in Theorem 1.8. □

Arguing as in the proof of Theorem 1.6, in order to show that $\dim \nu_\lambda = 1$ for all nonalgebraic λ, it would suffice to answer the following question in the affirmative.[5]

QUESTION 1.10. *Let* Π_n *denote the collection of polynomial of degree* $\leq n$ *with coefficients* $0, \pm 1$. *Does there exist a constant* $s > 0$ *such that for* α, β *that are roots of polynomials in* Π_n, *either* $\alpha = \beta$ *or* $|\alpha - \beta| > s^n$?

Classical bounds imply this for $s \sim 1/n$, but we have not found an answer to the question in the literature.

Another problem to which our methods apply is the Keane-Smorodinsky {0,1,3}-problem. For details about the problem, we refer to Pollicott-Simon [28] or Keane-Smorodinsky-Solomyak [18].

Finally, our methods also can be adapted with minor changes to IFSs that "contract on average" [23]. We restrict attention to a problem raised by Sinai [26] concerning the maps $\varphi_- : x \mapsto (1-\alpha)x - 1$ and $\varphi_+ : x \mapsto (1+\alpha)x + 1$. A composition of n of these maps chosen i.i.d. with probability $\frac{1}{2}, \frac{1}{2}$ asymptotically contracts by approximately $(1-\alpha^2)^{n/2}$, and so for each $0 < \alpha < 1$, there is a unique probability measure μ_α on $\mathbb{R}$ satisfying $\mu_\alpha = \frac{1}{2}\varphi_- \mu_\alpha + \frac{1}{2}\varphi_+ \mu_\alpha$. Little is known about the dimension or absolute continuity of μ_α

[5] In order to show that an "almost-root" of a polynomial is close to an acrual root one can rely on the classical transversality arguments; e.g., [33].

beyond upper bounds analogous to (4). Some results in a randomized analog of this model have been obtained by Peres, Simon, and Solomyak [26]. We prove

THEOREM 1.11. *There is a set* $E \subseteq (0,1)$ *of Hausdorff* (*and packing*) *dimension* 0 *such that* $\dim \mu_\alpha = \min\{1, \text{s-dim}\, \mu_\alpha\}$ *for* $\alpha \in (0,1) \setminus E$.

For further discussion of this problem, see Section 5.5.

1.5. *Absolute continuity*? There is a conjecture analogous to the one we began with, predicting that if μ is a self-similar measure, $\text{s-dim}\, \mu > 1$, and there are no exact overlaps, then μ should be absolutely continuous with respect to Lebesgue measure. The Bernoulli convolutions problem discussed above is a special case of this conjecture.

Our methods at present are not able to tackle this problem. At a technical level, whenever our methods give $\dim \mu = 1$ it is a consequence of showing that $H(\mu, \mathcal{D}_n) = n - o(n)$. In contrast, absolute continuity would require better asymptotics; e.g., $H(\mu, \mathcal{D}_n) = n - O(1)$ (see [13, Th. 1.5]). More substantially, our arguments do not distinguish between the critical $\text{s-dim}\, \mu = 1$, where the conclusion of the conjecture is generally false, and super-critical phase $\text{s-dim}\, \mu > 1$, so in their present form they cannot possibly give results about absolute continuity.

The discussion above notwithstanding, shortly after this paper appeared in preprint form, P. Shmerkin found an ingenious way to "amplify" our results on parametric families of self-similar measures and obtain results about absolute continuity. For instance,

THEOREM (Shmerkin [31]). *There is a set* $E \subseteq (\frac{1}{2}, 1)$ *of Hausdorff dimension* 0 *such that the Bernoulli convolution* ν_λ *is absolutely continuous for all* $\lambda \in (\frac{1}{2}, 1) \setminus E$.

The idea of the proof is to split ν_λ as a convolution $\nu'_\lambda * \nu''_\lambda$ of self-similar measures, with $\text{s-dim}\, \nu'_\lambda \geq 1$ and $\text{s-dim}\, \nu''_\lambda > 0$. By Theorem 1.8, $\dim \nu'_\lambda = 1$ outside a zero-dimensional set E' of parameters. On the other hand, a classical argument of Erdős and Kahane shows that, outside a zero-dimensional set E'' of parameters, the Fourier transform of ν''_λ has power decay. Taking $E = E' \cup E''$, Shmerkin shows that $\nu_\lambda = \nu'_\lambda * \nu''_\lambda$ is absolutely continuous for $\lambda \in (\frac{1}{2}, 1) \setminus E$.

At present the argument above is limited by the fact that E'' is completely noneffective, so, unlike Theorem 1.1, it does not give a condition that applies to *individual* self-similar measure and does not provide concrete new examples of parameters for which ν_λ is absolutely continuous. In contrast, Corollary 1.5 tells us that $\dim \nu_\lambda = 1$ whenever $\lambda \in (\frac{1}{2}, 1) \cap \mathbb{Q}$, as well as other algebraic examples. It remains a challenge to prove a similar result for absolute continuity.

1.6. *Notation and organization of the paper.* The main ingredient in the proofs are our results on the growth of convolutions of measures. We develop this subject in the next three sections. Section 2 introduces the statements and basic definitions, Section 3 contains some preliminaries on entropy and convolutions, and Section 4 proves the main results on convolutions. In Section 5 we prove Theorem 1.1 and the other main results.

We follow standard notational conventions. $\mathbb{N} = \{1, 2, 3, \dots\}$. All logarithms are to base 2. $\mathcal{P}(X)$ is the space of probability measures on X, endowed with the weak-* topology if appropriate. We follow standard "big O" notation: $O_\alpha(f(n))$ is an unspecified function bounded in absolute value by $C_\alpha \cdot f(n)$ for some constant C_α depending on α. Similarly $o(1)$ is a quantity tending to 0 as the relevant parameter $\to \infty$. The statement "for all s and $t > t(s), \dots$" should be understood as saying "there exists a function $t(\cdot)$ such that for all s and $t > t(s), \dots$". If we want to refer to the function $t(\cdot)$ outside the context where it is introduced, we will designate it as $t_1(\cdot)$, $t_2(\cdot)$, etc.

Acknowledgment. I am grateful to Pablo Shmerkin and Boris Solomyak for many contributions that have made this a better paper, and especially for their permission to include the derivation of Theorem 1.6. I also thank Nicolas de Saxce and Izabella Laba for their comments. This project began during a visit to Microsoft Research in Redmond, Washington, and I would like to thank Yuval Peres and the members of the theory group for their hospitality.

2. An inverse theorem for the entropy of convolutions

2.1. *Entropy and additive combinatorics.* As we saw in Section 1.3, a key ingredient in the proof of Theorem 1.3 is an analysis of the growth of measures under convolution. This subject is of independent interest and will occupy us for a large part of this paper.

It will be convenient to introduce the normalized scale-n entropy

$$H_n(\mu) = \frac{1}{n} H(\mu, \mathcal{D}_n).$$

Our aim is to obtain structural information about measures μ, ν for which $\mu * \nu$ is small in the sense that

$$H_n(\mu * \nu) \leq H_n(\mu) + \delta, \tag{16}$$

where $\delta > 0$ is small but fixed and n is large.

This problem is a relative of classical ones in additive combinatorics concerning the structure of sets A, B whose sumset $A+B = \{a+b \,:\, a \in A\,,\, b \in B\}$ is appropriately small. The general principle is that when the sum is small, the sets should have some algebraic structure. Results to this effect are known as inverse theorems. For example, the Freiman-Ruzsa theorem asserts that if

$|A+B| \leq C|A|$, then A, B are close, in a manner depending on C, to generalized arithmetic progressions[6] (the converse is immediate). For details and more discussion see, e.g., [36].

The entropy of a discrete measure corresponds to the logarithm of the cardinality of a set, and convolution is the analog for measures of the sumset operation. Thus the analog of the condition $|A+A| \leq C|A|$ is

$$H_n(\mu * \mu) \leq H_n(\mu) + O\left(\frac{1}{n}\right). \tag{17}$$

An entropy version of Freiman's theorem was recently proved by Tao [35], who showed that if μ satisfies (17), then it is close, in an appropriate sense, to a uniform measures on a (generalized) arithmetic progression.

The condition (16), however, is significantly weaker than (17) even when the latter is specialized to $\nu = \mu$, and it is harder to draw conclusions from it about the global structure of μ. Consider the following example. Start with an arithmetic progression of length n_1 and gap ε_1, and put the uniform measure on it. Now split each atom x into an arithmetic progression of length n_2 and gap $\varepsilon_2 < \varepsilon_1/n_2$, starting at x (so the entire gap fits in the space between x and the next atom). Repeat this procedure N times with parameters n_i, ε_i, and call the resulting measure μ. Let k be such that ε_N is of order 2^{-k}. It is not hard to verify that we can have $H_k(\mu) = 1/2$ but $|H_k(\mu) - H_k(\mu * \mu)|$ arbitrarily small. This example is actually the uniform measure on a (generalized) arithmetic progression, as predicted by Freiman-type theorems, but the rank N can be arbitrarily large. Furthermore, if one conditions μ on an exponentially small subset of its support, one gets another example with the similar properties that is quite far from a generalized arithmetic progression.

Our main contribution to this matter is Theorem 2.7 below, which shows that constructions like the one above are, in a certain statistical sense, the only way that (16) can occur. We note that there is a substantial existing literature on the growth condition $|A+B| \leq |A|^{1+\delta}$, which is the sumset analog of (16). Such a condition appears in the sum-product theorems of Bourgain-Katz-Tao [4] in the work of Katz-Tao [17], and in the Euclidean setting, more explicitly in Bourgain's work on the Erdős-Volkmann conjecture [2] and Marstrand-like projection theorems [3]. However we have not found a result in the literature that meets our needs and, in any event, we believe that the formulation given here will find further applications.

2.2. *Component measures.* The following notation will be needed in $\mathbb{R}^d$ as well as $\mathbb{R}$. Let $\mathcal{D}_n^d = \mathcal{D}_n \times \cdots \times \mathcal{D}_n$ denote the dyadic partition of $\mathbb{R}^d$; we often

[6] A generalized arithmetic progression is an affine image of a box in a higher-dimensional lattice.

suppress the superscript when it is clear from the context. Let $\mathcal{D}_n(x) \in \mathcal{D}_n$ denote the unique level-n dyadic cell containing x. For $D \in \mathcal{D}_n$, let $T_D : \mathbb{R}^d \to \mathbb{R}^d$ be the unique homothety mapping D to $[0,1)^d$. Recall that if $\mu \in \mathcal{P}(\mathbb{R})$, then $T_D\mu$ is the push-forward of μ through T_D .

Definition 2.1. For $\mu \in \mathcal{P}(\mathbb{R}^d)$ and a dyadic cell D with $\mu(D) > 0$, the (raw) D-component of μ is

$$\mu_D = \frac{1}{\mu(D)}\mu|_D$$

and the (rescaled) D-component is

$$\mu^D = \frac{1}{\mu(D)}T_D(\mu|_D).$$

For $x \in \mathbb{R}^d$ with $\mu(\mathcal{D}_n(x)) > 0$, we write

$$\begin{aligned}\mu_{x,n} &= \mu_{\mathcal{D}_n(x)},\\ \mu^{x,n} &= \mu^{\mathcal{D}_n(x)}.\end{aligned}$$

These measures, as x ranges over all possible values for which $\mu(\mathcal{D}_n(x)) > 0$, are called the level-$n$ components of μ.

Our results on the multi-scale structure of $\mu \in \mathbb{R}^d$ are stated in terms of the behavior of random components of μ, defined as follows.[7]

Definition 2.2. Let $\mu \in \mathcal{P}(\mathbb{R}^d)$.

(1) A random level-n component, raw or rescaled, is the random measure μ_D or μ^D, respectively, obtained by choosing $D \in \mathcal{D}_n$ with probability $\mu(D)$; equivalently, the random measure $\mu_{x,n}$ or $\mu^{x,n}$, respectively, with x chosen according to μ.

(2) For a finite set $I \subseteq \mathbb{N}$, a random level-I component, raw or rescaled, is chosen by first choosing $n \in I$ uniformly, and then (conditionally independently on the choice of n) choosing a raw or rescaled level-n component, respectively.

Notation 2.3. When the symbols $\mu^{x,i}$ and $\mu_{x,i}$ appear inside an expression $\mathbb{P}(\cdots)$ or $\mathbb{E}(\cdots)$, they will always denote random variables drawn according to the component distributions defined above. The range of i will be specified as needed.

[7]Definition 2.2 is motivated by Furstenberg's notion of a CP-distribution [10], [11], [15], which arises as limits as $N \to \infty$ of the distribution of components of level $1, \ldots, N$. These limits have a useful dynamical interpretation but in our finitary setting we do not require this technology.

The definition is best understood with some examples. For $A \subseteq \mathcal{P}([0,1]^d)$ we have

$$\mathbb{P}_{i=n}\left(\mu^{x,i} \in A\right) = \int 1_A(\mu^{x,n})\, d\mu(x),$$

$$\mathbb{P}_{0\le i\le n}\left(\mu^{x,i} \in A\right) = \frac{1}{n+1}\sum_{i=0}^{n}\int 1_A(\mu^{x,i})\, d\mu(x).$$

This notation implicitly defines x, i as random variables. Thus if $A_0, A_1, \ldots \subseteq \mathcal{P}([0,1]^d)$ and $D \subseteq [0,1]^d$, we could write

$$\mathbb{P}_{0\le i\le n}\left(\mu^{x,i} \in A_i \text{ and } x \in D\right) = \frac{1}{n+1}\sum_{i=0}^{n}\mu\left(x\,:\,\mu^{x,i} \in A_i \text{ and } x \in D\right).$$

Similarly, for $f : \mathcal{P}([0,1)^d) \to \mathbb{R}$ and $I \subseteq \mathbb{N}$, we have the expectation

$$\mathbb{E}_{i\in I}\left(f(\mu^{x,i})\right) = \frac{1}{|I|}\sum_{i\in I}\int f(\mu^{x,i})\, d\mu(x).$$

When dealing with components of several measures μ, ν, we assume all choices of components $\mu^{x,i}$, $\nu^{y,j}$ are independent unless otherwise stated. For instance,

$$\mathbb{P}_{i=n}\left(\mu^{x,i} \in A\,,\, \nu^{y,i} \in B\right) = \iint 1_A(\mu^{x,n}) \cdot 1_B(\nu^{y,n})\, d\mu(x)\, d\nu(y).$$

Here 1_A is the indicator function on A, given by $1_A(\omega) = 1$ if $\omega \in A$ and 0 otherwise.

We record one obvious fact, which we will use repeatedly.

LEMMA 2.4. *For $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $n \in \mathbb{N}$,*

$$\mu = \mathbb{E}_{i=n}\left(\mu_{x,i}\right).$$

We sometimes use similar notation to average a sequence $a_n, \ldots, a_{n+k} \in \mathbb{R}$:

$$\mathbb{E}_{n\le i\le n+k}\left(a_i\right) = \frac{1}{k+1}\sum_{i=n}^{n+k} a_i.$$

2.3. *An inverse theorem.* The approximate equality $H_n(\mu * \nu) \approx H_n(\mu)$ occurs trivially if either μ is uniform (Lebesgue) measure on $[0,1]$, or if $\nu = \delta_x$ is a point mass. As we saw in Section 2.1, there are other ways this can occur, but the theorem below shows that in a *statistical* sense, *locally* (i.e., for typical component measures) the two trivial scenarios are essentially the only ones. In order to state this precisely we require finite-scale and approximate versions of being uniform and being a point mass. There are many definitions to choose from. One possible choice is the following.

Definition 2.5. A measure $\mu \in \mathcal{P}([0,1])$ is *ε-atomic* if there is an interval I of length ε such that $\mu(I) > 1 - \varepsilon$.

Alternatively we could require that the entropy be small at a given scale, or that the random variable whose distribution is the given measure has small variance. Up to choice of parameters these definitions coincide, and we shall use all the definitions later. See Definition 3.9 and the discussion following it, and Lemma 4.4, below.

Definition 2.6. A measure $\mu \in \mathcal{P}([0,1])$ is (ε, m)-*uniform if* $H_m(\mu) > 1-\varepsilon$.

Again one can imagine many alternative definitions. For example, almost-uniformity of $\mu \in \mathcal{P}([0,1])$ at scale δ could mean that $|\mu(I) - |I|| < \delta^2$ for all intervals I of length $|I| \geq \delta$, or that the Fourier transform $\widehat{\mu}(\xi)$ is small at frequencies $|\xi| < 1/\delta$. Again, these definitions are essentially equivalent, up to adjustment of parameters, to the one above. We shall not use them here.

