Increase of entropy under convolution and self-similar sets with overlaps

Master's Thesis
Aleksi Pyörälä
2459196
Research Unit of Mathematical Sciences
University of Oulu
Spring 2019

Contents

Introduction and main results

This thesis has two major sections: the increase of entropy under convolution and the dimension theory of sets and measures. While they both have interesting results of their own and may be considered two separate subjects, the connection between them is later found when we consider the dimensions of self-similar sets and measures without assuming separation of any kind in the cylinders of the associated iterated function system.

The main result in our discussion regarding the entropy of a probability measure (Theorem 1.26) is the following: given measures μ and ν on the unit interval such that the normalized entropy of their convolution is not too large, i.e.

$$\tfrac{1}{n}H(\mu * \nu, \mathcal{D}_n) \leq \tfrac{1}{n}H(\mu, \mathcal{D}_n) + \delta$$

for a small δ and all large n, either the typical restriction of μ on a dyadic interval of length 2^{-n} has to be close to uniform or the typical restriction of ν has to be close to atomic. This theorem is presented by M. Hochman in [7] and may be considered a generalization of the Freiman theorem in additive combinatorics to the fractal regime.

In the area of fractal geometry which occupies us for the second half of this thesis, our motivation is to study the dimensions of self-similar sets and measures, later specializing ourselves in the case where the open set condition is not fulfilled. We discuss a conjecture stating that, for a self-similar set on the line, the only case in which the (Hausdorff) dimension of the set can be strictly less than the minimum of its similarity dimension and the dimension of the ambient space, is the occurrence of an exact overlap in its cylinders (Conjecture 4.8). The main result for us in this area (Theorem 4.9), given again by Hochman in [7], contributes to this conjecture by stating that a strict inequality of this kind implies a super-exponential decrease in the distance of the cylinders of the associated IFS. A major ingredient in the proof of this result is the conclusion of Section 1, Theorem 1.26. An immediate application proves the conjecture to hold in an IFS with exponentially separated cylinders.

The purpose of this thesis is to offer a comprehensive introduction to the results mentioned above. For this, it is necessary to begin with the general theory of measure-theoretical entropy and the theory of fractal dimensions, especially for self-similar sets and measures. We begin by recalling the standard probabilistic and measure-theoretical results and definitions, found in most text-books (e.g. [12]) and used throughout the thesis, after which we turn to multiscale analysis of the entropy of a measure. This discussion is based mainly on Hochman's paper [7]. After arriving at the desired conclusion on this subject, Theorem 1.26, we shift our focus onto fractal geometry and build up the subject from the basics, presenting all the tools required in understanding the more deep results located in Section 4. Most of the results in Sections 2 and 3 can be found in K. Falconer's book [4] or in most text-books concerning fractal geometry and geometric measure theory.

On notations

We give a few words on the notations used in this thesis, particularly on those on whose definitions one might find slight variation through the literature.

We will use the standard "big-O" notation; $O_a(f(n))$ is an unspecified real-valued function such that $|O_a(f(n))| \leq C_a \cdot f(n)$ for some constant C_a that depends on a. For example, writing $f(x) = O(1/k)$ means that f is bounded in absolute value by the function $k \mapsto 1/k$ multiplied by a scalar. In particular, we note that $f \to 0$ as $k \to \infty$. We also use the standard "little-o" notation; if $g(n) = o(f(n))$, then for all $\varepsilon > 0$ there is an integer N such that $|g(n)| \leq \varepsilon f(n)$ whenever $n \geq N$. For example, $1/n = o(1)$.

We write $k_0(\alpha, \beta)$ for a number k_0 that depends on parameters α and β; for example, when we say that something holds for integers $n \geq n_0(\alpha, \beta)$, there is an integer n_0 depending on α and β such that the property holds whenever $n \geq n_0$.

The set $\mathbb{N}$ is the set of natural numbers, $\mathbb{N} = \{1, 2, \ldots\}$. If $\mathcal{X}$ is a vector space over a field K and $A \subset \mathcal{X}$, then the translation of A by $x \in \mathcal{X}$ is defined by $A + x = \{a + x \mid a \in \mathcal{A}\}$ and the scaling by $s \in K$ is defined by $sA = \{sa \mid a \in A\}$. For all finite sets A, the notation $|A|$ stands for the cardinality of A. For intervals I, we use the same notation to denote the number of integers inside I, or the cardinality of $I \cap \mathbb{N}$. We define the subset notation in the following way: $A \subset B$ if and only if $x \in B$ for all $x \in A$. Given a point x in a metric space $(\mathcal{X}, d)$, we denote the open ball of radius r centered in x by $B(x, r) = \{y \in \mathcal{X} \mid d(x, y) < r\}$. The closed ball is denoted by $\overline{B}(x, r) = \{y \in \mathcal{X} \mid d(x, y) \leq r\}$. We denote the diameter of a set E by

$$\operatorname{diam}(E) := \sup\{d(x, y) \mid x, y \in E\}.$$

Note that the diameter of a set equals that of its closure.