THEOREM 2.7. *For every $\varepsilon > 0$ and integer $m \geq 1$, there is a $\delta = \delta(\varepsilon, m) > 0$ such that for every $n > n(\varepsilon, \delta, m)$, the following holds. If $\mu, \nu \in \mathcal{P}([0,1])$ and*

$$H_n(\mu * \nu) < H_n(\mu) + \delta,$$

then there are disjoint subsets $I, J \subseteq \{1, \ldots, n\}$ with $|I \cup J| > (1-\varepsilon)n$, such that

$$\begin{aligned} \mathbb{P}_{i=k}\left(\mu^{x,i} \text{ is } (\varepsilon, m)\text{-uniform}\right) \;&>\; 1-\varepsilon \text{ for } k \in I, \\ \mathbb{P}_{i=k}\left(\nu^{x,i} \text{ is } \varepsilon\text{-atomic}\right) \;&>\; 1-\varepsilon \text{ for } k \in J. \end{aligned}$$

From this it is easy to derive many variants of the theorem for the other notions of atomicity and uniformity discussed above. In Section 2.3 we give a marginally stronger statement in which atomicity is expressed in terms of entropy.

The proof is given in Section 4.4. The dependence of δ on ε, m is effective, but the bounds we obtain are certainly far from optimal, and we do not pursue this topic. The value of n depends among other things on the rate at which $H_m(\mu) \to \dim \mu$, which is currently not effective.

The converse direction of the theorem is false; that is, there are measures that satisfy the conclusion but also $H_n(\mu * \nu) > H_n(\mu) + \delta$. To see this begin with a measure $\mu \in [0,1]$ such that $\dim(\mu * \mu) = \dim \mu = 1/2$ and such that $\lim H_n(\mu) = \lim H_n(\mu * \mu) = \frac{1}{2}$. (Such measures are not hard to construct; see, e.g., [7] or the more elaborate constructions in [20], [30].) By Marstrand's theorem, for almost every t, the scaled measure $\nu(A) = \mu(tA)$ satisfies $\dim \mu * \nu = 1$ and hence $H_n(\mu * \nu) \to 1$. But it is easy to verify that as the conclusion of the theorem holds for the pair μ, μ, it holds for μ, ν as well.

Note that there is no assumption on the entropy of ν, but if $H_n(\nu)$ is sufficiently close to 0, the conclusion will automatically hold with I empty, and if $H_n(\nu)$ is not too close to 0, then J cannot be too large relative to n (see Lemma 3.4 below). We obtain the following useful conclusion.

THEOREM 2.8. *For every $\varepsilon > 0$ and integer m, there is a $\delta = \delta(\varepsilon, m) > 0$ such that for every $n > n(\varepsilon, \delta, m)$ and every $\mu \in \mathcal{P}([0,1])$, if*

$$\mathbb{P}_{0 \le i \le n}\left(H_m(\mu^{x,i}) < 1-\varepsilon\right) > 1-\varepsilon,$$

then for every $\nu \in \mathcal{P}([0,1])$,

$$H_n(\nu) > \varepsilon \implies H_n(\mu * \nu) \ge H_n(\mu) + \delta.$$

Specializing the above to self-convolutions we have the following result, which shows that constructions like the one described in Section 2.1 are, roughly, the only way that $H_n(\mu * \mu) = H_n(\mu) + \delta$ can occur. This should be compared with the results of Tao [35], who studied the condition $H_n(\mu * \mu) = H_n(\mu) + O(\frac{1}{n})$.

THEOREM 2.9. *For every $\varepsilon > 0$ and integer m, there is a $\delta = \delta(\varepsilon, m) > 0$ such that for every sufficiently large $n > n(\varepsilon, \delta, m)$ and every $\mu \in \mathcal{P}([0,1))$, if*

$$H_n(\mu * \mu) < H_n(\mu) + \delta,$$

then there are disjoint subsets $I, J \subseteq \{0, \ldots, n\}$ with $|I \cup J| \ge (1-\varepsilon)n$ and such that

$$\mathbb{P}_{i=k}\left(\mu^{x,i} \text{ is } (\varepsilon, m)\text{-uniform}\right) > 1-\varepsilon \text{ for } k \in I,$$
$$\mathbb{P}_{i=k}\left(\mu^{x,i} \text{ is } \varepsilon\text{-atomic}\right) > 1-\varepsilon \text{ for } k \in J.$$

These results hold more generally for compactly supported measures, but the parameters will depend on the diameter of the support. They can also be extended to measures with unbounded support under additional assumptions; see Section 5.5.

3. **Entropy, atomicity, uniformity**

3.1. *Preliminaries on entropy.* The Shannon entropy of a probability measure μ with respect to a countable partition $\mathcal{E}$ is given by

$$H(\mu, \mathcal{E}) = -\sum_{E \in \mathcal{E}} \mu(E) \log \mu(E),$$

where the logarithm is in base 2 and $0 \log 0 = 0$. The conditional entropy with respect to a countable partition $\mathcal{F}$ is

$$H(\mu, \mathcal{E}|\mathcal{F}) = \sum_{F \in \mathcal{F}} \mu(F) \cdot H(\mu_F, \mathcal{E}),$$

where $\mu_F = \frac{1}{\mu(F)}\mu|_F$ is the conditional measure on F. For a discrete probability measure μ, we write $H(\mu)$ for the entropy with respect to the partition into points, and for a probability vector $\alpha = (\alpha_1, \dots, \alpha_k)$, we write

$$H(\alpha) = -\sum \alpha_i \log \alpha_i.$$

We collect here some standard properties of entropy.

LEMMA 3.1. *Let μ, ν be probability measures on a common space, $\mathcal{E}, \mathcal{F}$ partitions of the underlying space and $\alpha \in [0, 1]$.*

(1) $H(\mu, \mathcal{E}) \geq 0$, *with equality if and only if μ is supported on a single atom of $\mathcal{E}$.*
(2) *If μ is supported on k atoms of $\mathcal{E}$, then $H(\mu, \mathcal{E}) \leq \log k$.*
(3) *If $\mathcal{F}$ refines $\mathcal{E}$ (i.e., for all $F \in \mathcal{F}$, there exists $E \in \mathcal{E}$ such that $F \subseteq E$), then $H(\mu, \mathcal{F}) \geq H(\mu, \mathcal{E})$.*
(4) *If $\mathcal{E} \vee \mathcal{F} = \{E \cap F : E \in \mathcal{E}, F \in \mathcal{F}\}$ is the join of $\mathcal{E}$ and $\mathcal{F}$, then*

$$H(\mu, \mathcal{E} \vee \mathcal{F}) = H(\mu, \mathcal{F}) + H(\mu, \mathcal{E}|\mathcal{F}).$$

(5) $H(\cdot, \mathcal{E})$ *and* $H(\cdot, \mathcal{E}|\mathcal{F})$ *are concave.*
(6) $H(\cdot, \mathcal{E})$ *obeys the "convexity" bound*

$$H\left(\sum \alpha_i \mu_i, \mathcal{E}\right) \leq \sum \alpha_i H(\mu_i, \mathcal{E}) + H(\alpha).$$

In particular, we note that for $\mu \in \mathcal{P}([0, 1]^d)$, we have the bounds $H(\mu, \mathcal{D}_m) \leq md$ (hence $H_n(\mu) \leq 1$) and $H(\mu, \mathcal{D}_{n+m}|\mathcal{D}_n) \leq md$.

Although the function $(\mu, m) \mapsto H(\mu, \mathcal{D}_m)$ is not weakly continuous, the following estimates provide usable substitutes.

LEMMA 3.2. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, $\mathcal{E}, \mathcal{F}$ be partitions of $\mathbb{R}^d$, and let $m, m' \in \mathbb{N}$.*

(1) *Given a compact $K \subseteq \mathbb{R}^d$ and $\mu \in \mathcal{P}(K)$, there is a neighborhood $U \subseteq \mathcal{P}(K)$ of μ such that $|H(\nu, \mathcal{D}_m) - H(\mu, \mathcal{D}_m)| = O_d(1)$ for $\nu \in U$.*
(2) *If each $E \in \mathcal{E}$ intersects at most k elements of $\mathcal{F}$ and vice versa, then $|H(\mu, \mathcal{E}) - H(\mu, \mathcal{F})| = O(\log k)$.*
(3) *If $f, g : \mathbb{R}^d \to \mathbb{R}^k$ and $\|f(x) - g(x)\| \leq C2^{-m}$ for $x \in \mathbb{R}^d$, then $|H(f\mu, \mathcal{D}_m) - H(g\mu, \mathcal{D}_m)| \leq O_{C,k}(1)$.*
(4) *If $\nu(\cdot) = \mu(\cdot + x_0)$, then $|H(\mu, \mathcal{D}_m) - H(\nu, \mathcal{D}_m)| = O_d(1)$.*
(5) *If $C^{-1} \leq m'/m \leq C$, then $|H(\mu, \mathcal{D}_m) - H(\mu, \mathcal{D}_{m'})| \leq O_{C,d}(1)$.*

Recall that the total variation distance between $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ is

$$\|\mu - \nu\| = \sup_A |\mu(A) - \nu(A)|,$$

where the supremum is over Borel sets A. This is a complete metric on $\mathcal{P}(\mathbb{R}^d)$. It follows from standard measure theory that for every $\varepsilon > 0$, there is a $\delta > 0$ such that if $\|\mu - \nu\| < \delta$, then there are probability measures τ, μ', ν' such that $\mu = (1 - \varepsilon)\tau + \varepsilon\mu'$ and $\nu = (1 - \varepsilon)\tau + \varepsilon\nu'$. Combining this with Lemma 3.1(5) and (6), we have

LEMMA 3.3. *For every $\varepsilon > 0$, there is a $\delta > 0$ such that if $\mu,\nu \in \mathcal{P}(\mathbb{R}^d)$ and $\|\mu - \nu\| < \delta$, then for any finite partition $\mathcal{A}$ of $\mathbb{R}^d$ with k elements,*

$$|H(\mu,\mathcal{A}) - H(\nu,\mathcal{A})| < \varepsilon \log k + H(\varepsilon).$$

In particular, if $\mu,\nu \in \mathcal{P}([0,1]^d)$, then

$$|H_m(\mu) - H_m(\nu)| < d\varepsilon + \frac{H(\varepsilon)}{m}.$$

3.2. *Global entropy from local entropy.* Recall from Section 2.2 the definition of the raw and rescaled components $\mu_{x,n}$, $\mu^{x,n}$, and note that

$$H(\mu^{x,n},\mathcal{D}_m) = H(\mu_{x,n},\mathcal{D}_{n+m}). \tag{18}$$

Also, note that

$$\begin{aligned}\mathbb{E}_{i=n}\left(H_m(\mu^{x,i})\right) &= \int \frac{1}{m} H(\mu^{x,n},\mathcal{D}_m)\,d\mu(x)\\ &= \frac{1}{m}\int H(\mu_{x,n},\mathcal{D}_{n+m})\,d\mu(x)\\ &= \frac{1}{m}\sum_{D\in\mathcal{D}_n} \mu(D) H(\mu_D,\mathcal{D}_{m+n})\\ &= \frac{1}{m} H(\mu,\mathcal{D}_{n+m}\,|\,\mathcal{D}_n).\end{aligned}$$

LEMMA 3.4. *For $r \geq 1$ and $\mu \in \mathcal{P}([-r,r]^d)$ and integers $m < n$,*

$$H_n(\mu) = \mathbb{E}_{0\le i\le n}\left(H_m(\mu^{x,i})\right) + O\left(\frac{m}{n} + \frac{\log r}{n}\right).$$

Proof. By the paragraph before the lemma, the statement is equivalent to

$$H_n(\mu) = \frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{m}H(\mu,\mathcal{D}_{i+m}|\mathcal{D}_i) + O\left(\frac{m}{n} + \frac{\log r}{n}\right).$$

At the cost of adding $O(m/n)$ to the error term we can delete up to m terms from the sum. Thus without loss of generality we may assume that $n/m \in \mathbb{N}$. When $m = 1$, iterating the conditional entropy formula and using $H(\mu,\mathcal{D}_0) = O(\log r)$ gives

$$\sum_{i=0}^{n-1} H(\mu,\mathcal{D}_{i+1}\,|\,\mathcal{D}_i) = H(\mu,\mathcal{D}_n|\mathcal{D}_0) = H(\mu,\mathcal{D}_n) - O(\log r).$$

The result follows on dividing by n. For general m, first decompose the sum according to the residue class of $i \bmod m$ and apply the above to each one:

$$\begin{aligned}\sum_{i=0}^{n-1}\frac{1}{m}H(\mu,\mathcal{D}_{i+m}\,|\,\mathcal{D}_i) &= \frac{1}{m}\sum_{p=0}^{m-1}\left(\sum_{k=0}^{n/m-1} H(\mu,\mathcal{D}_{(k+1)m+p}\,|\,\mathcal{D}_{km+p})\right)\\ &= \frac{1}{m}\sum_{p=0}^{m-1} H(\mu,\mathcal{D}_{n+p}\,|\,\mathcal{D}_p).\end{aligned}$$

Dividing by n, the result follows from the bound

$$\left|\frac{1}{n}H(\mu,\mathcal{D}_{n+p}|\mathcal{D}_p)-H_n(\mu)\right|<\frac{2m+O(\log r)}{n},$$

which can be derived from the identities

$$\begin{aligned}H(\mu,\mathcal{D}_n)+H(\mu,\mathcal{D}_{n+p}|\mathcal{D}_n)&=H(\mu,\mathcal{D}_{n+p})\\&=H(\mu,\mathcal{D}_p)+H(\mu,\mathcal{D}_{n+p}|\mathcal{D}_p)\end{aligned}$$

together with the fact that $H(\mu,\mathcal{D}_p)\le p+\log r$ and $H(\mu,\mathcal{D}_{rm+p}|\mathcal{D}_{rm})\le p$, and recalling that $0\le p<m$. □

We have a similar lower bound for the entropy of a convolution in terms of convolutions of its components at each level.

LEMMA 3.5. *Let $r>0$ and $\mu,\nu\in\mathcal{P}([-r,r]^d)$. Then for $m<n\in\mathbb{N}$,*

$$\begin{aligned}H_n(\mu*\nu)&\ge\mathbb{E}_{0\le i\le n}\left(\frac{1}{m}H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_{i+m}|\mathcal{D}_i)\right)+O\left(\frac{m+\log r}{n}\right)\\&\ge\mathbb{E}_{0\le i\le n}\left(H_m(\mu^{x,i}*\nu^{y,i})\right)+O\left(\frac{1}{m}+\frac{m}{n}+\frac{\log r}{n}\right).\end{aligned}$$

Proof. As in the previous proof, by introducing an error of $O(m/n)$ we can assume that m divides n, and by the conditional entropy formula,

$$\begin{aligned}H(\mu*\nu,\mathcal{D}_n)&=\sum_{k=0}^{n/m-1}H(\mu*\nu,\mathcal{D}_{(k+1)m}|\mathcal{D}_{km})+H(\mu*\nu,\mathcal{D}_0)\\&=\sum_{k=0}^{n/m-1}H(\mu*\nu,\mathcal{D}_{(k+1)m}|\mathcal{D}_{km})+O(\log r)\end{aligned}$$

since $\mu*\nu$ is supported on $[-2r,2r]^d$. Apply the linear map $(x,y)\mapsto x+y$ to the trivial identity $\mu\times\nu=\mathbb{E}_{i=k}(\mu_{x,i}\times\nu_{y,i})$ (Lemma (2.4) for the product measure). We obtain the identity $\mu*\nu=\mathbb{E}_{i=k}(\mu_{x,i}*\nu_{xy,i})$. By concavity of conditional entropy (Lemma 3.1 (5)),

$$\begin{aligned}H(\mu*\nu,\mathcal{D}_n)&=\sum_{k=0}^{n/m-1}H\left(\mathbb{E}_{i=km}(\mu_{x,i}*\nu_{xy,i}),\mathcal{D}_{(k+1)m}|\mathcal{D}_{km}\right)+O(\log r)\\&\ge\sum_{k=0}^{n/m-1}\mathbb{E}_{i=km}\left(H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_{(k+1)m}|\mathcal{D}_{km})\right)+O(\log r).\end{aligned}$$

Dividing by n, we have shown that

$$H_n(\mu*\nu)\ge\frac{m}{n}\sum_{k=0}^{n/m-1}\mathbb{E}_{i=k}\left(H_m(\mu^{x,i}*\nu^{xy,i})\right)+O\left(\frac{m}{n}+\frac{\log r}{n}\right).$$

Now do the same for the sum $k = p$ to $n/m+p$ for $p = 0, 1, \dots, m-1$. Averaging the resulting expressions gives the first inequality. The second inequality follows from the first using

$$\begin{aligned} H(\mu_{x,i} * \nu_{x,i}, \mathcal{D}_{(k+1)m}|\mathcal{D}_{km}) &= H(\mu^{x,i} * \nu^{y,i}, \mathcal{D}_m|\mathcal{D}_0) \\ &= H(\mu^{x,i} * \nu^{y,i}, \mathcal{D}_m) + O(1) \\ &= mH_m(\mu^{x,i} * \nu^{y,i}) + O(1), \end{aligned}$$

where the $O(1)$ error term arises because $\mu^{x,i} * \nu^{x,i}$ is supported on $[0,2)^d$ and hence meets $O(1)$ sets in $\mathcal{D}_0$. □

3.3. *Covering lemmas.* We will require some simple combinatorial lemmas.

Lemma 3.6. *Let $I \subseteq \{0, \dots, n\}$ and $m \in \mathbb{N}$ be given. Then there is a subset $I' \subseteq I$ such that $I \subseteq I' + [0, m]$ and $[i, i+m] \cap [j, j+m] = \emptyset$ for distinct $i, j \in I'$.*

Proof. Define I' inductively. Begin with $I' = \emptyset$ and, at each successive stage, if $I \setminus \bigcup_{i \in I'}[i, i+m] \neq \emptyset$, then add its least element to I'. Stop when $I \subseteq \bigcup_{i \in I'}[i, i+m]$. □

Lemma 3.7. *Let $I, J \subseteq \{0, \dots, n\}$ and $m \in \mathbb{N}$, $\delta > 0$. Suppose that $|[i, i+m] \cap J| \geq (1-\delta)m$ for $i \in I$. Then there is a subset $J' \subseteq J$ such that $|J' \cap (J' - \ell)| \geq (1 - \delta - \frac{\ell}{m})|I|$ for $0 \leq \ell \leq m$.*

Proof. Let $I' \subseteq I$ be the collection obtained by applying the previous lemma to I,m. Let $J' = J \cap (\bigcup_{i \in I'}[i, i+m])$. Then

$$J' \cap (J' - \ell) \supseteq J \cap \bigcup_{i \in I'} ([i, i+m] \cap [i - \ell, i + m - \ell]) = \bigcup_{i \in I'} (J \cap [i, i + m - \ell]).$$

Also, $|J \cap [i, i+m-\ell]| \geq (1 - \delta - \frac{\ell}{m})m$ for $i \in I'$, and $I \subseteq \bigcup_{i \in I'}[i, i+m]$, so by the above,

$$|J' \cap (J' - \ell)| \geq \left(1 - \delta - \frac{\ell}{m}\right) \cdot |\bigcup_{i \in I'}[i, i+m]| \geq \left(1 - \delta - \frac{\ell}{m}\right)|I|. \qquad \square$$

Lemma 3.8. *Let m, δ be given, and let I_1, J_1 and I_2, J_2 be two pairs of subsets of $\{0, \dots, n\}$ satisfying the assumptions of the previous lemma. Suppose also that $I_1 \cap I_2 = \emptyset$. Then there exist $J_1' \subseteq J_1$ and $J_2' \subseteq J_2$ with $J_1' \cap J_2' = \emptyset$ and such that $|J_1' \cup J_2'| \geq (1-\delta)^2 |I_1 \cup I_2|$.*

Proof. Define $I_1' \subseteq I_1$ and $J_1' = J_1 \cap \bigcup_{i \in I'}[i, i+m]$ as in the previous proof, so taking $\ell = 0$ in its conclusion, $|J_1'| \geq (1-\delta)|I_1|$. Let $U = \bigcup_{i \in I_1'}[i, i+m]$, and recall that $|J_1'| = |U \cap J_1| \geq (1-\delta)|U|$. Since $I_1 \subseteq U$ and $I_1 \cap I_2 = \emptyset$,

$$|J_1' \cap I_2| \leq |U| - |I_1| \leq \frac{1}{1-\delta}|J_1'| - |I_1|.$$

Hence, using $|J_1'| \geq (1-\delta)|I_1|$,

$$\begin{aligned}|J_1' \cup I_2| &= |J_1'| + |I_2| - |J_1' \cap I_2| \\ &\geq |J_1'| + |I_2| - \left(\frac{1}{1-\delta}|J_1'| - |I_1|\right) \\ &\geq |I_1| + |I_2| - \frac{\delta}{1-\delta}|J_1'| \\ &\geq (1-\delta)|I_1| + |I_2|.\end{aligned}$$

Now perform the analysis above with $I_2 \setminus J_1', J_2$ in the role of I_1, J_1 and with J_1' in the role of I_2. (Thus $(I_2 \setminus J_1') \cap J_1' = \emptyset$ as required.) We obtain $J_2' \subseteq J_2$ such that

$$\begin{aligned}|J_2' \cup J_1'| &\geq (1-\delta)|I_2 \setminus J_1'| + |J_1'| \\ &= (1-\delta)|J_1' \cup I_2|.\end{aligned}$$

Substituting the previous bound $|J_1' \cup I_2| \geq (1-\delta)|I_1| + |I_2|$ gives the claim, except for disjointness of J_1', J_2', but clearly if they are not disjoint, we can replace J_1' with $J_1' \setminus J_2$. □

3.4. *Atomicity and uniformity of components.* We need to know that almost-atomicity and almost-uniformity passes to component measures. It will be convenient to replace the notion of ε-atomic measures, introduced in Section 2.3, with one that is both stronger and more convenient to work with.

Definition 3.9. A measure $\mu \in \mathcal{P}([0,1])$ is (ε, m)-*atomic* if $H_m(\mu) < \varepsilon$.

Recall that $H_m(\mu) = 0$ if and only if μ is supported on a single interval $I \in \mathcal{D}_m$ of length 2^{-m}. Thus, by continuity of the entropy function $(p_i) \mapsto -\sum p_i \log p_i$, if ε is small compared to m, then any (ε, m)-atomic measure is 2^{-m}-atomic. The reverse implication is false: indeed, a measure may be ε-atomic for arbitrarily small ε and at the same time have its mass divided evenly between two (adjacent) intervals $I, I' \in \mathcal{D}_m$, in which case $H_m(\mu) = \frac{1}{m}$. Thus, for ε small compared to m, the most one can say in general about an ε-atomic measure is that it is $(\frac{1}{m}, m)$-atomic. Thus the definition above is slightly stronger.

LEMMA 3.10. *If $\mu \in \mathcal{P}([0,1])$ is (ε, m)-atomic, then for $k < m$,*

$$\mathbb{P}_{0 \leq i \leq m}\left(\mu^{x,i} \text{ is } (\varepsilon', k)\text{-atomic}\right) > 1 - \varepsilon'$$

for $\varepsilon' = \sqrt{\varepsilon + O(\frac{k}{m})}$.

Proof. By Lemma 3.4,

$$\mathbb{E}_{0 \leq i \leq m}(H_k(\mu^{i,x})) \leq H_m(\mu) + O\left(\frac{k}{m}\right) < \varepsilon + O\left(\frac{k}{m}\right).$$

Since $H_k(\mu^{i,x}) \geq 0$, the claim follows by Markov's inequality. □

LEMMA 3.11. *If $\mu \in \mathcal{P}([0,1])$ is (ε, n)-uniform, then for every $1 \leq m < n$,*

$$\mathbb{P}_{0 \leq i \leq n}\left(\mu^{x,i} \text{ is } (\varepsilon', m)\text{-uniform}\right) > 1 - \varepsilon',$$

where $\varepsilon' = \sqrt{\varepsilon + O(\frac{m}{n})}$.

Proof. The proof is the same as the previous lemma and we omit it. □

We also will repeatedly use the following consequence of Chebychev's inequality.

LEMMA 3.12. *Suppose that $\mathcal{A} \subseteq \mathcal{P}([0,1])$ and that*

$$\mathbb{P}_{0 \leq i \leq n}(\mu^{x,i} \in \mathcal{A}) > 1 - \varepsilon.$$

Then there is a subset $I \subseteq \{0, \ldots, n\}$ with $|I| > (1 - \sqrt{\varepsilon})n$ and

$$\mathbb{P}_{i=q}(\mu^{x,i} \in \mathcal{A}) > 1 - \sqrt{\varepsilon} \qquad \text{for } q \in I.$$

Proof. Consider the function $f : \{0, \ldots, m\} \to [0,1]$ given by $f(q) = \mathbb{P}_{i=q}(\mu^{x,i} \in \mathcal{A})$. By assumption, $\mathbb{E}_{0 \leq q \leq n}(f(q)) > 1 - \varepsilon$. By Chebychev's inequality, there is a subset $I \subseteq \{0, \ldots, n\}$ with $|I| \geq (1 - \sqrt{\varepsilon})n$ and $f(q) > 1 - \sqrt{\varepsilon}$ for $q \in I$, as desired. □

4. **Convolutions**

4.1. *The Berry-Esseen theorem and an entropy estimate.* For $\mu \in \mathcal{P}(\mathbb{R})$, let $m(\mu)$ denote the mean, or barycenter, of μ, given by

$$\langle \mu \rangle = \int x \, d\mu(x),$$

and let $\mathrm{Var}(\mu)$ denote its variance:

$$\mathrm{Var}(\mu) = \int (x - \langle \mu \rangle)^2 \, d\mu(x).$$

Recall that if $\mu_1, \ldots, \mu_k \in \mathcal{P}(\mathbb{R})$, then $\mu = \mu_1 * \cdots * \mu_k$ has mean $\langle \mu \rangle = \sum_{i=1}^{k} \langle \mu_i \rangle$ and $\mathrm{Var}(\mu) = \sum_{i=1}^{k} \mathrm{Var}(\mu_i)$.

The Gaussian with mean m and variance σ^2 is given by $\gamma_{m,\sigma^2}(A) = \int_A \varphi((x - m)/\sigma^2) dx$, where $\varphi(x) = \sqrt{2\pi} \exp(-\frac{1}{2}|x|^2)$. The central limit theorem asserts that, for $\mu_1, \mu_2, \cdots \in \mathcal{P}(\mathbb{R}^d)$ of positive variance, the convolutions $\mu_1 * \cdots * \mu_k$ can be rescaled so that the resulting measure is close in the weak sense to a Gaussian measure. The Berry-Esseen inequalities quantify the rate of this convergence. We use the following variant from [8].

THEOREM 4.1. *Let $\mu_1, \ldots, \mu_k$ be probability measures on $\mathbb{R}$ with finite third moments $\rho_i = \int |x|^3 \, d\mu_i(x)$. Let $\mu = \mu_1 * \cdots * \mu_k$, and let γ be the*

Gaussian measure with the same mean and variance as μ. *Then*[8] *for any interval* $I \subseteq \mathbb{R}$,

$$|\mu(I) - \gamma(I)| \leq C_1 \cdot \frac{\sum_{i=1}^{k} \rho_i}{\mathrm{Var}(\mu)^{3/2}},$$

where $C_1 = C_1(d)$. *In particular, if* $\rho_i \leq C$ *and* $\sum_{i=1}^{k} \mathrm{Var}(\mu_i) \geq ck$ *for constants* $c, C > 0$, *then*

$$|\mu(I) - \gamma(I)| = O_{c,C}(k^{-1/2}).$$

4.2. *Multiscale analysis of repeated self-convolutions.* In this section we show that for any measure μ, every $\delta > 0$, every integer scale $m \geq 2$, and appropriately large k, the following holds. Typical levels-i components of the convolution μ^{*k} are (δ, m)-uniform, unless in μ the level-i components are typically (δ, m)-atomic. The main idea is to apply the Berry-Esseen theorem to convolutions of component measures.

PROPOSITION 4.2. *Let* $\sigma > 0$, $\delta > 0$, *and let* $m \geq 2$ *be an integer. Then there exists an integer* $p = p_0(\sigma, \delta, m)$ *such that for all* $k \geq k_0(\sigma, \delta, m)$, *the following holds. Let* $\mu_1, \ldots, \mu_k \in \mathcal{P}([0,1])$, *let* $\mu = \mu_1 * \cdots * \mu_k$, *and suppose that* $\mathrm{Var}(\mu) \geq \sigma k$. *Then*

$$\mathbb{P}_{i=p-[\log\sqrt{k}]}\left(\mu^{x,i} \textit{ is } (\delta, m)\textit{-uniform}\right) > 1 - \delta. \tag{19}$$

Note that $p - [\log \sqrt{k}]$ will generally be negative. Dyadic partitions of level q with $q < 0$ are defined in the same manner as for positive q; that is, by $\mathcal{D}_{-q} = \{[r2^{-q}, (r+1)2^{-q})\}_{r \in \mathbb{Z}}$. For $q < 0$, this partition consists of intervals of length is $2^{|q|}$ with integer endpoints. Thus, the conclusion of the proposition concerns the μ-probabilities of nearby intervals of length $O_p(\sqrt{k}) = O_{\sigma,\delta,m}(\sqrt{k})$ (since $p = p_0(\sigma, \delta, m)$). This is the natural scale at which we can expect to control such probabilities: indeed, μ is close to a Gaussian γ of variance σk, but only in the sense that for any c, if k is large enough, μ and γ closely agree on the mass that they give to intervals of length $c\sqrt{\mathrm{Var}(\mu)} = c\sqrt{k}$.

Proof. Let us first make some elementary observations. Suppose that $\gamma \in \mathcal{P}(\mathbb{R})$ is a probability measure with continuous density function f and $x \in \mathbb{R}$ is such that $f(x) \neq 0$. Since $\gamma(I) = \int_I f(y)dy$, for any interval I, we

[8]In the usual formulation one considers the measure μ' defined by scaling μ by $\mathrm{Var}(\mu)$, and γ' the Gaussian with the same mean and variance $1 = \mathrm{Var}(\mu')$, and gives a similar bound for $|\mu'(J) - \gamma'(J)|$ as J ranges over intervals. The two formulations are equivalent since $\mu(I) - \gamma(I) = \mu'(J) - \gamma'(J)$, where J is an interval depending in the obvious manner on I, and $I \to J$ is a bijection.

have

$$\left|\frac{\gamma(I)}{|I|} - f(x)\right| \leq \sup_{z\in I} |f(x) - f(z)|,$$

where $|I|$ is the length of I. By continuity, the right-hand side tends to 0 uniformly as the endpoints of I approach x. In particular, if n is large enough, for any $I \subseteq \mathcal{D}_n(x)$, the ratio $\frac{\gamma(x)}{|I|}$ will be arbitrarily close to $f(x)$. Therefore, since $f(x) \neq 0$, for any fixed m, if n is large enough, then $|\frac{\gamma(I)}{\gamma(J)} - 1| = |\frac{\gamma(I)/|I|}{\gamma(J)/|J|} - 1|$ for all intervals $I, J \in \mathcal{D}_{n+m}$ with $I, J \subseteq \mathcal{D}_n(x)$. In other words, the distribution of $\gamma^{x,n}$ on the level-m dyadic subintervals of $[0, 1)$ approaches the uniform one as $n \to \infty$. Now,

$$H_m(\mu^{x,n}) = - \sum_{I\in\mathcal{D}_{n+m}, I\subseteq\mathcal{D}_n(x)} \mu(I) \log \mu(I),$$

and the function $t \log t$ is continuous for $t \in (0, 1)$. Therefore, writing u for the uniform measure on $[0, 1)$, we conclude that

$$\lim_{n\to\infty} H_m(\gamma^{x,n}) = H_m(u) = 1.$$

This in turn implies that $\mathbb{E}_{i=p}(H_m(\gamma^{x,p})) \to 1$ as $p \to \infty$. Finally, the rate of convergence in the limits above is easily seen to depend only on the value $f(x)$ and the modulus of continuity of f at x.

Fix $0 < \sigma, \delta < 1$, and consider the family $\mathcal{G}$ of Gaussians with mean 0 and variance in the interval $[\sigma, 1]$. For every interval $I = [-R, R]$, the restriction to I of the density functions of measures in $\mathcal{G}$ form an equicontinuous family. Also, by choosing a large enough R we can ensure that $\inf_{g\in\mathcal{G}} \gamma([-R, R])$ is arbitrarily close to 1. Therefore, by the previous discussion, there is a $p = p_0(\sigma, \delta, m)$ such that $\mathbb{P}_{i=p}(H_m(\gamma^{x,i}) > 1 - \delta) > 1 - \delta$ for all $\gamma \in \mathcal{G}$.

Now, if μ_i and μ are as in the statement and μ' is μ scaled by $2^{-[\log\sqrt{k}]}$ (which is up to a constant factor the same as $1/\sqrt{k}$), then by the Berry-Esseen theorem (Theorem 4.1), μ' agrees with the Gaussian of the same mean and variance on intervals of length 2^{-p-m} to a degree that can be made arbitrarily small by making k large in a manner depending on σ, p. In particular, for large enough k, this guarantees that $\mathbb{P}_{i=p}(H_m((\mu')^{x,i}) > 1 - \delta) > 1 - \delta$.

All that remains is to adjust the scale by a factor of $2^{[\log\sqrt{k}]}$. Then the same argument applied to μ instead of the scaled μ' gives $\mathbb{P}_{i=p-[\log\sqrt{k}]}(H_m((\mu)^{x,i}) > 1 - \delta) > 1 - \delta$, which is (19). □

We turn to repeated self-convolutions.