0 Measures and probability

Let $\mathcal{X}$ be a set. If Γ is a collection of subsets of $\mathcal{X}$, we call Γ a *σ-algebra* if it has the following properties:

(i.) $\mathcal{X} \in \Gamma$

(ii.) For all $A \in \Gamma$, $\mathcal{X} \setminus A \in \Gamma$

(iii.) If $A_1, A_2, \ldots \in \Gamma$, then $\bigcup_{i=1}^{\infty} A_i \in \Gamma$.

The pair $(\mathcal{X}, \Gamma)$ is called a *measurable space* and the elements of Γ are called *measurable sets*. We call a set mapping $\mu : \Gamma \to [0, \infty]$ a *measure* on (X, Γ) if it has the following properties:

(i.) $\mu(\emptyset) = 0$

(ii.) The mapping μ satisfies σ-additivity: if $A_1, A_2, \ldots$ are disjoint sets in Γ, then $\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$.

If μ is a measure on $(\mathcal{X}, \Gamma)$, the tuple (X, Γ, μ) is called a *measure space.*

We call a set mapping μ^* an *outer measure* on $\mathcal{X}$, if

(i.) $\mu^*(\emptyset) = 0$

(ii.) $\mu^*(A) \leq \mu^*(B)$ for all $A, B \subset \mathcal{X}$

(iii.) $\mu^*(A) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$ if $A \subset \bigcup_{i=1}^{\infty} A_i$.

Sometimes it is more convenient to adapt the one above as the definition of a measure; since a measure can always be extended to an outer measure on the whole space and an outer measure is a measure when restricted to a certain σ-agebra, it is not always necessary to draw a distinction between the two. We do so because in proving certain results we need to assign a measure on a space without any pre-defined σ-algebra. The following is an elementary continuity result for measures.

Theorem 0.1. *Let (X, Γ, μ) be a measure space and $A_1, A_2, \ldots \in \Gamma$ be such that*

(i.) $A_1 \subset A_2 \subset \ldots$. *Then*

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

(ii.) $A_1 \supset A_2 \supset \ldots$ *and* $\mu(A_k) < \infty$ *for some* k*. Then*

$$\mu\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to\infty} \mu(A_n).$$

If Γ is a Borel σ-algebra, that is, the σ-algebra of a topological space $\mathcal{X}$ generated by the collection of open sets and if μ is a measure on $(\mathcal{X}, \Gamma)$, we then say that μ is a Borel measure. In this thesis, we will be dealing exclusively with Borel measures and from now on assume all σ-algebras and measures to be the Borel ones unless stated otherwise. When we say that μ is a measure on a topological space $\mathcal{X}$, the space in question is always equipped with the Borel σ-algebra.

The support of a measure μ, denoted by $\text{supp}(\mu)$, is the largest closed set S such that every open neighbourhood of every point in S has positive μ-measure. We point out as a fact that although $\mu(\mathcal{X} \setminus \text{supp}(\mu)) = 0$ does not necessarily hold in a general case, it is true if μ is a finite Borel measure on a Euclidean space. If the support of μ is contained in a bounded subset of $\mathcal{X}$ and $0 < \mu(\mathcal{X}) < \infty$, we then say that μ is a *mass distribution*. Given a property P defined on a set of points of $\mathcal{X}$, we say that the property P holds *μ-almost everywhere* if the set of points where P does not hold has μ-measure 0. We say that μ is *atomic*, if there is a countable partition of $\mathcal{X}$ such that every element of the partition is either μ-null set (i.e. a set with μ-measure 0) or an atom, that is, a set of positive μ-measure such that every subset of the atom is either a null set or has measure equal to that of the atom. In Euclidian spaces, it can be shown that a measure is atomic if and only if it is a finite or countably infinite linear combination of Dirac measures δ_x, defined by

$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{otherwise} \end{cases}$$

for all A. If μ is a measure on a metric space $(\mathcal{X}, d)$, we call it *uniform* if it is finite and the measure of an open ball depends only on its radius and not on its center, i.e. $\mu(B(x,r)) = \mu(B(y,r))$ for all $r > 0$ and $x, y \in \mathcal{X}$.

If the pairs $(\mathcal{X}, \Gamma)$ and $(\mathcal{Y}, \Delta)$ are measurable spaces and $f : \mathcal{X} \to \mathcal{Y}$ is a mapping such that $f^{-1}(A) \in \Gamma$ for all $A \in \Delta$, we say that f is (Γ-)*measurable*. Measurability without any reference to a σ-algebra should be understood as Borel-measurability. Particularly, if Γ and Δ are Borel σ-algebras, we call f a Borel function.