PROPOSITION 4.3. *Let $\sigma, \delta > 0$, and let $m \geq 2$ be an integer. Then there exists $p = p_1(\sigma, \delta, m)$ such that for sufficiently large $k \geq k_1(\sigma, \delta, m)$, the following holds. Let $\mu \in \mathcal{P}([0, 1])$, fix an integer $i_0 \geq 0$, and write*

$$\lambda = \mathbb{E}_{i=i_0}\left(\mathrm{Var}(\mu^{x,i})\right).$$

If $\lambda > \sigma$, then for $j_0 = i_0 - [\log \sqrt{k}] + p$ and $\nu = \mu^{*k}$, we have

$$\mathbb{P}_{j=j_0}\left(\nu^{x,j} \text{ is } (\delta, m)\text{-uniform}\right) > 1 - \delta.$$

Proof. Let μ, λ and m be given. Fix p and k. (We will later see how large they must be.) Let i_0 be as in the statement and $j_0 = i_0 - [\log \sqrt{k}] + p$.

Let $\widetilde{\mu}$ denote the k-fold self-product $\widetilde{\mu} = \mu \times \cdots \times \mu$, and let $\pi : (\mathbb{R})^k \to \mathbb{R}$ denote the addition map

$$\pi(x_1, \ldots, x_k) = \sum_{i=1}^{k} x_i.$$

Then $\nu = \pi\widetilde{\mu}$ and, since $\widetilde{\mu} = \mathbb{E}_{i=i_0}(\widetilde{\mu}_{x,i})$, we also have by linearity $\nu = \mathbb{E}_{i=i_0}(\pi\widetilde{\mu}_{x,i})$. By concavity of entropy and an application of Markov's inequality, there is a $\delta_1 > 0$, depending only on δ, such that the proposition will follow if we show that with probability $> 1 - \delta_1$ over the choice of the component $\widetilde{\mu}_{x,i_0}$ of $\widetilde{\mu}$, the measure $\eta = \pi\widetilde{\mu}_{x,i_0}$ satisfies

$$\mathbb{P}_{j=j_0}\left(\eta^{y,j} \text{ is } (\delta_1, m)\text{-uniform}\right) > 1 - \delta_1. \tag{20}$$

The random component $\widetilde{\mu}_{x,i_0}$ is itself a product measure $\widetilde{\mu}_{x,i} = \mu_{x_1,i_0} \times \cdots \times \mu_{x_k,i_0}$, and the marginal measures μ_{x_j,i_0} of this product are distributed independently according to the distribution of the raw components of μ at level i_0. Note that these components differ from the rescaled components by a scaling factor of 2^{i_0}, so the expected variance of the raw components is $2^{-2i_0}\lambda$. Recall that

$$\mathrm{Var}(\pi(\mu_{x_1,i_0} \times \cdots \times \mu_{x_k,i_0})) = \sum_{j=1}^{k} \mathrm{Var}(\mu_{x_j,i_0}).$$

Thus for any $\delta_2 > 0$, by the weak law of large numbers, if k is large enough in a manner depending on δ_2, then with probability $> 1 - \delta_2$ over the choice of $\widetilde{\mu}_{x,i_0}$ we will have[9]

$$\left| \frac{1}{k} \mathrm{Var}(\pi\widetilde{\mu}_{x,i_0}) - 2^{-2i_0}\lambda \right| < 2^{-2i_0}\delta_2. \tag{21}$$

We can choose δ_2 small in a manner depending on σ, so (21) implies

$$\mathrm{Var}(\pi\widetilde{\mu}_{x,i_0}) > 2^{-2i_0} \cdot k\sigma/2. \tag{22}$$

But now inequality (20) follows from an application of Proposition 4.2 with proper choice of parameters. □

[9] We use here the fact that we have a uniform bound for the rate of convergence in the weak law of large numbers for i.i.d. random variables $X_1, X_2, \ldots$. In fact, the rate can be bounded in terms of the mean and variance of X_1. Here X_1 is distributed like the variance $\mathrm{Var}(\mu_{x,i_0})$ of a random component of level i_0, and the mean and variance of X_1 are bounded independently of $\mu \in \mathcal{P}([0,1])$.

LEMMA 4.4. *Fix $m \in \mathbb{N}$. If* $\mathrm{Var}(\mu)$ *is small enough, then* $H_m(\mu) \leq \frac{2}{m}$. *If* $H_m(\mu)$ *is small enough, then* $\mathrm{Var}(\mu) < 2^{-m}$.

Proof. If $\mathrm{Var}(\mu)$ is small, then most of the μ-mass sits on an interval of length 2^{-m}, hence on at most two intervals from $\mathcal{D}_m$, so $H_m(\mu)$ is roughly $\frac{1}{m}$ (certainly $< \frac{2}{m}$). Conversely, if $H_m(\mu)$ is small, then most of the μ-mass sits on one interval from $\mathcal{D}_m$, whose length is 2^{-m}, so $\mathrm{Var}(\mu)$ is of this order. □

Recall Definitions 2.6 and 3.9.

COROLLARY 4.5. *Let $m \in \mathbb{N}$ and $\varepsilon > 0$. For $N > N(m, \varepsilon)$ and $0 < \delta < \delta(m, \varepsilon, N)$, if $\mu \in \mathbb{P}([0,1])$ and* $\mathrm{Var}(\mu) < \delta$, *then*

$$\mathbb{P}_{0 \leq i \leq N}(\mathrm{Var}(\mu^{x,i}) < \varepsilon \text{ and } \mu^{x,i} \text{ is } (\varepsilon, m)\text{-atomic}) > 1 - \varepsilon.$$

Proof. Using the previous lemma choose m', ε' such that $H_{m'}(\theta) < \varepsilon'$ implies $\mathrm{Var}(\theta) < \varepsilon$. Then it suffices to find N, δ such that $\mathrm{Var}(\mu) < \delta$ implies

$$\mathbb{P}_{0 \leq i \leq N}(H_{m'}(\mu^{x,i}) < \varepsilon' \text{ and } H_m(\mu^{x,i}) < \varepsilon) > 1 - \varepsilon.$$

By Lemma 3.10 (applied twice), if $\varepsilon'' > 0$ is small enough, then for large enough N, the last inequality follows from $H_N(\mu) < \varepsilon''$. Finally, by the last lemma again, if N is large enough, this follows from $\mathrm{Var}(\mu) < \delta$ if δ is sufficiently small. □

THEOREM 4.6. *Let $\delta > 0$, and let $m \geq 2$ be an integer. Then for $k \geq k_2(\delta, m)$ and all sufficiently large $n \geq n_2(\delta, m, k)$, the following holds. For any $\mu \in \mathcal{P}([0,1])$, there are disjoint subsets $I, J \subseteq \{1, \ldots, n\}$ with $|I \cup J| > (1-\delta)n$ such that, writing $\nu = \mu^{*k}$,*

$$\mathbb{P}_{i=q}\left(\nu^{x,i} \text{ is } (\delta, m)\text{-uniform}\right) \geq 1 - \delta \quad \text{for } q \in I, \tag{23}$$

$$\mathbb{P}_{i=q}\left(\mu^{x,i} \text{ is } (\delta, m)\text{-atomic}\right) \geq 1 - \delta \quad \text{for } q \in J. \tag{24}$$

Proof. Let δ and $m \geq 0$ be given. We may assume $\delta < 1/2$. The proof is given in terms of a function $\widetilde{\rho} : (0,1] \to (0,1]$ with $\widetilde{\rho}(\sigma)$ depending on σ, δ, m. The exact requirements will be given in the course of the proof. The definition of $\widetilde{\rho}$ uses the functions $k_1(\cdot)$, and $p_1(\cdot)$ from Proposition 4.3 and we assume, without loss of generality, that these functions are monotone in each of their arguments.

Our first requirement of $\widetilde{\rho}$ will be that $\widetilde{\rho}(\sigma) < \sigma$. Consider the decreasing sequence $\sigma_0 > \sigma_1 > \cdots$ defined by $\sigma_0 = 1$ and $\sigma_i = \widetilde{\rho}(\sigma_{i-1})$. Assume that $k \geq k_1(\sigma_{\lceil 1+2/\delta \rceil}, \delta, m)$; this expression can be taken for $k_2(\delta, m)$.

Fix μ and n large; we shall later see how large an n is desirable. For $0 \leq q \leq n$, write

$$\lambda_q = \mathbb{E}_{i=q}\left(\mathrm{Var}(\mu^{x,i})\right).$$

Since the intervals $(\sigma_i, \sigma_{i-1}]$ are disjoint, there is an integer $1 \leq s \leq 1 + \frac{2}{\delta}$ such that $\mathbb{P}_{0 \leq q \leq n}(\lambda_q \in (\sigma_s, \sigma_{s-1}]) < \frac{\delta}{2}$. For this s, define

$$\sigma = \sigma_{s-1},$$
$$\rho = \widetilde{\rho}(\sigma) = \sigma_s$$

and set

$$I' = \{0 \leq q \leq n \,:\, \lambda_q > \sigma\},$$
$$J' = \{0 \leq q \leq n \,:\, \lambda_q < \rho\}.$$

Then by our choice of s,

$$|I' \cup J'| > \left(1 - \frac{\delta}{2}\right) n. \tag{25}$$

Let $\ell \geq 0$ be the integer

$$\ell = [\log \sqrt{k}] - p_1(\sigma, \delta, m).$$

Since we may take n large relative to ℓ, by deleting at most ℓ elements of I' we can assume that $I' \subseteq [\ell, n]$ and that (25) remain valid. Let

$$I = I' - \ell.$$

Since $k \geq k_1(\sigma, \delta, m)$, by our choice of parameters and the previous proposition,

$$\mathbb{P}_{i=q}\left(\nu^{x,i} \text{ is } (\delta, m)\text{-uniform}\right) > 1 - \delta \qquad \text{for } q \in I,$$

which is (23).

We now turn to the slightly harder task of choosing n (i.e., determining the appropriate condition $n \geq n_2$). By definition of J',

$$\mathbb{E}_{i=q}\left(\mathrm{Var}(\mu^{x,i})\right) = \lambda_q < \rho \qquad \text{for } q \in J'.$$

This and Markov's inequality imply

$$\mathbb{P}_{i=q}\left(\mathrm{Var}(\mu^{x,i}) < \sqrt{\rho}\right) > 1 - \sqrt{\rho} \qquad \text{for } q \in J'. \tag{26}$$

Fix a small number $\rho' = \rho'(\delta, \sigma)$ and a large integer $N = N(\ell, \delta, \rho')$ upon which we place constraints in due course. Since we can take n large relative to N, we can assume $I', J' \subseteq \{\ell, \ldots, n - N\}$ without affecting the size bounds. Assuming ρ is small enough, Corollary 4.5 tells us that any measure $\theta \in \mathcal{P}([0,1])$ satisfying $\mathrm{Var}(\theta) < \sqrt{\rho}$ also satisfies

$$\mathbb{P}_{0 \leq i \leq N}\left(\mathrm{Var}(\theta^{y,i}) < \sigma \text{ and } \theta^{y,i} \text{ is } (\delta, m)\text{-atomic}\right) > 1 - \rho'.$$

Assuming again that $\sqrt{\rho} < \rho'$, the last equation and (26) give

$$\begin{aligned}\mathbb{P}_{q \leq i \leq q+N}\left(\mathrm{Var}(\mu^{x,i}) < \sigma \text{ and } \mu^{x,i} \text{ is } (\delta, m)\text{-atomic}\right) &> (1 - \sqrt{\rho})(1 - \rho') \\ &> 1 - 2\rho' \qquad \text{for } q \in J'.\end{aligned}$$

Let

$$U = \left\{ q \in \mathbb{N} \,:\, \mathbb{P}_{i=q}(\mathrm{Var}(\theta^{y,i}) < \frac{\sigma}{2} \text{ and } \theta^{y,i} \text{ is } (\delta, m)\text{-atomic}) > 1 - \sqrt{2\rho'} \right\}.$$

By Lemma 3.12 (i.e., Chebychev's inequality),

$$|U \cap [q, q+N]| \geq (1 - \sqrt{2\rho'})N \qquad \text{for } q \in J'.$$

Apply Lemma 3.7 to J' and U to obtain $U' \subseteq U$ satisfying $|U'| > (1-\sqrt{2\rho'})|J'|$ and $|U' \cap (U' - \ell)| > (1 - 2\sqrt{2\rho'} - \frac{\ell}{N})|U'|$. Defining

$$J = U' \cap (U' - \ell)$$

and assuming that $\frac{\ell}{N} < 2\sqrt{\rho'}$, we conclude that

$$|J| \geq (1 - 3\sqrt{2\rho'})|J'|.$$

We claim that $I \cap J = \emptyset$. Indeed, suppose $q \in I \cap J$. Then $q + \ell \in I'$, so $\lambda_{q+\ell} \geq \sigma$. On the other hand, $q \in J \subseteq U' - \ell$ implies $q + \ell \in U' \subseteq U$, so by definition of U and assuming that $3\sqrt{3\rho'} < \sigma$,

$$\begin{aligned}
\lambda_{q+\ell} &= \mathbb{E}_{i=q+\ell}(\mathrm{Var}(\mu^{x,i})) \\
&\leq \frac{\sigma}{2} \cdot \mathbb{P}_{i=q+\ell}\left(\mathrm{Var}(\mu^{x,i}) < \frac{\sigma}{2}\right) + 1 \cdot \mathbb{P}_{i=q+\ell}\left(\mathrm{Var}(\mu^{x,i}) \geq \frac{1}{2}\right) \\
&< \frac{\sigma}{2} \cdot 1 + 1 \cdot 3\sqrt{3\rho'} \\
&< \sigma.
\end{aligned}$$

This contradiction shows that $I \cap J = \emptyset$.

Finally, $I' \cap J' = \emptyset$ and $|I' \cup J'| > (1 - \frac{\delta}{2})n$ so, assuming that $3\sqrt{3\rho'} < \delta$,

$$|I \cup J| = |I| + |J| \geq |I| + (1 - 3\sqrt{3\rho'})|J'| > \left(1 - \frac{\delta}{2}\right)|I' \cup J'| > \left(1 - \frac{\delta}{2}\right)^2 n.$$

This completes the proof. □

4.3. *The Kaĭmanovich-Vershik lemma.* The Plünnecke-Ruzsa inequality in additive combinatorics roughly states that if $A, B \subseteq \mathbb{Z}$ and $|A + B| \leq C|A|$, then there is a subset $A_0 \subseteq A$ of size comparable to A such that $|A_0 + B^{\oplus k}| \leq C^k|A|$. The second ingredient in our proof of Theorem 2.7 is the following elegant analog for entropy.

LEMMA 4.7. *Let Γ be a countable abelian group, and let $\mu, \nu \in \mathcal{P}(\Gamma)$ be probability measures with $H(\mu) < \infty$, $H(\nu) < \infty$. Let*

$$\delta_k = H(\mu * (\nu^{*(k+1)})) - H(\mu * (\nu^{*k})).$$

Then δ_k is nonincreasing in k. In particular,

$$H(\mu * (\nu^{*k})) \leq H(\mu) + k \cdot (H(\mu * \nu) - H(\nu)).$$

Lemma 4.7 first appears in a study of random walks on groups by Kaĭmanovich and Vershik [16]. It was more recently rediscovered by Madiman and his co-authors [21], [22] and, in a weaker form, by Tao [35], who later made the connection to additive combinatorics. For completeness, we give the short proof here.

Proof. Let X_0 be a random variable distributed according to μ, let Z_n be distributed according to ν, and let all variables be independent. Set $X_n = X_0 + Z_1 + \cdots + Z_n$, so the distribution of X_n is just $\mu * \nu^{*n}$. Furthermore, since G is abelian, given $Z_1 = g$, the distribution of X_n is the same as the distribution of $X_{n-1} + g$ and hence $H(X_n|Z_1) = H(X_{n-1})$. We now compute

$$\begin{aligned} H(Z_1|X_n) &= H(Z_1, X_n) - H(X_n) \\ &= H(Z_1) + H(X_n|Z_1) - H(X_n) \\ &= H(\nu) + H(\mu * \nu^{*(n-1)}) - H(\mu * \nu^{*n}). \end{aligned} \tag{27}$$

Since X_n is a Markov process, a short calculation shows that $Z_1 = X_1 - X_0$ is independent of X_{n+1} when conditioned on X_n, so

$$H(Z_1 \mid X_n) = H(Z_1 \mid X_n, X_{n+1}) \le H(Z_1 \mid X_{n+1}).$$

Using (27) in both sides of the inequality above, we find that

$$H(\mu * \nu^{*(n-1)}) - H(\mu * \nu^{*n}) \le H(\mu * \nu^{*n}) - H(\mu * \nu^{*(n+1)}),$$

which is the what we claimed. □

For the analogous statement for the scale-n entropy of measures on $\mathbb{R}$, we use a discretization argument. For $m \in \mathbb{N}$, let

$$M_m = \left\{ \frac{k}{2^m} \,:\, k \in \mathbb{Z} \right\}$$

denote the group of 2^m-adic rationals. Each $D \in \mathcal{D}_m$ contains exactly one $x \in M_m$. Define the m-discretization map $\sigma_m : \mathbb{R} \to M_m$ by $\sigma_m(x) = v$ if $\mathcal{D}_m(x) = \mathcal{D}_m(v)$, so that $\sigma_m(x) \in \mathcal{D}_m(x)$.

We say that a measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is m-discrete if it is supported on M_m. For arbitrary μ its m-discretization is its push-forward $\sigma_m \mu$ through σ_m, given explicitly by

$$\sigma_m \mu = \sum_{v \in M_m^d} \mu(\mathcal{D}_m(v)) \cdot \delta_v.$$

Clearly $H_m(\mu) = H_m(\sigma_m \mu)$.

LEMMA 4.8. *Given* $\mu_1, \ldots, \mu_k \in \mathcal{P}(\mathbb{R})$ *with* $H(\mu_i) < \infty$ *and* $m \in \mathbb{N}$,

$$|H_m(\mu_1 * \mu_2 * \cdots * \mu_k) - H_m(\sigma_m \mu_1 * \cdots * \sigma_m \mu_k)| = O(k/m).$$

Proof. Let $\pi : \mathbb{R}^k \to \mathbb{R}$ denote the map $(x_1,\ldots,x_k) \mapsto \sum_{i=1}^k x_i$. Then $\mu_1 * \cdots * \mu_k = \pi(\mu_1 \times \cdots \times \mu_k)$ and $\mu_1^{(m)} * \cdots * \mu_k^{(m)} = \pi \circ \sigma_m^k(\mu_1 \times \cdots \times \mu_k)$. (Here $\sigma_m^k : (x_1,\ldots,x_k) \mapsto (\sigma_m x_1,\ldots,\sigma_m x_k)$.) Now, it is easy to check that

$$|\pi(x_1,\ldots,x_k) - \pi \circ \sigma_m^k(x_1,\ldots,x_k)| = O(k),$$

so the desired entropy bound follows from Lemma 3.2(3). □

Proposition 4.9. *Let $\mu,\nu \in \mathcal{P}(\mathbb{R})$ with $H_n(\mu), H_n(\nu) < \infty$. Then*

$$H_n(\mu * (\nu^{*k})) \le H_n(\mu) + k \cdot (H_n(\mu * \nu) - H_n(\mu)) + O\left(\frac{k}{n}\right). \tag{28}$$

Proof. Writing $\widetilde{\mu} = \sigma_n(\mu)$ and $\widetilde{\nu} = \sigma_n(\nu)$, Theorem 4.7 implies

$$H(\widetilde{\mu} * (\widetilde{\nu}^{*k})) \le H(\widetilde{\mu}) + k \cdot (H(\widetilde{\mu} * \widetilde{\nu}) - H(\widetilde{\nu})).$$

For n-discrete measures, the entropy of the measure coincides with its entropy with respect to $\mathcal{D}_n$, so dividing this inequality by n gives (28) for $\widetilde{\mu}, \widetilde{\nu}$ instead of μ,ν, and without the error term. The desired inequality follows from Lemma 4.8. □

We also will later need the following simple fact.

Corollary 4.10. *For $m \in \mathbb{N}$ and $\mu,\nu \in \mathcal{P}([-r,r]^d)$ with $H_n(\mu), H_n(\nu) < \infty$,*

$$H_m(\mu * \nu) \ge H_m(\mu) - O\left(\frac{1}{m}\right).$$

Proof. This is immediate from the identity $\mu * \nu = \int \mu * \delta_y \, d\nu(y)$, concavity of entropy, and Lemma 3.2(4). (Note that $\mu * \delta_y$ is a translate of μ.) □

4.4. *Proof of the inverse theorem.* Recall Definitions 2.6 and 3.9.

Theorem 4.11. *For every $\varepsilon_1, \varepsilon_2 > 0$ and integers $m_1, m_2 \ge 2$, there exists a $\delta = \delta(\varepsilon_1,\varepsilon_2,m_1,m_2)$ such that for all $n > n(\varepsilon_1,\varepsilon_2,m_1,m_2,\delta)$, if $\nu,\mu \in \mathcal{P}([0,1])$, then either $H_n(\mu * \nu) \ge H_n(\mu) + \delta$, or there exist disjoint subsets $I, J \subseteq \{0,\ldots,n\}$ with $|I \cup J| \ge (1-\varepsilon)n$ and*

$$\mathbb{P}_{i=k}\left(\mu^{x,i} \text{ is } (\varepsilon_1, m_1)\text{-uniform}\right) > 1 - \varepsilon \text{ for } k \in I,$$

$$\mathbb{P}_{i=k}\left(\nu^{x,i} \text{ is } (\varepsilon_2, m_2)\text{-atomic}\right) > 1 - \varepsilon \text{ for } k \in J.$$

Remark. Since, given ε_1, for a suitable choice of ε_2, m_2 any (ε', m')-atomic measure is ε_1-atomic, the statement above implies Theorem 2.7.

Proof. We begin with $\varepsilon_1 = \varepsilon_2 = \varepsilon$ and $m_1 = m_2 = m$ and assume that m is large with respect to ε. (We shall see how large below.) We later explain how to remove this assumption. Choose $k = k_2(\varepsilon, m)$ as in Theorem 4.6, with $\delta = \varepsilon/2$. We shall show that the conclusion holds if n is large relative to the previous parameters.

Let $\mu, \nu \in \mathcal{P}([0,1))$. Denote

$$\tau = \nu^{*k}.$$

Assuming n is large enough, Theorem 4.6 provides us with disjoint subsets $I, J \subseteq \{0, \ldots, n\}$, with $|I \cup J| > (1 - \varepsilon/2)n$ such that

$$\mathbb{P}_{i=k}\left(\tau^{x,i} \text{ is } \left(\frac{\varepsilon}{2}, m\right)\text{-uniform}\right) > 1 - \frac{\varepsilon}{2} \qquad \text{for } k \in I \tag{29}$$

and

$$\mathbb{P}_{i=k}\left(\nu^{x,i} \text{ is } (\varepsilon, m)\text{-atomic}\right) \geq 1 - \frac{\varepsilon}{2} \qquad \text{for } k \in J. \tag{30}$$

Let $I_0 \subseteq I$ denote the set of k such that

$$\mathbb{P}_{i=k}\left(\mu^{x,i} \text{ is } (\varepsilon, m)\text{-uniform}\right) > 1 - \varepsilon \qquad \text{for } k \in I. \tag{31}$$

If $|I_0| > (1-\varepsilon)n$, we are done since by (30) and (31), the pair I_0, J satisfy the second alternative of the theorem.

Otherwise, let $I_1 = I \setminus I_0$, so that $|I_1| = |I| - |I_0| > \varepsilon n/2$. We have

$$\mathbb{P}_{i=k}\left(\tau^{x,i} \text{ is } \left(\frac{\varepsilon}{2}, m\right)\text{-uniform and } \mu^{y,i} \text{ is not } (\varepsilon, m)\text{-uniform}\right) > \frac{\varepsilon}{2} \quad \text{for } k \in I_1.$$

For $\mu^{x,i}, \tau^{y,i}$ in the event above, this just means that $H_m(\tau^{y,i}) > H_m(\mu^{x,i}) + \varepsilon/2$ and hence $H_m(\mu^{x,i} * \tau^{y,i}) \geq H_m(\mu^{x,i}) + \varepsilon/2 - O(1/m)$. For any other pair $\mu^{x,i}, \tau^{y,i}$, we have the trivial bound $H_m(\mu^{x,i} * \tau^{y,i}) \geq H_m(\mu^{x,i}) - O(1/m)$. Thus, using Lemmas 3.4, 3.5, and 4.10,

$$\begin{aligned}
H_n(\mu * \tau) &= \mathbb{E}_{0 \leq i \leq n}(H_m(\mu^{x,i} * \tau^{y,i})) + O\left(\frac{m}{n}\right) \\
&= \frac{|I_1|}{n+1}\mathbb{E}_{i \in I_1}(H_m(\mu^{x,i} * \tau^{y,i})) \\
&\quad + \frac{n+1-|I_1|}{n+1}\mathbb{E}_{i \in I_1^c}(H_m(\mu^{x,i} * \tau^{y,i})) + O\left(\frac{m}{n}\right) \\
&> \frac{|I_1|}{n+1}\left(\mathbb{E}_{i \in I_1}(H_m(\mu^{x,i})) + \left(\frac{\varepsilon}{2}\right)^2\right) \\
&\quad + \frac{n+1-|I_1|}{n+1}\mathbb{E}_{i \in I_1^c}(H_m(\mu^{x,i})) + O\left(\frac{1}{m} + \frac{m}{n}\right) \\
&= \mathbb{E}_{0 \leq i \leq n}(H_m(\mu^{x,i})) + \left(\frac{\varepsilon}{2}\right)^3 + O\left(\frac{1}{m} + \frac{m}{n}\right) \\
&= H_n(\mu) + \left(\frac{\varepsilon}{2}\right)^3 + O\left(\frac{1}{m} + \frac{m}{n}\right).
\end{aligned}$$

So, assuming that ε was sufficiently small to begin with, m large with respect to ε and n large with respect to m, we have

$$H_n(\mu * \tau) > H_n(\mu) + \frac{\varepsilon^3}{10}.$$

On the other hand, by Proposition 4.9,

$$H_n(\mu*\tau) = H_n(\mu*\nu^{*k}) \leq H_n(\mu) + k\cdot(H_n(\mu*\nu) - H_n(\mu)) + O\left(\frac{k}{n}\right).$$

Assuming that n is large enough in a manner depending on ε and k, this and the previous inequality give

$$H_n(\mu*\nu) \geq H_n(\mu) + \frac{\varepsilon^3}{100k}.$$

This is the desired conclusion, with $\delta = \varepsilon^3/100k$.

We now remove the largeness assumption on m. Let ε, m_1, m_2 be given and choose $\varepsilon' > 0$ small compared to ε, and m' appropriately large for ε, m_1, m_2. Applying what we just proved for a large enough n, we obtain corresponding $I, J \subseteq [0,n]$. It will be convenient to denote $U_1 = I$ and $U_2 = J$. Now, for $i \in U_1$, by definition of U_1 and Lemma 3.11, and assuming m_1/m' small enough,

$$\mathbb{P}_{i\leq j\leq i+m'}(\mu^{x,j} \text{ is } (\sqrt{2\varepsilon'}, m_1)\text{-uniform}) > 1-\sqrt{2\varepsilon'}.$$

Thus, assuming as we may that $\varepsilon < \sqrt{2\varepsilon'}$, if we set

$$V_1 = \{j \in [0,n] \,:\, \mathbb{P}_{u=j}(\mu^{x,u} \text{ is } (\varepsilon, m_2)\text{-uniform}) > 1-\varepsilon\},$$

then by Lemma 3.12 (Chebychev's inequality), $|[i, i+m'] \cap V_1| > (1-(2\varepsilon)^{1/4})m'$. Similarly, defining

$$V_2 = \{j \in [0,n] \,:\, \mathbb{P}_{u=j}(\mu^{x,u} \text{ is } (\varepsilon, m)\text{-atomic}) > 1-\varepsilon\}$$

and using Lemma 3.10, if m_2/m is small enough, then $|[j, j+m'] \cap V_2| > (1-(2\varepsilon)^{1/4})m'$ for all for $j \in U_2$. Now, applying Lemma 3.8 to U_1, V_1 and U_2, V_2, we find $U_1' \subseteq U_1$ and $U_2' \subseteq U_2$ as in that lemma. Taking $I' = U_1'$ and $J' = U_2'$, these are the desired sets.

Lastly, to allow for different parameters $\varepsilon_1, \varepsilon_2$, just take $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$ and apply what we have already seen. Then any (ε, m_1)-uniform measure is (ε_1, m_1)-uniform and any (ε, m_2)-atomic measure is also (ε_2, m_2)-atomic, and we are done. □

Theorems 2.8 and 2.9 are formal consequences of Theorem 2.7, as discussed in Section 2.3.

5. **Self-similar measures**

5.1. *Uniform entropy dimension and self-similar measures.* The entropy dimension of a measure $\theta \in \mathcal{P}(\mathbb{R})$ is the limit $\lim_{n\to\infty} H_n(\theta)$, assuming it exists; by Lemma 3.4, this limit is equal to $\lim_{n\to\infty} \mathbb{E}_{0\leq i\leq n}(H_m(\theta^{x,i}))$ for all integers m. The convergence of the averages does not, however, imply that the entropies of the components $\theta^{x,i}$ concentrate around their mean, and examples show that they need not. We introduce the following stronger notion.

Definition 5.1. A measure $\theta \in \mathcal{P}(\mathbb{R})$ has *uniform entropy dimension* α if for every $\varepsilon > 0$, for large enough m,

$$\liminf_{n\to\infty} \mathbb{P}_{0\le i\le n}(|H_m(\theta^{x,i}) - \alpha| < \varepsilon) > 1 - \varepsilon. \tag{32}$$

Our main objective in this section is to prove

PROPOSITION 5.2. *Let $\mu \in \mathcal{P}(\mathbb{R})$ be a self-similar measure and $\alpha = \dim \mu$. Then μ has uniform entropy dimension α.*

For simplicity we first consider the case that all the contractions in the IFS contract by the same ratio r. Thus, consider an IFS $\Phi = \{\varphi_i\}_{i\in\Lambda}$ with $\varphi_i(x) = r(x - a_i)$, $0 < r < 1$. We denote the attractor by X and without loss of generality assume that $0 \in X \subseteq [0, 1]$, which can always be arranged by a change of coordinates and may be seen not to affect the conclusions. Let $\mu = \sum_{i\in\Lambda} p_i \cdot \varphi_i\mu$ be a self-similar measure, and as usual write $\varphi_i = \varphi_{i_1} \circ \cdots \circ \varphi_{i_n}$ and $p_i = p_{i_1} \cdot \ldots \cdot p_{i_n}$ for $i \in \Lambda^n$.

Let

$$\alpha = \dim \mu.$$

As we have already noted, self-similar measures are exact dimensional [9], and for such measures, the dimension and entropy dimension coincide:

$$\lim_{n\to\infty} H_n(\mu) = \alpha. \tag{33}$$

Fix $\widetilde{x} \in X$, and define probability measures

$$\mu_{x,k}^{[n]} = c \cdot \sum \{p_i \cdot \varphi_i\mu \,:\, i \in \Lambda^n \,,\, \varphi_i\widetilde{x} \in \mathcal{D}_k(x)\} \,,$$

where $c = c(x, \widetilde{x}, k, n)$ is a normalizing constant. Thus $\mu_{x,k}^{[n]}$ differs from $\mu_{x,k}$ in that, instead of restricting $\mu = \sum_{i\in\Lambda^n} p_i \cdot \varphi_i\mu$ to $\mathcal{D}_k(x)$, we include or exclude each term in its entirety depending on whether $\varphi_i\widetilde{x} \in \mathcal{D}_k(x)$. Since $\varphi_i\mu$ may not be supported entirely on either $\mathcal{D}_k(x)$ or its complement, in general we have neither $\mu_{x,k}^{[n]} \ll \mu_{x,k}$ nor $\mu_{x,k} \ll \mu_{x,k}^{[n]}$. Note that the definition of $\mu_{x,k}^{[n]}$ depends on the point $\widetilde{x}$, but this will not concern us.

For $0 < \rho < 1$, it will be convenient to write

$$\ell(\rho) = \lceil \log\rho / \log r \rceil \,,$$

so $\rho, r^{\ell(\rho)}$ differ by a multiplicative constant. Recall that $\|\cdot\|$ denotes the total variation norm; see Section 3.1.

LEMMA 5.3. *For every $\varepsilon > 0$ there is a $0 < \rho < 1$ such that, for all k and $n = \ell(\rho 2^{-k})$,*

$$\mathbb{P}_{i=k}\left(\left\|\mu_{x,i} - \mu_{x,i}^{[n]}\right\| < \varepsilon\right) > 1 - \varepsilon. \tag{34}$$

Furthermore, ρ can be chosen independently of $\widetilde{x}$ and of the coordinate system on $\mathbb{R}$. (So the same bound holds for any translate of μ.)

Proof. It is elementary that if μ is atomic then it consists of a single atom. In this case the statement is trivial, so assume μ is nonatomic. Then[10] given $\varepsilon > 0$, there is a $\delta > 0$ such that every interval of length δ has μ-mass $< \varepsilon^2/2$. Choose an integer q so that $r^q < \delta/2$, and let $\rho = r^q$.