For all $A \subset \mathcal{X}$, define the *indicator function* $1_A : \mathcal{X} \to \{0, 1\}$, $1_A(x) = \delta_x(A)$.

Definition 0.2. Let $(\mathcal{X}, \Gamma, \mu)$ be a measure space. A function $\tilde{f} : \mathcal{X} \to [0, \infty]$ is *simple*, if $\tilde{f} = \sum_{i \in \mathbb{N}} a_i 1_{A_i}$ for some $a_i \in \mathbb{R}$ and $A_i \in \Gamma$. For all $B \in \Gamma$, the *integral* of $\tilde{f}$ over B is defined by

$$\int_B \tilde{f}(x)\ d\mu(x) = \sum_{i \in \mathbb{N}} a_i \mu(B \cap A_i).$$

For any measurable $f : \mathcal{X} \to [0, \infty]$, define

$$\int_B f(x)\ d\mu(x) = \sup \left\{ \int_B \tilde{f}(x)\ d\mu(x) \mid f \geq \tilde{f} \text{ is simple} \right\}$$

and for measurable $g : \mathcal{X} \to \mathbb{R}$, define

$$\int_B g(x)\ d\mu(x) = \int_B g^+(x)\ d\mu(x) - \int_B -g^-(x)\ d\mu(x),$$

where $g^+ \geq 0$, $g^- \leq 0$ and $g = g^+ + g^-$. If $\int |g|\ d\mu < \infty$, we say that g is integrable.

We often write the integral over $\mathcal{X}$ as $\int f\ d\mu$ when the underlying space is clear from the context. We define the integral for a non-real valued function with countable set of values the same way we define it for a simple function.

Lemma 0.3 (Fatou's lemma)**.** *Let $f_1, f_2, \ldots$ be non-negative, real-valued measurable functions on $\mathcal{X}$ and define the function $f : \mathcal{X} \to [0, \infty)$ by setting $f(x) = \liminf_{n\to\infty} f_n(x)$. Then f is measurable and*

$$\int_{\mathcal{X}} f(x)\ d\mu(x) \leq \liminf_{n\to\infty} \int_{\mathcal{X}} f_n(x)\ d\mu(x).$$

If $f : \mathcal{X} \to \mathcal{Y}$ is measurable, define the *push-forward* of μ through f as the measure $f\mu = \mu \circ f^{-1}$ on $\mathcal{Y}$. The following is an equivalent characterization and is easily seen to be true by inspecting simple functions.

Theorem 0.4. *If μ is a measure on $\mathcal{X}$ and $F : \mathcal{X} \to \mathcal{Y}$ is measurable, then for every measurable $f : \mathcal{Y} \to \mathbb{R}$,*

$$\int_{\mathcal{Y}} f(y)\ dF\mu(y) = \int_{\mathcal{X}} f \circ F(x)\ d\mu(x).$$

Definition 0.5. Let μ and ν be measures on $(\mathcal{X}, \Gamma)$. We call a measurable function $f : \mathcal{X} \to [0, \infty]$ the *density of μ with respect to ν* if for all $A \in \Gamma$,

$$\mu(A) = \int_A f(x)\ d\nu(x).$$

Given measures μ and ν on a measurable space $(\mathcal{X}, \Gamma)$, we say that μ is absolutely continuous with respect to ν, if $\nu(A) = 0$ implies $\mu(A) = 0$ for every $A \in \Gamma$.

Theorem 0.6 (Radon-Nikodym theorem)**.** *If μ and ν are finite measures on $(\mathcal{X}, \Gamma)$ and μ is absolutely continuous with respect to ν, then μ has a density with respect to ν.*

We call μ a *probability measure* on $\mathcal{X}$ if $\mu(\mathcal{X}) = 1$. The measure space $(\mathcal{X}, \Gamma, \mu)$ is then correspondingly called a *probability space*. The elements of Γ are called *events* and for any event E, $\mu(E)$ is called the probability of E. If f is a measurable function on $(\mathcal{X}, \Gamma, \mu)$, we call it a *random variable* distributed according to μ and its *distribution* is defined as the push-forward of μ through f.

Definition 0.7. Let f be a random variable distributed according to μ. The *expected value* of f is defined by

$$\mathbb{E}(f) = \int_{\mathcal{X}} f(x)\ d\mu(x).$$

The *variance* of f is then defined by

$$\operatorname{Var}(f) = \mathbb{E}(f^2) - \mathbb{E}(f)^2 = \mathbb{E}\left((f - \mathbb{E}(f))^2\right).$$

The following result regarding the expected value of a random variable is known as Markov's inequality and will prove very useful.

Theorem 0.8 (Markov's inequality)**.** *Assume f is a positive real-valued random variable on a probability space $(\mathcal{X}, \Gamma, \mu)$. Then for all $a > 0$,*

$$\mu(\{x \in \mathcal{X} \mid f(x) \geq a\}) \leq \frac{\mathbb{E}(f)}{a}.$$

As an application of the theorem above, we bring up the weaker version of the law of large numbers.

Theorem 0.9 (Weak law of large numbers)**.** *Assume $f, f_1, f_2, \ldots$ are independent, real-valued random variables on $(\mathcal{X}, \Gamma, \mu)$ with finite expected values $\mathbb{E}(f_1) = \mathbb{E}(f_2) = \ldots = \mathbb{E}(f)$. Denote by $\overline{f_n}$ the mean value, $\overline{f_n} = \frac{1}{n}\sum_{i=1}^n f_i$. Then, for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mu\left(\{x \in \mathcal{X} \mid |\overline{f_n}(x) - \mathbb{E}(f)| > \varepsilon\}\right) = 0.$$

We may also define the expected value (or mean) and variance of a probability measure.

Definition 0.10. For a probability measure μ on $\mathcal{X}$, define the *mean* by

$$\langle \mu \rangle = \int_{\mathcal{X}} x\ d\mu(x).$$

The variance of μ is then defined by

$$\operatorname{Var}(\mu) = \int_{\mathcal{X}} (x - \langle \mu \rangle)^2\ d\mu(x).$$

We denote the set of all probability measures on $\mathcal{X}$ by $\mathcal{P}(\mathcal{X})$.

Consider the space $C(\mathcal{X})$ of all continuous, bounded real-valued functions on $\mathcal{X}$. This clearly forms a norm space over $\mathbb{R}$ with the norm $\|f\| = \sup_{x\in\mathcal{X}} |f(x)|$. By Riesz representation theorem, we may identify $\mathcal{P}(\mathcal{X})$ as a subspace of the Banach dual of continuous linear functionals on $C(\mathcal{X})$ by letting μ operate on f by $\int f\, d\mu$. We may often consider the weak-* topology on $\mathcal{P}(\mathcal{X})$ generated by $C(\mathcal{X})$; in this topology, a sequence of probability measures $(\mu_n)_{n\in\mathbb{N}}$ converges to μ if and only if $\lim_{n\to\infty} \int f\, d\mu_n = \int f\, d\mu$ for all $f \in C(\mathcal{X})$. We then say that $(\mu_n)_n$ converges weakly to μ.

We remark that if $\mathcal{X}$ is compact, then the space $\mathcal{P}(\mathcal{X})$ is compact in the weak-* topology. This is because by Banach-Alaoglu theorem the unit ball of $(C(\mathcal{X}))^*$ is compact in the weak-* topology and since $\mathcal{P}(\mathcal{X})$ belongs to the unit sphere (in the dual norm $\|\phi\| = \sup_{\|f\|=1} |\phi(f)|$), it suffices to see that it is weak-* closed. This is, however, easy to see: if μ is the weak limit of $(\mu_n)_n$, then $\mu(\mathcal{X}) = \int 1\, d\mu = \lim_{n\to\infty} \int 1\, d\mu_n = 1$. The following theorem gives a well-known equivalent definition for the weak convergence, one that will be more useful for us in practice.

Theorem 0.11. *If $\mu, \mu_1, \mu_2, \dots$ are probability measures on a topological space $\mathcal{X}$, the sequence (μ_n) converges weakly (converges in the weak-* topology) to the probability measure μ if*

$$\liminf_{n\to\infty} \mu_n(U) \geq \mu(U)$$

for all open sets $U \subset \mathcal{X}$.

Another way to give structure to the space $\mathcal{P}(\mathcal{X})$ is by extending it into a metric space.

Definition 0.12. Let μ and ν be probability measures on $(\mathcal{X}, \Gamma)$. Define the *total variation distance* between μ and ν as

$$\|\mu - \nu\| = \sup_{E\in\Gamma} |\mu(E) - \nu(E)|.$$

Theorem 0.13. *The space $(\mathcal{P}(\mathcal{X}), \|\cdot\|)$ is a complete metric space.*

Proof. It is clear that $\|\cdot\|$ is indeed a metric in $\mathcal{P}(\mathcal{X})$. Let $(\mu_n)_{n\in\mathbb{N}}$ be a Cauchy sequence in $\mathcal{P}(\mathcal{X})$, i.e. for every $\varepsilon > 0$ there exists an integer N such that $\|\mu_n - \mu_m\| < \varepsilon$ for all $n, m \geq N$. For every $E \in \Gamma$, the sequence $(\mu_n(E))_{n\in\mathbb{N}}$ is then a Cauchy sequence in $\mathbb{R}$. By completeness of $\mathbb{R}$, we find a converging subsequence $(\mu_{n_k}(E))_{k\in\mathbb{N}}$. Let μ be the set function $E \mapsto \lim_{k\to\infty} \mu_{n_k}(E)$ for all $E \in \Gamma$; note that the convergence in k is uniform over $E \in \Gamma$. We claim that μ is a probability measure.