Let $k \in \mathbb{N}$ and $\ell = \ell(2^{-k})$ so that $2^{-k} \cdot r \leq r^\ell \leq 2^{-k}$. Let $i \in \Lambda^\ell$, and consider those $j \in \Lambda^q$ such that $\varphi_{ij}\mu$ is not supported on an element of $\mathcal{D}_k$. Then $\varphi_{ij}\mu$ is supported on the interval J of length δ centered at one of the endpoints of an element of $\mathcal{D}_k$. Since $\varphi_i\mu$ can give positive mass to at most two such intervals J, and $\varphi_i\mu(J) < \varepsilon^2/2$ for each such J, we conclude that in the representation $\mu_i = \frac{1}{p_i}\sum_{j\in\Lambda^q} p_{ij}\cdot(\varphi_{ij}\mu)$, at least $1-\varepsilon^2$ of the mass comes from terms that are supported entirely on just one element of $\mathcal{D}_k$. Therefore the same is true in the representation $\mu = \sum_{u\in\Lambda^{\ell+q}} p_u \cdot \varphi_u\mu$. The inequality (34) now follows by an application of the Markov inequality. Finally, Since our choice of parameters did not depend on $\widetilde{x}$ and is invariant under translation of μ and of the IFS, the last statement holds. □

LEMMA 5.4. *For $\varepsilon > 0$, for large enough m and all k,*

$$\mathbb{P}_{i=k}\left(H_m(\mu^{x,i}) > \alpha - \varepsilon\right) > 1 - \varepsilon,$$

and the same holds for any translate of μ.

Proof. Let $\varepsilon > 0$ be given. Choose $0 < \varepsilon' < \varepsilon$ sufficiently small that $\|\nu - \nu'\| < \varepsilon'$ implies $|H_m(\nu) - H_m(\nu')| < \varepsilon/2$ for every m and every $\nu, \nu' \in \mathcal{P}([0,1]^d)$ (Lemma 3.3). Let ρ be as in the previous lemma chosen with respect to ε'. Assume that m is large enough that $|H_m(\mu') - \alpha| < \varepsilon/2$ whenever μ' is μ scaled by a factor of at most ρ (m exists by (33) and Lemma 3.2 (5)). Now fix k and let $\ell = \ell(\rho 2^{-k})$. By the previous lemma and choice of ε', it is enough to show that $\frac{1}{m}H(\mu_{x,k}^{[\ell]}, \mathcal{D}_{k+m}) > \alpha - \varepsilon/2$. But this follows from the fact that $\mu_{x,k}^{[\ell]}$ is a convex combination of measures μ_j for $j \in \Lambda^\ell$, our choice of m and ℓ, and concavity of entropy. □

We now prove Proposition 5.2. Let $0 < \varepsilon < 1$ be given, and fix an auxiliary parameter $\varepsilon' < \varepsilon/2$. We first show that this holds for m large in a manner depending on ε. Specifically, let m be large enough that the previous lemma applies for the parameter ε'. In particular, for any n,

$$\mathbb{P}_{0\leq i\leq n}\left(H_m(\mu^{x,i}) > \alpha - \varepsilon'\right) > 1 - \varepsilon'. \tag{35}$$

[10]This is the only part of the proof of Theorem 1.3 that is not effective, but with a little more work one could make it effective in the sense that, if $\liminf -\log\Delta^{(n)} = M < \infty$, then at arbitrarily small scales one can obtain estimates of the continuity of μ in terms of M.

By (33), for n large enough we have $|H_n(\mu) - \alpha| < \varepsilon'/2$, so by Lemma 3.4, for large enough n we have

$$|\mathbb{E}_{0\le i\le n}\left(H_m(\mu^{x,i})\right) - \alpha| < \varepsilon'.$$

Since $H_m(\mu^{x,i}) \ge 0$, the last two equalities imply

$$\mathbb{P}_{0\le i\le n}\left(H_m(\mu^{x,i}) < \alpha + \varepsilon''\right) > 1 - \varepsilon''$$

for some ε'' that tend to 0 with ε'. Thus, choosing ε' small enough, the last inequality and (35) give (32), as desired.

When the contraction ratios are not uniform, $\varphi_i = r_i x + a_i$, some minor changes are needed in the proof. Given n, let $\Lambda^{(n)}$ denote the set of $i \in \Lambda^* = \bigcup_{m=1}^{\infty} \Lambda^m$ such that $r_i < r^n \le r_j$, where j is the same as i but with the last symbol deleted. (So its length is one less than i.) This ensures that $\{r_i\}_{i\in\Lambda^{(n)}}$ are all within a multiplicative constant of each other. (This constant is $\min\{r_j \,:\, j \in \Lambda\}$.) It is easy to check that $\Lambda^{(n)}$ is a section of Λ^* in the sense that every sequence $i \in \Lambda^*$ with $r_i < r^n$ has a unique prefix in $\Lambda^{(n)}$. Now define $\mu_{x,k}^{[n]}$ as before, but using $\varphi_i\mu$ for $i \in \Lambda^{(n)}$; i.e.,

$$\mu_{x,k}^{[n]} = c \cdot \sum \left\{p_i \cdot \varphi_i\mu \,:\, i \in \Lambda^{(n)}\,,\, \varphi_i\widetilde{x} \in \mathcal{D}_k(x)\right\}.$$

With this modification all the previous arguments now go through.

Finally, let us note the following consequence of the inverse theorem (Theorem 2.8).

Corollary 5.5. *For every measure $\mu \in \mathcal{P}(\mathbb{R})$ with uniform entropy dimension $0 < \alpha < 1$, and for every $\varepsilon > 0$, there is a $\delta > 0$ and such that for all large enough n and every $\nu \in \mathcal{P}([0,1])$,*

$$H_n(\nu) > \varepsilon \implies H_n(\mu * \nu) \ge H_n(\mu) + \delta.$$

Similar conclusions hold for dimension.

5.2. *Proof of Theorem* 1.3. We again begin with the uniformly contracting case, $\varphi_i = rx + a_i$, and continue with the notation from the previous section; in particular, assume that 0 is in the attractor. Recall from the introduction that

$$\nu^{(n)} = \sum_{i\in\Lambda^n} p_i \cdot \delta_{\varphi_i(0)}.$$

Define

$$\tau^{(n)}(A) = \mu(r^{-n}A).$$

One may verify easily, using the assumption $0 \in X$, that

$$\mu = \nu^{(n)} * \tau^{(n)}. \tag{36}$$

As in the introduction, write

$$n' = [n\log(1/r)].$$

Thus $\tau^{(n)}$ is μ scaled down by a factor of $r^n = 2^{-n'}$ and translated. Using (33), Lemma 3.2, and the fact that $\tau^{(n)}$ is supported on an interval of order $r^n = 2^{-n'}$, we have

$$\lim_{n\to\infty} \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{n'}) = \lim_{n\to\infty} \frac{1}{n'} H(\mu, \mathcal{D}_{n'}) = \dim \mu = \alpha.$$

Suppose now that $\alpha < 1$. Fix a large q, and consider the identity

$$\begin{aligned}\frac{1}{qn} H(\mu, \mathcal{D}_{qn}) &= \frac{n'}{qn} \cdot \left(\frac{1}{n'} H(\mu, \mathcal{D}_{n'})\right) + \frac{qn-n'}{qn} \cdot \left(\frac{1}{qn-n'} H(\mu, \mathcal{D}_{qn}|\mathcal{D}_{n'})\right) \\ &= \frac{[\log(1/r)]}{q} \left(\frac{1}{n'} H(\mu, \mathcal{D}_{n'})\right) + \frac{q-[\log(1/r)]}{q} \left(\frac{1}{qn-n'} H(\mu, \mathcal{D}_{qn}|\mathcal{D}_{n'})\right).\end{aligned}$$

The left-hand side and the term $\frac{1}{n'} H(\mu, \mathcal{D}_{n'})$ on the right-hand side both tend to α as $n \to \infty$. Since r, q are independent of n, we conclude that

$$\lim_{n\to\infty} \frac{1}{qn-n'} H(\mu, \mathcal{D}_{qn}|\mathcal{D}_{n'}) = \alpha. \tag{37}$$

From the identity $= \mathbb{E}_{i=n'}(\nu_{y,i}^{(n)})$ and linearity of convolution,

$$\mu = \nu^{(n)} * \tau^{(n)} = \mathbb{E}_{i=n'} \left(\nu_{y,i}^{(n)} * \tau^{(n)}\right).$$

Also, each measure $\nu_{y,i}^{(n)} * \tau^{(n)}$ is supported on an interval of length $O(2^{-n'})$ so

$$|H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{n'}) - H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn})| = O(1).$$

By concavity of conditional entropy (Lemma 3.1 (5)),

$$\begin{aligned}H(\mu, \mathcal{D}_{qn}|\mathcal{D}_{n'}) &= H(\nu^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{n'}) \\ &\geq \mathbb{E}_{i=n'} \left(H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{n'})\right) \\ &= \mathbb{E}_{i=n'} \left(H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn})\right) + O(1),\end{aligned}$$

so by (37),

$$\limsup_{n\to\infty} \frac{1}{qn-n'} \mathbb{E}_{i=n'} \left(H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn})\right) \leq \alpha. \tag{38}$$

Now, we also know that

$$\lim_{n\to\infty} \frac{1}{qn-n'} H(\tau^{(n)}, \mathcal{D}_{qn}) = \alpha \tag{39}$$

since, up to a re-scaling, this is just (33). (We again used the fact that $\tau^{(n)}$ is supported on intervals of length $2^{-n'}$.) By Lemma 4.10, for every component $\nu_{y,i}^{(n)}$,

$$\frac{1}{qn-n'} H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) \geq \frac{1}{qn-n'} H(\tau^{(n)}, \mathcal{D}_{qn}) + O\left(\frac{1}{qn-n'}\right).$$

Therefore for every $\delta > 0$,

$$\lim_{n\to\infty} \mathbb{P}_{i=n'}\left(\frac{1}{qn-n'}H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) > \alpha - \delta\right) = 1$$

which, combined with (38), implies that for every $\delta > 0$,

$$\lim_{n\to\infty} \mathbb{P}_{i=n'}\left(\left|\frac{1}{qn-n'}H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) - \alpha\right| < \delta\right) = 1,$$

and replacing α with the limit in (39), we have that for all $\delta > 0$,

(40)
$$\lim_{n\to\infty} \mathbb{P}_{i=n'}\left(\left|\frac{1}{qn-n'}H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) - \frac{1}{qn-n'}H(\tau^{(n)}, \mathcal{D}_{qn})\right| < \delta\right) = 1.$$

Now let $\varepsilon > 0$. By Proposition 5.2 and the assumption that $\alpha < 1$, for small enough ε, large enough m and all sufficiently large n,

$$\begin{aligned}\mathbb{P}_{n'<i\leq qn'}\left(H_m((\tau^{(n)})^{x,i}) < 1-\varepsilon\right) &\geq \mathbb{P}_{n'<i\leq qn'}\left(H_m((\tau^{(n)})^{x,i}) < \alpha+\varepsilon\right). \\ &> 1-\varepsilon\end{aligned}$$

Choose $\delta > 0$ smaller than the constant of the same name in the conclusion of Theorem 2.8. Then, for sufficiently large n, we can apply Theorem 2.8 to the components $\nu_{y,i}^{(n)}$ in the event in equation (40). (For this we rescale by $2^{n'}$ and note that the measures $\nu_{y,n'}^{(n)}$ are supported on level-n' dyadic cells and $\tau^{(n)}$ is supported on an interval of the same order of magnitude.) We conclude that every component $\nu_{y,i}^{(n)}$ in the event in question satisfies $\frac{1}{qn-n'}H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn}) < \varepsilon$, and hence by (40),

$$\lim_{n\to\infty} \mathbb{P}_{i=n'}\left(\frac{1}{qn-n'}H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn}) < \varepsilon\right) = 1.$$

Thus, from the definition of conditional entropy and the last equation,

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{qn-n'}H(\nu^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{n'}) &= \lim_{n\to\infty}\frac{1}{qn-n'}\mathbb{E}_{i=n'}\left(H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn})\right) \\ &= \lim_{n\to\infty}\mathbb{E}_{i=n'}\left(\frac{1}{qn-n'}H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn})\right) \\ &< \varepsilon.\end{aligned}$$

Since ε was arbitrary, this is Theorem 1.3.

5.3. *Proof of Theorem* 1.4 (*the nonuniformly contracting case*). We now consider the situation for general IFS, in which the contraction r_i of φ_i is not constant. Again assume that 0 is in the attractor. Let $r = \prod_{i\in\Lambda} r_i^{p_i}$, $n' = \log_2(1/r)$ as in the introduction, and define $\widetilde{\nu}^{(n)}$ as before. Given n, let

$$R_n = \{r_i \,:\, i \in \Lambda^n\}.$$

Note that $|R_n| = O(n^{|\Lambda|})$. Therefore $H(\widetilde{\nu}^{(n)}, \{\mathbb{R}\} \times \mathcal{F}) = O(\log n)$, and consequently for all k,

$$H(\widetilde{\nu}^{(n)}, \widetilde{\mathcal{D}}_k) = H(\nu^{(n)}, \mathcal{D}_k) + O(\log n).$$

Thus

$$H(\widetilde{\nu}^{(n)}, \widetilde{\mathcal{D}}_{qn}|\widetilde{\mathcal{D}}_{n'}) = H(\nu^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_n) + O(\log n),$$

and our goal reduces to proving that for every $q > 1$,

$$\frac{1}{qn} H(\nu^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{n'}) \to 0 \qquad \text{as } n \to \infty.$$

Furthermore, for every $\varepsilon > 0$,

$$H(\nu^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{(1-\varepsilon)n'}) = H(\nu^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{n'}) - O(\varepsilon n),$$

so it will suffice for us to prove that

$$\limsup_{n\to\infty} \frac{1}{qn} H(\nu^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{(1-\varepsilon)n}) = o(1) \qquad \text{as } \varepsilon \to 0.$$

Fix $\varepsilon > 0$. For $t \in R_n$, let

$$\Lambda^{n,t} = \{i \in \Lambda^n \,:\, r_i = t\},$$
$$p^{n,t} = \sum_{i \in \Lambda^{n,t}} p_i,$$

so $\{p^{n,t}\}_{t\in R_n}$ is a probability vector. It will sometimes be convenient to consider $i \in \Lambda^n$, $i \in \Lambda^{n,t}$ and $t \in R_n$ as random elements drawn according to the probabilities p_i, $p_i/p^{n,t}$, and $p^{n,t}$, respectively. Then we interpret expressions such as $\mathbb{P}_{i\in\Lambda^n}(A)$, $\mathbb{P}_{i\in\Lambda^{n,t}}(A)$ and $\mathbb{P}_{t\in R_n}(A)$ in the obvious manner, and similarly expectations. With this notation, we can define

$$\nu^{(n,t)} = \mathbb{E}_{i\in\Lambda^{n,t}}(\delta_{\varphi_i(0)}) = \frac{1}{p^{n,t}} \sum_{i\in\Lambda^{n,t}} p_i \cdot \delta_{\varphi_i(0)}.$$

This a probability measure on $\mathbb{R}$ representing the part of $\nu^{(n)}$ coming from contractions by t; indeed,

$$\nu^{(n)} = \mathbb{E}_{t\in R_n}(\nu^{(n,t)}). \tag{41}$$

For $t > 0$, let $\tau^{(t)}$ be the measure

$$\tau^{(t)}(A) = \tau(tA).$$

(Note that we are no longer using logarithmic scale, so the measure that was previously denoted$\tau^{(n)}$ is now $\tau^{(2^{-n})}$.) We then have

$$\mu = \mathbb{E}_{t\in R_n}(\nu^{(n,t)} * \tau^{(t)}). \tag{42}$$

Fix $\varepsilon > 0$. Arguing as in the previous section, using equation (42) and concavity of entropy, we have

$$\begin{aligned}(43)\qquad \alpha &= \lim_{n\to\infty} \frac{1}{qn-(1-\varepsilon)n'} H(\mu, \mathcal{D}_{qn}|\mathcal{D}_{(1-\varepsilon)n'}) \\ &\geq \limsup_{n\to\infty} \frac{1}{qn-(1-\varepsilon)n'} \mathbb{E}_{t\in R_n}\left(H(\nu^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn}|\mathcal{D}_{(1-\varepsilon)n'})\right).\end{aligned}$$

By the law of large numbers,

$$\lim_{n\to\infty} \mathbb{P}_{i\in\Lambda^n}\left(2^{-(1+\varepsilon)n'} < r_i < 2^{-(1-\varepsilon)n'}\right) = 1$$

or, equivalently,

$$(44)\qquad \lim_{n\to\infty} \mathbb{P}_{t\in R_n}\left(2^{-(1+\varepsilon)n'} < t < 2^{-(1-\varepsilon)n'}\right) = 1.$$

Using $H_k(\mu) \to \alpha$ and the definition of $\tau^{(t)}$, we conclude that

$$\lim_{n\to\infty} \mathbb{P}_{t\in R_n}\left(\frac{1}{qn-(1-\varepsilon)n'} H(\tau^{(t)}, \mathcal{D}_{qn}) \geq (1-\varepsilon)\alpha\right) = 1.$$

Also, since $\tau^{(t)}$ is supported on an interval of order t, from (44), (43) and concavity of entropy,

$$\begin{aligned}(45)\quad \alpha &\geq \limsup_{n\to\infty} \frac{1}{qn-(1-\varepsilon)n'} \mathbb{E}_{t\in R_n}\mathbb{E}_{i=n'}\left(H(\nu^{(n,t)}_{y,i} * \tau^{(t)}, \mathcal{D}_{qn}|\mathcal{D}_{(1-\varepsilon)n'})\right) \\ &= \limsup_{n\to\infty} \frac{1}{qn-(1-\varepsilon)n'} \mathbb{E}_{t\in R_n}\mathbb{E}_{i=n'}\left(H(\nu^{(n,t)}_{y,i} * \tau^{(t)}, \mathcal{D}_{qn}\right).\end{aligned}$$

This is the analogue of equation 38 in the proof of the uniformly contracting case, and from here one proceeds exactly as in that proof to conclude that there is a function $\delta(\varepsilon)$, tending to 0 as $\varepsilon \to 0$, such that

$$\mathbb{P}_{t\in R_n}\left(\mathbb{P}_{i=n'}\left(\frac{1}{qn-(1-\varepsilon)n} H(\nu^{(n,t)}, \mathcal{D}_{qn}) < \delta(\varepsilon)\right)\right) = 1.$$

Now, using equation (41) and the fact that the entropy of the distribution $\{p^{(n,t)}\}_{t\in R_n}$ is $o(n)$ as $n \to \infty$, by Lemma 3.1(6) one concludes that

$$\limsup_{n\to\infty} H(\nu^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{(1-\varepsilon)n'}) \leq \delta(\varepsilon),$$

which is what we wanted to prove.