Clearly $\mu(\emptyset) = 0$ and $\mu(\mathcal{X}) = 1$. Let $A_1, A_2, \dots$ belong to Γ and be disjoint. Since the convergence of $(\mu_{n_k}(E))_k$ is uniform in E, for any $\varepsilon > 0$ we may choose k_0 so that $|\mu(E) - \mu_{n_k}(E)| < \varepsilon$ for all $E \in \Gamma$ and $k \geq k_0$. Clearly μ is finitely additive. Let N_0 be so big that $\sum_{i=N_0}^{\infty} \mu_{n_{k_0}}(A_i) < \varepsilon$. Then we have

$$\begin{aligned}\mu\left(\bigcup_{i=1}^{\infty} A_i\right) &= \mu\left(\bigcup_{i=1}^{N_0} A_i\right) + \mu\left(\bigcup_{i=N_0+1}^{\infty} A_i\right) \\ &\leq \sum_{i=1}^{N_0} \mu(A_i) + \sum_{i=N_0}^{\infty} \mu_{n_{k_0}}(A_i) + \varepsilon \\ &\leq \sum_{i=1}^{N_0} \mu(A_i) + 2\varepsilon.\end{aligned}$$

From the first inequality we see that also $\mu(\bigcup_{i=1}^{\infty} A_i) \geq \sum_{i=1}^{N_0} \mu(A_i)$. Therefore letting $\varepsilon \to 0$ and $N_0 \to \infty$ we obtain countable additivity. Moreover, μ_n converges to μ in the total variation metric: if $\varepsilon' > 0$, then

$$\begin{aligned}\|\mu - \mu_n\| = \|\mu - \mu_{n_k} + \mu_{n_k} - \mu_n\| &\leq \|\mu - \mu_{n_k}\| + \|\mu_{n_k} - \mu_n\| \\ &\leq 2\varepsilon \leq \varepsilon',\end{aligned}$$

when $\varepsilon \leq \varepsilon'/2$ and n, k are so large that $\|\mu - \mu_{n_k}\| = \sup_{E \in \Gamma} |\mu(E) - \mu_{n_k}(E)| < \varepsilon$ and $\|\mu_{n_k} - \mu_n\| < \varepsilon$. Therefore $(\mathcal{P}(\mathcal{X}), \|\cdot\|)$ is complete. □

Given multiple measure spaces, we can define a measure on their product space as follows.

Definition 0.14. Let $(\mathcal{X}_i, \Gamma_i, \mu_i)$ be measure spaces for $i = 1, \dots, k$. Denoting by $\bigotimes_{i=1}^{k} \Gamma_i$ the σ-algebra on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_k$ generated by sets $\mathcal{E}_1 \times \cdots \times \mathcal{E}_k$ where $\mathcal{E}_i \in \Gamma_i$, the *product measure* $\mu_1 \times \cdots \times \mu_k$ is a measure on $\mathcal{X}_1 \times \cdots \times \mathcal{X}_k$ that satisfies

$$(\mu_1 \times \cdots \times \mu_k)(E_1 \times \cdots \times E_k) = \prod_{i=1}^{k} \mu_i(E_i)$$

for all $E_i \in \Gamma_i$.

If f and g are random variables on $\mathcal{X}$ and $\mathcal{Y}$, distributed according to probability measures μ and ν respectively, we call the random variable (f, g) on $\mathcal{X} \times \mathcal{Y}$ their *joint*. The random variables f and g are *independent*, if their joint is distributed according to $\mu \times \nu$.

Theorem 0.15 (Fubini's theorem). *Let $(\mathcal{X}_1, \Gamma_1, \mu_1)$ and $(\mathcal{X}_2, \Gamma_2, \mu_2)$ be σ-finite measure spaces, let $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be $\mu_1 \times \mu_2$ -measurable and suppose that $\int_{\mathcal{X}_1} \left(\int_{\mathcal{X}_2} f(x, y)\ d\mu_2(y) \right) d\mu_1(x) < \infty$. Then*

$$\begin{aligned}\int_{\mathcal{X}_1 \times \mathcal{X}_2} f(x, y)\ d(\mu_1 \times \mu_2)(x, y) &= \int_{\mathcal{X}_1} \left(\int_{\mathcal{X}_2} f(x, y)\ d\mu_2(y) \right) d\mu_1(x) \\ &= \int_{\mathcal{X}_2} \left(\int_{\mathcal{X}_1} f(x, y)\ d\mu_1(x) \right) d\mu_2(y).\end{aligned}$$

Corollary 0.16. *Assume the space $\mathcal{X}_1 \times \mathcal{X}_2$ is σ-finite and $A \in \Gamma_1 \otimes \Gamma_2$. Then*

$$(\mu \times \nu)(A) = \int_{\mathcal{X}_2} \mu(A_y)\ d\nu = \int_{\mathcal{X}_1} \nu(A^x)\ d\mu = \int_{\mathcal{X}_1} \int_{\mathcal{X}_2} 1_A(x, y)\ d\nu(y) d\mu(x),$$

where $A_y = \{x \in \mathcal{X}_1 \mid (x, y) \in A\}$ and $A^x = \{y \in \mathcal{X}_2 \mid (x, y) \in A\}$.

Some of the main results in this thesis concern the push-forward of the product measure through the addition map, known as the convolution measure.

Definition 0.17. Let $\mu_1, \ldots, \mu_k$ be finite measures on a vector space $\mathcal{X}$ and let $\pi : \mathcal{X} \times \cdots \times \mathcal{X} \to \mathcal{X}$ be the addition mapping, $\pi(x_1, \ldots, x_k) = \sum_{i=1}^k x_i$. The *convolution* of $\mu_1, \ldots, \mu_k$ is the push-forward measure

$$\mu_1 * \cdots * \mu_k = F(\mu_1 \times \cdots \times \mu_k) = (\mu_1 \times \cdots \times \mu_k) \circ F^{-1}.$$

For repeated self-convolutions, we use the notation $\mu^{*k} = \mu * \cdots * \mu$. Clearly convolution is a commutative and multilinear operator and the convolution of probability measures is a probability measure. Note that if f and g are independent random variables with distributions μ and ν respectively, it follows from the definition that the distribution of $X + Y$ is $\mu * \nu$. The proposition below is immediate from the definition and Fubini's theorem.

Proposition 0.18. *For every $A \in \Gamma$, $a, b \in \mathbb{R}$, the convolution $\mu * \nu$ satisfies the following.*

(i.) $(\mu * \nu)(A) = \int \int 1_A(x + y)\ d\mu(x) d\nu(y)$.

(ii.) $(\mu * \nu)(A) = \int \mu(A - x)\ d\nu(x)$

The following proposition gives a way to calculate the mean and variance of a convolution measure.

Proposition 0.19. *If $\mu_1, \ldots, \mu_k \in \mathcal{P}(\mathcal{X})$, the convolution $\mu = \mu_1 * \cdots * \mu_k$ has a mean*

$$\langle \mu \rangle = \sum_{i=1}^{k} \langle \mu_i \rangle$$

and variance

$$Var(\mu) = \sum_{k=1}^{k} Var(\mu_i).$$

One particular probability measure that will be of use to us is the Gaussian measure.

Definition 0.20. The *Gaussian measure* with mean m and variance σ^2 is given by

$$\gamma_{m,\sigma^2}(A) = \int_A \varphi\left(\frac{x-m}{\sigma^2}\right)\, dx,$$

where $\varphi(x) = \sqrt{2\pi}\exp(-\frac{1}{2}|x|^2)$, for every $A \in \mathbb{R}$.

An application of the central limit theorem asserts that given a large number of probability measures of positive variance, their convolution can be rescaled so that the resulting measure approaches the Gaussian measure. We bring up the Berry-Esseen theorem that quantifies the rate of this convergence. The following variant is found in [2].

Theorem 0.21. *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 for any interval $I \subset \mathbb{R}$,*

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

where $C \in \mathbb{R}$ is independent of k. In particular, if $\rho_i \leq C'$ and $\sum_{i=1}^{k} Var(\mu) \geq ck$ for constants $c, C' > 0$, then

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

0.1 Components of a measure

Throughout this thesis, a common technique we use in dealing with the properties of a general measure is to study its small-scale components. It turns out that most of the properties we are interested in, if shown to hold for a randomly chosen component of the measure, can be generalized to hold for the original measure with probability close to 1.

Indeed, we define a *(countable) partition* of a set $\mathcal{X}$ as a countable family $\mathcal{E} = \{E_i\}_{i\in\mathbb{N}}$ such that the sets E_i are disjoint and $\mathcal{X} = \bigcup_i E_i$. The sets E_i are called the *atoms* of the partition $\mathcal{E}$. Since we are primarily moving in Euclidean spaces, we introduce the dyadic partition of $\mathbb{R}^d$. Denote by $\mathcal{D}_n$ the partition of the line into subintervals of length 2^{-n}:

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

For any $t \in \mathbb{R}$, write $\mathcal{D}_t = \mathcal{D}_{\lfloor t \rfloor}$, where $\lfloor t \rfloor = \max\{k \in \mathbb{Z} \mid k \leq t\}$. The dyadic partition of $\mathbb{R}^d$ is then defined as $\mathcal{D}_n^d = \mathcal{D}_n \times \cdots \times \mathcal{D}_n$. We will sometimes omit the superscript d when the dimension of the space is clear from the context.