5.4. *Transversality and the dimension of exceptions.* In this section we prove Theorem 1.8. Let $I \subseteq \mathbb{R}$ be a compact interval for $t \in I$, and let $\Phi_t = \{\varphi_{i,t}\}_{i\in\Lambda}$ be an IFS, $\varphi_{i,t}(x) = r_i(t)(x - a_i(t))$. We define $\varphi_{i,t}$ and $r_i(t)$ for $i \in \Lambda^n$ as usual, set $\Delta_{i,j}(t) = \varphi_{i,t}(0) - \varphi_{j,t}(0)$ when $i, j \in \Lambda^n$, and for $i, j \in \Lambda^{\mathbb{N}}$, define $\Delta_{i,j}(t) = \lim \Delta_{i_1\cdots i_n, j_1\cdots j_n}(t)$. (This is well defined since $\lim \varphi_{i_1\cdots i_n}(0)$ converges, in fact exponentially, as $n \to \infty$.)

For $i,j \in \Lambda^n$ or $i,j \in \Lambda^{\mathbb{N}}$, let $i \wedge j$ denote the longest common initial segment of i,j, and $|i \wedge j|$ its length, so $|i \wedge j| = \min\{k : i_k \neq j_k\} - 1$. Let

$$r_{\min} = \min_{i \in \Lambda} \min_{t \in I} |r_i(t)|,$$

so $0 < r_{\min} < 1$. For a C^k-function $F : I \to \mathbb{R}$, write $F^{(p)} = \frac{d^p}{dt^p} F$ and

$$\|F\|_{I,k} = \max_{p \in \{0,\ldots,k\}} \max_{t \in I} |F^{(p)}(t)|.$$

In particular, we write

$$R_k = \max_{i \in \Lambda} \|r_i\|_{I,k}.$$

Definition 5.6. The family $\{\Phi_t\}_{t \in I}$ is *transverse of order k* if $r_i(\cdot), a_i(\cdot)$ are k-times continuously differentiable and there is a constant $c > 0$ such that for every $n \in \mathbb{N}$ and distinct $i,j \in \Lambda^n$,

(46)

$$\forall\, t_0 \in I \quad \exists\, p \in \{0,1,2,\ldots,k\} \quad \text{such that} \quad |\Delta_{i,j}^{(p)}(t_0)| \geq c \cdot |i \wedge j|^{-p} \cdot r_{i \wedge j}(t_0).$$

The classical notion of transversality roughly corresponds to the case $k = 1$ in this definition; see, e.g., [24, Def. 2.7]. Unlike the classical notion, which either fails or is difficult to verify in many cases of interest, higher-order transversality holds almost automatically. To begin with, let $i,j \in \Lambda^n$ and observe that

$$\Delta_{i,j}(t) = r_{i \wedge j}(t) \widetilde{\Delta}_{i,j}(t)$$

where, writing u,v for the sequences obtained from i,j after deleting the longest initial segment,

$$\widetilde{\Delta}_{i,j}(t) = \Delta_{u,v}(t).$$

Differentiating p times,

$$\begin{aligned} \widetilde{\Delta}_{i,j}^{(p)}(t) &= \frac{d^p}{dt^p}(r_{i \wedge j}(t)^{-1} \cdot \Delta_{i,j}(t)) \\ &= \sum_{q=0}^{p} \binom{p}{q} \cdot \frac{d^q}{dt^q}(r_{i \wedge j}(t)^{-1}) \cdot \Delta_{i,j}^{(p-q)}(t). \end{aligned}$$

A calculation shows that

$$\left| \frac{d^q}{dt^q}(r_{i \wedge j}(t)^{-1}) \right| \leq O_{q, r_{\min}, R_q}(|i \wedge j|^q \cdot r_{i \wedge j}(t)^{-1}).$$

Thus we have the bound

$$|\widetilde{\Delta}_{i,j}^{(p)}(t)| = O_{p, r_{\min}, R_p}\left(\max_{0 \leq q \leq p} \left(|i \wedge j|^q \cdot r_{i \wedge j}(t)^{-1} \cdot |\Delta_{i,j}^{(q)}(t)| \right) \right).$$

Proposition 5.7. *Suppose $r_i(\cdot), a_i(\cdot)$ are real-analytic on I. Suppose that for $i,j \in \Lambda^{\mathbb{N}}$, $\Delta_{i,j} \equiv 0$ on I if and only if $i = j$. Then the associated family $\{\Phi_t\}_{t \in I}$ is transverse of order k for some k.*

Proof. First, for $x \in I$, we can extend r_i, a_i analytically to a complex neighborhood U_x of x on which $|r_i|$ are still bounded uniformly away from 1. Define $\Delta_{i,j}(z)$ as before for $i, j \in \Lambda^n$ and $z \in U_x$, and note that for $i, j \in \Lambda^{\mathbb{N}}$, the limit $\Delta_{i,j}(z) = \lim \Delta_{i_1 \cdots i_n, j_1 \cdots j_n}(z)$ is uniform for $z \in U_x$. This shows that $\Delta_{i,j}(t)$ is also real-analytic on I

Given k, from the expression for $\widetilde{\Delta}^{(p)}_{i,j}$ above, we see that if $c > 0$ and there exists $t_0 \in I$ such that $|\Delta^{(p)}_{i,j}(t_0)| \le c \cdot |i \wedge j|^{-p} \cdot r_{i\wedge j}(t_0)$ for all $0 \le p \le k$, then $|\widetilde{\Delta}^{(p)}_{i,j}(t_0)| \le c'$ for all $0 \le p \le k$, where $c' = O_{k,R_k}(c)$. For each k, choose $c_k > 0$ such that the associated c'_k satisfies $c'_k < 1/k$.

Suppose that for all k the family $\{\Phi_t\}$ is not transverse of order k. Then by assumption we can choose $n(k)$ and distinct $i^{(k)}, j^{(k)} \in \Lambda^{n(k)}$, and a point $t_k \in I$, such that $|\Delta^{(p)}_{i^{(k)},j^{(k)}}(t_k)| \le c_k \cdot |i^{(k)} \wedge j^{(k)}|^{-p} \cdot r_{i^{(k)}\wedge j^{(k)}}(t_k)$ for $0 \le p \le k$, and hence $\widetilde{\Delta}^{(p)}_{i^{(k)},j^{(k)}}(t_k) \le c'_k$. Let $u^{(k)}$ and $v^{(k)}$ denote the sequences obtained from $i^{(k)}$ and $j^{(k)}$ by deleting the first $|i^{(k)} \wedge j^{(k)}|$ symbols, so that the first symbols of $u^{(k)}$ and $v^{(k)}$ now differ and $\Delta_{u^{(k)},v^{(k)}} = \widetilde{\Delta}_{i^{(k)},j^{(k)}}$. Hence we have

$$|\Delta^{(p)}_{u^{(k)},v^{(k)}}(t_k)| \le c'_k < 1/k \quad \text{for all} \quad 0 \le p \le k. \tag{47}$$

Passing to a subsequence k_ℓ, we may assume that $t_{k_\ell} \to t_0$ and that $u^{(k_\ell)} \to u \in \Lambda^{\mathbb{N}}$ and $v^{(k_\ell)} \to v \in \Lambda^{\mathbb{N}}$ (the latter in the sense that all coordinates stabilize eventually to the corresponding coordinate in the limit sequence). Note that $u \ne v$, because $u^{(k_\ell)}, v^{(k_\ell)}$ differ in their first symbol for all ℓ, hence so do u, v. It follows that $\Delta_{u^{(k_\ell)},v^{(k_\ell)}} \to \Delta_{u,v}$ uniformly and that the same holds for p-th derivatives. Hence for all $p \ge 0$, using uniform convergence and (47),

$$|\Delta^{(p)}_{u,v}(t_0)| = \lim_{\ell\to\infty} |\Delta^{(p)}_{u^{(k_\ell)},v^{(k_\ell)}}(t_{k_\ell})| = 0.$$

But $\Delta_{u,v}$ is real analytic, so the vanishing of its derivatives implies $\Delta_{u,v} \equiv 0$ on I, contrary to the hypothesis. □

We turn now to the implications of transversality. The key implication is provided by the following simple lemma.

Lemma 5.8. *Let $k \in \mathbb{N}$, and let F be a k-times continuously differentiable function on a compact interval $J \subseteq \mathbb{R}$. Let $M = \|F\|_{J,k}$, and let $0 < b < 1$ be such that for every $x \in J$ there is a $p \in \{0, \dots, k\}$ with $|F^{(p)}(x)| > b$. Then for every $0 < \rho < (b/2)^{2^k}$, the set $F^{-1}(-\rho, \rho) \subseteq J$ can be covered by $O_{k,M,|J|}(1/b^k)$ intervals of length $\le 2(\rho/b)^{1/2^k}$ each.*

Proof. For brevity, we suppress dependence on the parameters $k, M, |J|$, so throughout this proof, $O(\cdot) = O_{k,M,|J|}(\cdot)$.

The proof is by induction on k. For $k = 0$, the hypothesis is that $|F^{(0)}(x)| = |F(x)| > b$ for all $x \in J$, hence $F^{-1}(-\rho,\rho) = \emptyset$ for $0 < \rho < b/2 = (b/2)^{2^0}$, and the assertion is trivial.

Assume that we have proved the claim for $k-1$, and consider the case k. Let J' be a maximal closed interval in $F^{-1}[-b,b]$, and let $G = F'|_{J'}$. Note that G satisfies the hypothesis for $k-1$ and the same value of b and M, and $\sqrt{b\rho} < \sqrt{\rho} < (b/2)^{2^{k-1}}$, so from the induction hypothesis we find that $G^{-1}(-\sqrt{b\rho},\sqrt{b\rho})$ can be covered by $O(1/b^{k-1})$ intervals of length $< 2(\sqrt{b\rho}/b)^{1/2^{k-1}} = 2(\rho/b)^{1/2^k}$ each. Let U denote the union of this cover, and consider the intervals J'_i that are the closures of the maximal subintervals in $J' \setminus U$. By the above, the number of such intervals J'_i is $\leq O(1/b^{k-1})$. Now, on each J'_i we have $|F'| \geq \sqrt{b\rho}$, so by continuity of F' either $F' \geq \sqrt{b\rho}$ or $F' \leq -\sqrt{b\rho}$ in all of J'_i. An elementary consequence of this is that $J'_i \cap F^{-1}(-\rho,\rho)$ is an interval of length at most $2\rho/\sqrt{b\rho} = 2\sqrt{\rho/b} \leq 2(\rho/b)^{1/2^k}$. In summary we have covered $J' \cap F^{-1}(-\rho,\rho)$ by $O(1/b^{k-1})$ intervals of length $2(\rho/b)^{1/2^k}$ each.

It remains to show that there are $O(1/b)$ maximal intervals $J' \subseteq F^{-1}[-b,b]$ as in the paragraph above. In fact, we only need to bound the number of such J' that intersect $F^{-1}(-\rho,\rho)$. For J' of this kind, if $J' = J$, we are done, since this means there is just one such interval. Otherwise there is an endpoint $a \in J'$ with $|F(a)| = b$. There is also a point $a' \in J'$ with $|F(a')| < \rho < (b/2)^{2^k}$. Since $|F'| \leq M$, we conclude that $|J'| \geq |a'-a| \geq (b-\rho)/M \geq b/2M$. Thus, since the intervals J' are disjoint, their number is $\leq |J|/(b/2M) = O(1/b)$, completing the induction step. □

Let $\operatorname{bdim} X$ denote the upper box dimension of a set X, defined by

$$\operatorname{bdim} X = \limsup_{r\to 0} \frac{\log \#\min\{\ell \,:\, X \text{ can be covered by } \ell \text{ balls of radius } r\}}{\log(1/r)}.$$

One always has $\dim X \leq \operatorname{bdim} X$. The packing dimension is defined by

$$\operatorname{pdim} X = \inf\left\{\sup_n \operatorname{bdim} X_n \,:\, X \subseteq \bigcup_{n=1}^{\infty} X_n\right\}.$$

Note that $\dim X \leq \operatorname{pdim} X$, and $Y \subseteq X$ implies $\operatorname{pdim} Y \leq \operatorname{pdim} X$.

Theorem 5.9. *If $\{\Phi_t\}_{t\in I}$ satisfies transversality of order $k \geq 1$ on the compact interval I, then the set E of "exceptional" parameters in Theorem* 1.7 *has packing* (*and hence Hausdorff*) *dimension* 0.

Proof. Write

$$M = \sup_n \sup_{i,j\in\Lambda^n} \|\Delta_{i,j}\|_{I,k}\,.$$

That $M < \infty$ follows from k-fold continuous differentiability of $r_i(\cdot), a_i(\cdot)$ and the fact that $|r_i|$ are bounded away from 1 on I. By transversality there is a constant $c > 0$ such that for every $t \in I$, every n and all distinct $i, j \in \Lambda^n$,

$$\left|\frac{\partial^p}{\partial t^p}\Delta_{i,j}(t)\right| > c \cdot |i \wedge j|^{-p} \cdot r_{\min}^{|i\wedge j|} \qquad \text{for some } p \in \{0, \ldots, k\}.$$

We may assume that $c < 1$ and $k \geq 2$. In what follows we suppress the dependence on k, M, c and $|I|$ in the $O(\cdot)$ notation: $O(\cdot) = O_{k,M,c,|I|}(\cdot)$.

Fix n and distinct $i, j \in \Lambda^n$. Let $b = b(n) = cn^{-k} r_{\min}^n$ so that the hypothesis of the previous lemma is satisfied for the function $F = \Delta_{i,j}$ and this b. Therefore, for all $0 < \rho < (b/2)^{2^k}$, the set $\{t \in I : |\Delta_{i,j}| < \rho\}$ can be covered by at most $O(1/b^k)$ intervals of length $2(\rho/b)^{1/2^k}$ each.

Now let $\varepsilon > 0$ be such that $\rho = \varepsilon^n$ satisfies $\rho < (b(n)/2)^{2^k} = (cn^{-k}r_{\min}^n)^{2^k}$ for all n. (This holds for all sufficiently small $\varepsilon > 0$.) Fixing n again, the discussion above applies to $(\Delta_{i,j})^{-1}(-\varepsilon^n, \varepsilon^n)$ for every distinct pair $i, j \in \Lambda^n$, so ranging over all such pairs we find that

$$E_{\varepsilon,n} = \bigcup_{i,j\in\Lambda^n,\, i\neq j} (\Delta_{i,j})^{-1}(-\varepsilon^n, \varepsilon^n)$$

can be covered by $O(|\Lambda|^n/b(n)^k)$ intervals of length at most $O((\varepsilon^n/b(n))^{1/2^k})$. Now, $E \subseteq E_\varepsilon$ where

$$E_\varepsilon = \bigcup_{N=1}^{\infty} \bigcap_{n>N} E_{\varepsilon,n}. \tag{48}$$

By the above, for each ε and N, we have

$$\begin{aligned}
\operatorname{bdim}\left(\bigcap_{n>N} E_{\varepsilon,n}\right) &\leq \lim_{n\to\infty} \frac{\log\left(|\Lambda|^n/b(n)^k\right)}{\log\left((\varepsilon^n/b(n))^{1/2^k}\right)} \\
&= \lim_{n\to\infty} \frac{\log\left(n^{O(1)} \cdot |\Lambda|^{kn}/r_{\min}^{kn}\right)}{\log\left(n^{O(1)} \cdot \varepsilon^{n/2^k}/r_{\min}^{n/2^k}\right)} \\
&= O_k\left(\frac{\log(|\Lambda|/r_{\min})}{\log(\varepsilon/r_{\min})}\right).
\end{aligned}$$

The last expression tends to 0 as $\varepsilon \to 0$, uniformly in N. Thus by (48), the same is true of E_ε, and $E \subseteq E_\varepsilon$ for all ε, so E has packing (and Hausdorff) dimension 0. □

Theorem 1.8 now follows by combining Proposition 5.7 and Theorem 5.9.