For all x in $\mathbb{R}^d$, let $\mathcal{D}_n(x)$ denote the unique cell of the partition $\mathcal{D}_n$ that contains x. For an atom D, let $T_D : \mathbb{R}^d \to \mathbb{R}^d$ denote the unique homothety mapping D to $[0,1)^d$; for example, if $D = [k/2^n, (k+1)/2^n)^d$ is an atom in the dyadic partition of $\mathbb{R}^d$, then the homothety T_D is the affine mapping $x \mapsto 2^n x - (k, \dots, k)$. Given a probability measure on an atom of $\mathcal{D}_n$, we may rescale it into a measure on the unit cube by taking its push-forward through this homothety.

Definition 0.22. For a probability measure μ on $\mathbb{R}^d$ and $D \in \mathcal{D}_n$ with $\mu(D) > 0$, the *raw D-component of μ* is

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

and the *rescaled D-component of μ* 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 are called the (raw or rescaled) *level-n components of μ* for every x with $\mu(\mathcal{D}_n(x)) > 0$.

We will mostly be interested in random selection of a small-scale component, each one drawn with the probability that μ assigns to the atom in question.

Definition 0.23. Let μ be a probability measure on $\mathbb{R}^d$.

(i.) Define the raw (or rescaled) random level-n component of μ as the discrete random variable $\mathbb{R}^d \to \mathcal{P}(\mathbb{R}^d)$, $x \mapsto \mu_{x,n}$ (or $x \mapsto \mu^{x,n}$), distributed according to μ.

(ii.) If $I \subset \mathbb{N}$ is a finite set of indices, then the random level-I component, raw or rescaled, is chosen by first drawing an n uniformly from I and then choosing the level-n component independently on the choice of n, as described above.

Note that the raw and rescaled components of a product measure $\mu \times \nu$ are the products of the components of μ and ν; if $\mathcal{D}_n^2 = \mathcal{D}_n \times \mathcal{D}_n$ is the dyadic partition of $\mathbb{R}^{2d}$ and A is a subset of $\mathbb{R}^{2d}$, then

$$\begin{aligned}(\mu \times \nu)_{(x,y),n}(A) &= \frac{1}{(\mu \times \nu)(\mathcal{D}_n(x,y))}(\mu \times \nu)(\mathcal{D}_n(x,y) \cap A) \\ &= \frac{1}{\mu(\mathcal{D}_n(x))\nu(\mathcal{D}_n(y))} \int_{\mathbb{R}^d \cap \mathcal{D}_n(y)} \mu(\mathcal{D}_n(x) \cap A_t)\ d\nu(t) \\ &= \frac{1}{\mu(\mathcal{D}_n(x))\nu(\mathcal{D}_n(y))}(\mu_{|\mathcal{D}_n(x)} \times \nu_{|\mathcal{D}_n(y)})(A) \\ &= (\mu_{x,n} \times \nu_{y,n})(A).\end{aligned}$$

Similarly, if $T_{\mathcal{D}_n(x)}$ and $T_{\mathcal{D}_n(y)}$ are the homotheties mapping $\mathcal{D}_n(x)$ and $\mathcal{D}_n(y)$ to $[0,1)^d$, then $(a,b) \mapsto (T_{\mathcal{D}_n(x)}(a), T_{\mathcal{D}_n(y)}(b))$ is the corresponding homothety of $\mathcal{D}_n(x,y)$ and

$$\begin{aligned}(\mu \times \nu)^{(x,y),n}(D) &= \frac{1}{(\mu \times \nu)(\mathcal{D}_n(x,y))}(\mu \times \nu)(\mathcal{D}_n(x,y) \cap T^{-1}_{\mathcal{D}_n(x,y)}(D)) \\ &= \frac{1}{\mu(\mathcal{D}_n(x))\nu(\mathcal{D}_n(y))}(\mu_{|\mathcal{D}_n(x)} \times \nu_{|\mathcal{D}_n(y)})(T^{-1}_{\mathcal{D}_n(x,y)}(D)) \\ &= (\mu^{x,n} \times \nu^{y,n})(D).\end{aligned}$$

For a family $\mathcal{A}$ of probability measures on $[0,1)^d$, we introduce the notation

$$\mathbb{P}_{i=n}(\mu^{x,i} \in \mathcal{A}) = \int_{\mathbb{R}^d} 1_{\mathcal{A}}(\mu^{x,n})\ d\mu(x).$$

If instead of fixing the scale i we restrict it to a set of integers, it is chosen according to the uniform distribution u and by the law of total probability,

$$\mathbb{P}_{0\le i\le n}(\mu^{x,i} \in \mathcal{A}) = \sum_{k=0}^{n} \mathbb{P}_{i=k}(\mu^{x,i} \in \mathcal{A})u(i=l) = \frac{1}{n+1}\sum_{i=0}^{n} \int_{\mathbb{R}^d} 1_{\mathcal{A}}(\mu^{x,i})\ d\mu(x).$$

Similarly, if $f : \mathcal{P}([0,1)^d) \to \mathbb{R}$ and $I \subset \mathbb{N}$, then by the law of total expectation

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

Generally, when dealing with components of several measures μ, ν, we assume all the choices of components $\mu^{x,i}, \nu^{y,j}$ to be independent;

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

When we fix the scale n, the random level-n component gives averagely correct estimates for μ, as seen in the following lemma.