5.5. *Miscellaneous proofs.* To complete the proof of Corollary 1.5, we have

LEMMA 5.10. *Let $A \subseteq \mathbb{R}$ be a finite set of algebraic numbers over $\mathbb{Q}$. Then there is a constant $0 < s < 1$ such that any polynomial expression x of degree n in the elements of A, either $x = 0$ or $|x| > s^n$.*

Proof. Choose an algebraic integer α such that $A \subseteq \mathbb{Q}(\alpha)$. Since the statement is unchanged if we multiply all elements of A by an integer, we can assume that the elements of A are integer polynomials in α of degree $\leq d$ with coefficients bounded by N for some d, N that depend only on α. Substituting these polynomials into the expression for x, we have an expression $x = \sum_{k=0}^{dn} n_k \alpha^k$, where $n_k \in \mathbb{N}$ and $|n_k| \leq N$. It suffices to prove that any such expression is either 0 or $\geq s^n$ for $0 < s < 1$ independent of n (but which may depend on α and hence on d, N). In proving this last statement we may assume that $d = 1$. (Replace s by $s^{1/d}$ and change variables to $n' = dn$.)

Let $\alpha = \alpha_1, \alpha_2, \ldots, \alpha_u$ denote the algebraic conjugates of α, and let $\sigma_1, \sigma_2, \ldots, \sigma_u$ denote the automorphisms of $\mathbb{Q}(\alpha)$, with $\sigma_i \alpha = \alpha_i$. If $x \neq 0$, then $\prod_{i=1}^{u} \sigma_i(x) \in \mathbb{Z}$, so

$$1 \leq \left| \prod_{i=1}^{u} \sigma_i(x) \right| = x \cdot \prod_{i=2}^{u} \left| \sum_{k=0}^{n} n_k \sigma_i(x)^k \right| \leq x \cdot \prod_{i=2}^{u} \sum_{k=0}^{n} n_k |\alpha_i|^k \leq x \cdot (n \cdot N \cdot \alpha_{\max}^n)^u,$$

where $\alpha_{\max} = \max\{|\alpha_2|, \ldots, |\alpha_u|\}$. Dividing out gives the lemma. $\square$

We finish with some comments on Sinai's problem, Theorem 1.11. We first state a generalization of Theorem 1.7 needed to treat families of IFSs that contract only on average.

Suppose that for $t \in I$ we have a family $\Phi_t = \{\varphi_{i,t}\}_{i \in \Lambda}$ of (not necessarily contracting) similarities of $\mathbb{R}$, and as usual write $\varphi_{i,t} = r_{i,t} U_{i,t} + a_{i,t}$. Let p be a fixed probability vector, and suppose that for each t we have $\sum p_i' \log r_i < 0$; i.e., the systems contract on average. One can then show that there is a unique probability measure μ_t on $\mathbb{R}$ satisfying $\mu_t = \sum_{i \in \Lambda} p_i \cdot \varphi_{i,t} \mu_t$ [23], that $H(\mu_t, \mathcal{D}_m) < \infty$ for every t and m, and that $\mu_t([-R, R]) \to 1$ as $R \to \infty$ uniformly in t. Under these conditions one can verify the stronger property that for every $t \in I$, we have

$$\left| H_m(\mu_t) - H_m((\mu_t)_{[-R,R]}) \right| = o(1) \qquad \text{as } R \to \infty$$

uniformly in t and m.

THEOREM 5.11. *Let $(\Phi_t)_{t \in I}$, p, and μ_t be as in the preceding paragraph. Let $\widetilde{\mu}$ denote the product measure on $\Lambda^{\mathbb{N}}$ with marginal p, and suppose that $A \subseteq \Lambda^{\mathbb{N}}$ is a Borel set such that $\widetilde{\mu}(A) > 0$. Write*

$$E = \bigcap_{\varepsilon > 0} \left(\bigcup_{N=1}^{\infty} \bigcap_{n > N} \left(\bigcup_{i,j \in A} (\Delta_{i,j})^{-1}((-\varepsilon^n, \varepsilon^n)) \right) \right).$$

Then $\dim \mu_t = \min\{d, \text{s-}\dim \mu_t\}$ *for every* $t \in I \setminus E$. *Furthermore, suppose that $I \subseteq \mathbb{R}$ is compact and connected and that the parametrization is analytic in the sense of Theorem* 1.8. *If*

$$\forall i, j \in A \quad (\ \Delta_{i,j} \equiv 0 \text{ on } I \iff i = j\ ),$$

then the set E above is of packing (*and Hausdorff*) *dimension at most* $k-1$ *and*, *in particular*, *of Lebesgue measure* 0.

The proof is the same as the proofs of Theorems 1.7 and 1.8, except that in analyzing the resulting convolution one must approximate μ_t by $(\mu_t)_{[-R,R]}$ for an appropriately large R that is fixed in advance, with the scale n large relative to R. We omit the details.

Let us see how this applies to Theorem 1.11, where $\varphi_{-1,\alpha}(x) = (1-\alpha)x-1$ and $\varphi_{1,\alpha}(x) = (1+\alpha)x+1$ for $\alpha \in (0,1]$, and $p = (1/2, 1/2)$. It suffices to consider the system for $\alpha \in [s,1]$ for some $s > 0$. Let A be the set of $i \in \Lambda^{\mathbb{N}}$ such that $|\frac{1}{N}\sum_{n=1}^{N} i_n - \frac{1}{2}| < \delta$ for $n > N(\delta)$, where $\delta > 0$ small enough to ensure that $|\varphi_{i_1\cdots i_n}| < 1$ when this condition holds, and $N(\delta)$ large enough that $\widetilde{\mu}(A) > 0$; in fact we can make $\widetilde{\mu}(A)$ arbitrarily close to 1, by the law of large numbers. It remains to verify for $i, j \in A$ that $\Delta_{i,j}$ vanishes on $[s,1]$ if and only if $i = j$. Note that for $i \in \{-1,1\}^n$,

$$\varphi_{i,\alpha}(0) = 1 + (1+i_1\alpha) + (1+i_1\alpha)(1+i_2\alpha) + \cdots + \prod_{k=1}^{n}(1+i_k\alpha).$$

Thus $\Delta_{i,j}$ is a series whose terms are of the form $c_{k,m}(1-\alpha)^k(1+\alpha)^m$ for some $c_{k,m} \in \{0, \pm 1\}$, and $i = j$ if and only if all terms are 0. Furthermore, there is an n_0 such that if $k + m \geq n_0$ and $c_{k,m} \neq 0$, then $k > (1-\delta)m$. Thus since $s \leq \alpha \leq 1$ and δ was chosen small enough, the series converges uniformly on $[s,1]$ and, furthermore, there is an $\varepsilon > 0$ such that the series converges uniformly on some larger interval $[s, 1+\varepsilon]$, and even in a neighborhood of 1 in the complex plane. Hence $\Delta_{i,j}(\cdot)$ is real-analytic on $[s, 1+\varepsilon]$ and is given by this series. Now, if $i \neq j$, we can divide out by the highest power $(1-\alpha)^{k_0}$ that is common to all the terms (possibly $k_0 = 0$) and evaluate the resulting function at $\alpha = 1$. We get a finite sum of the form $\sum_{(k,m)\in U} c_{m,k}2^m$ for some finite set of indices $U \in \mathbb{N}^2$ such that $c_{m,k} \in \{\pm 1\}$ for $(k,m) \in U$. Such a sum cannot vanish, hence by analyticity $\Delta_{i,j} \not\equiv 0$ on every sub-interval of $[s, 1+\varepsilon]$ and, in particular, $\Delta_{i,j} \not\equiv 0$ on $[s,1]$, as desired.

References

[1] J. C. Alexander and J. A. Yorke, Fat baker's transformations, *Ergodic Theory Dynam. Systems* **4** (1984), 1–23. MR 0758890. Zbl 0553.58020. http://dx.doi.org/10.1017/S0143385700002236.

[2] J. Bourgain, On the Erdős-Volkmann and Katz-Tao ring conjectures, *Geom. Funct. Anal.* **13** (2003), 334–365. MR 1982147. Zbl 1115.11049. http://dx.doi.org/10.1007/s000390300008.

[3] J. Bourgain, The discretized sum-product and projection theorems, *J. Anal. Math.* **112** (2010), 193–236. MR 2763000. Zbl 1234.11012. http://dx.doi.org/10.1007/s11854-010-0028-x.

[4] J. Bourgain, N. Katz, and T. Tao, A sum-product estimate in finite fields, and applications, *Geom. Funct. Anal.* **14** (2004), 27–57. MR 2053599. Zbl 1145.11306. http://dx.doi.org/10.1007/s00039-004-0451-1.

[5] P. Erdös, On a family of symmetric Bernoulli convolutions, *Amer. J. Math.* **61** (1939), 974–976. MR 0000311. Zbl 0022.35402. http://dx.doi.org/10.2307/2371641.

[6] P. Erdös, On the smoothness properties of a family of Bernoulli convolutions, *Amer. J. Math.* **62** (1940), 180–186. MR 0000858. Zbl 0022.35403. http://dx.doi.org/10.2307/2371446.

[7] P. Erdős and B. Volkmann, Additive Gruppen mit vorgegebener Hausdorffscher Dimension, *J. Reine Angew. Math.* **221** (1966), 203–208. MR 0186782. Zbl 0135.10202.

[8] C.-G. Esseen, On the Liapounoff limit of error in the theory of probability, *Ark. Mat. Astr. Fys.* **28A** (1942), 19. MR 0011909. Zbl 0027.33902.

[9] D.-J. Feng and H. Hu, Dimension theory of iterated function systems, *Comm. Pure Appl. Math.* **62** (2009), 1435–1500. MR 2560042. Zbl 1230.37031. http://dx.doi.org/10.1002/cpa.20276.

[10] H. Furstenberg, Intersections of Cantor sets and transversality of semigroups, in *Problems in Analysis* (Sympos. Salomon Bochner, Princeton Univ., Princeton, N.J., 1969), Princeton Univ. Press, Princeton, N.J., 1970, pp. 41–59. MR 0354562. Zbl 0208.32203.

[11] H. Furstenberg, Ergodic fractal measures and dimension conservation, *Ergodic Theory Dynam. Systems* **28** (2008), 405–422. MR 2408385. Zbl 1154.37322. http://dx.doi.org/10.1017/S0143385708000084.

[12] A. M. Garsia, Arithmetic properties of Bernoulli convolutions, *Trans. Amer. Math. Soc.* **102** (1962), 409–432. MR 0137961. Zbl 0103.36502. http://dx.doi.org/10.2307/1993615.

[13] A. M. Garsia, Entropy and singularity of infinite convolutions, *Pacific J. Math.* **13** (1963), 1159–1169. MR 0156945. Zbl 0126.14901. http://dx.doi.org/10.2140/pjm.1963.13.1159.

[14] M. Hochman, Self-similar sets with overlaps and sumset phenomena for entropy, the multidimensional case, 2012, preprint.

[15] M. Hochman and P. Shmerkin, Equidistribution from fractals, 2011. arXiv 1302.5792.

[16] V. A. Kaĭmanovich and A. M. Vershik, Random walks on discrete groups: boundary and entropy, *Ann. Probab.* **11** (1983), 457–490. MR 0704539. Zbl 0641.60009. http://dx.doi.org/10.1214/aop/1176993497.

[17] N. H. Katz and T. Tao, Some connections between Falconer's distance set conjecture and sets of Furstenburg type, *New York J. Math.* **7** (2001), 149–187. MR 1856956. Zbl 0991.28006. Available at http://nyjm.albany.edu:8000/j/2001/7_149.html.

[18] M. Keane, M. Smorodinsky, and B. Solomyak, On the morphology of γ-expansions with deleted digits, *Trans. Amer. Math. Soc.* **347** (1995), 955–966. MR 1290723. Zbl 0834.11033. http://dx.doi.org/10.2307/2154880.

[19] R. Kenyon, Projecting the one-dimensional Sierpinski gasket, *Israel J. Math.* **97** (1997), 221–238. MR 1441250. Zbl 0871.28006. http://dx.doi.org/10.1007/BF02774038.

[20] T. W. Körner, Hausdorff dimension of sums of sets with themselves, *Studia Math.* **188** (2008), 287–295. MR 2429824. Zbl 1200.28007. http://dx.doi.org/10.4064/sm188-3-4.

[21] M. Madiman, On the entropy of sums, in *Information Theory Workshop*, 2008. *ITW '*08, *IEEE*, 2008, pp. 303–307. http://dx.doi.org/10.1109/ITW.2008.4578674.

[22] M. Madiman, A. W. Marcus, and P. Tetali, Entropy and set cardinality inequalities for partition-determined functions, *Random Structures Algorithms* **40** (2012), 399–424. MR 2925304. Zbl 1244.05024. http://dx.doi.org/10.1002/rsa.20385.

[23] M. Nicol, N. Sidorov, and D. Broomhead, On the fine structure of stationary measures in systems which contract-on-average, *J. Theoret. Probab.* **15** (2002), 715–730. MR 1922444. Zbl 1026.28006. http://dx.doi.org/10.1023/A:1016224000145.

[24] Y. Peres and W. Schlag, Smoothness of projections, Bernoulli convolutions, and the dimension of exceptions, *Duke Math. J.* **102** (2000), 193–251. MR 1749437. Zbl 0961.42007. http://dx.doi.org/10.1215/S0012-7094-00-10222-0.

[25] Y. Peres, W. Schlag, and B. Solomyak, Sixty years of Bernoulli convolutions, in *Fractal Geometry and Stochastics*, *II* (Greifswald/Koserow, 1998), *Progr. Probab.* **46**, Birkhäuser, Basel, 2000, pp. 39–65. MR 1785620. Zbl 0961.42006.

[26] Y. Peres, K. Simon, and B. Solomyak, Absolute continuity for random iterated function systems with overlaps, *J. London Math. Soc.* **74** (2006), 739–756. MR 2286443. Zbl 1122.37018. http://dx.doi.org/10.1112/S0024610706023258.

[27] Y. Peres and B. Solomyak, Problems on self-similar sets and self-affine sets: an update, in *Fractal Geometry and Stochastics*, *II* (Greifswald/Koserow, 1998), *Progr. Probab.* **46**, Birkhäuser, Basel, 2000, pp. 95–106. MR 1785622. Zbl 0946.28003.

[28] M. Pollicott and K. Simon, The Hausdorff dimension of λ-expansions with deleted digits, *Trans. Amer. Math. Soc.* **347** (1995), 967–983. MR 1290729. Zbl 0831.28005. http://dx.doi.org/10.2307/2154881.

[29] F. Przytycki and M. Urbański, On the Hausdorff dimension of some fractal sets, *Studia Math.* **93** (1989), 155–186. MR 1002918. Zbl 0691.58029.

[30] J. Schmeling and P. Shmerkin, On the dimension of iterated sumsets, in *Recent Developments in Fractals and Related Fields*, *Appl. Numer. Harmon. Anal.*, Birkhäuser, Boston, MA, 2010, pp. 55–72. MR 2742987. Zbl 1216.28011. http://dx.doi.org/10.1007/978-0-8176-4888-6_5.

[31] P. Shmerkin, On the exceptional set for absolute continuity of Bernoulli convolutions, *Geom. Funct. Anal.* **24**, 946–958. MR 3213835. http://dx.doi.org/10.1007/s00039-014-0285-4.

[32] P. Shmerkin and B. Solomyak, Zeros of $\{-1, 0, 1\}$ power series and connectedness loci for self-affine sets, *Experiment. Math.* **15** (2006), 499–511. MR 2293600. Zbl 1122.30002. http://dx.doi.org/10.1080/10586458.2006.10128977.

[33] B. Solomyak, On the random series $\sum \pm\lambda^n$ (an Erdős problem), *Ann. of Math.* **142** (1995), 611–625. MR 1356783. Zbl 0837.28007. http://dx.doi.org/10.2307/2118556.

[34] G. Świątek and J. J. P. Veerman, On a conjecture of Furstenberg, *Israel J. Math.* **130** (2002), 145–155. MR 1919376. Zbl 1022.37016. http://dx.doi.org/10.1007/BF02764075.

[35] T. Tao, Sumset and inverse sumset theory for Shannon entropy, *Combin. Probab. Comput.* **19** (2010), 603–639. MR 2647496. Zbl 1239.11015. http://dx.doi.org/10.1017/S0963548309990642.

[36] T. Tao and V. Vu, *Additive Combinatorics*, *Cambridge Stud. Adv. Math.* **105**, Cambridge Univ. Press, Cambridge, 2006. MR 2289012. Zbl 1127.11002. http://dx.doi.org/10.1017/CBO9780511755149.

(Received: March 6, 2013)
(Revised: November 14, 2013)

Einstein Institute of Mathematics, The Hebrew University of Jerusalem, Givat Ram, Jerusalem, Israel
E-mail: mhochman@math.huji.ac.il