Lemma 0.24. *If μ is a probability measure on $\mathbb{R}^d$ and $n \in \mathbb{N}$,*

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

Proof. The component $\mu_{x,i}$ is a random variable $\mathbb{R}^d \to \mathcal{P}(\mathbb{R}^d)$ admitting a constant value μ_D at each $D \in \mathcal{D}_n$. Hence by definition

$$\mathbb{E}_{i=n}(\mu_{x,i}) = \int_{\mathbb{R}^d} \mu_{x,n}\ d\mu(x) = \sum_{D\in\mathcal{D}_n} \mu_D\mu(D) = \sum_{D\in\mathcal{D}_n} \mu_{|D} = \mu.$$

□

1 Entropy of a measure

Entropy is often conceived as a measurement of chaos, or uncertainty, in a system. In our context, we are interested in how evenly a probability measure emphasizes the atoms of a given (usually the dyadic) partition.

Take, for example, a roll of a die with k faces. If we wanted the outcome of a throw to be as random as possible, we would emphasize the faces so that each of them would have an equal probability, namely $1/k$. Bearing this intuition in mind, we replace the die with an arbitrary set $\mathcal{X}$, the faces represented by a partition of the set, $\mathcal{E}$. Then we emphasize the atoms of this partition by assigning a probability measure on $\mathcal{X}$ and define a quantity that takes into account how evenly μ is distributed on the atoms of $\mathcal{E}$.

Definition 1.1. Let μ be a probability measure on $\mathcal{X}$ and let $\mathcal{E}$ be a countable, measurable partition of $\mathcal{X}$. The *entropy* of μ with respect to $\mathcal{E}$ is defined by

$$H(\mu,\mathcal{E}) = -\sum_{E\in\mathcal{E}} \mu(E)\log_2 \mu(E).$$

Here the logarithm is in base 2 and we define $0 \log_2 0 = 0$. The *conditional entropy* with respect to a countable partition $\mathcal{F}$ is

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

where $\mu_F = \dfrac{1}{\mu(F)} \mu|_F$ is the conditional measure on F.

If the probability measure μ is discrete, we write $H(\mu)$ for the entropy with respect to the partition into its atoms, and for a probability vector $\alpha = (\alpha_1, \ldots, \alpha_k)$ we write $H(\alpha) = -\sum \alpha_i \log_2 \alpha_i$.

We define the entropy of a random variable as the entropy of its distribution. Since we will only encounter it in the context of discrete random variables, we will only give the definition in that case, omitting the definition of differential entropy.

Indeed, if X and Y are discrete random variables on $\mathcal{X}$ with distributions μ and ν, respectively, the entropy of X is then defined by

$$H(X) = -\sum_{x_i} \mu(x_i) \log_2 \mu(x_i),$$

where x_i ranges over all possible values of X. If η is the distribution of the joint of X and Y, the conditional entropy of X given Y is defined by

$$H(X|Y) = \sum_{y_i} \nu(y_i) H(X|Y = y_i),$$

where $X|Y = y_i$ is the random variable $\omega \mapsto X(\omega)$ with the conditioned distribution $\gamma : A \mapsto \dfrac{\eta(A, y_i)}{\nu(y_i)}$ for every A in the associated σ-algebra. The following lemma collects some properties of the conditional entropy of a random variable.

Lemma 1.2. *Let X and Y be discrete random variables on $\mathcal{X}$ and $\mathcal{Y}$ with distributions μ and ν, respectively.*

(i.) If η is the distribution of the joint (X, Y),

$$H(X|Y) = \sum_{x_i, y_j} \eta(x_i, y_j) \log_2 \frac{\nu(y_j)}{\eta(x_i, y_j)}.$$

(ii.) Conditioning a random variable never increases its entropy:

$$H(X|Y) \leq H(X).$$

If X and Y are independent, the equality holds.

(iii.) If $H(X)$ and $H(Y)$ are finite, the conditional entropy obeys the following "chain rule":

$$H(X|Y) = H(X, Y) - H(Y).$$

In the proof of this lemma, we require a general result regarding convex functions, known as Jensen's inequality.

Lemma 1.3 (Jensen's inequality). *Assume f is a real-valued convex function, $x_1, \ldots, x_n$ are in its domain and $a_i > 0$ for all i. Then,*

$$f\left(\frac{\sum_{i=1}^n a_i x_i}{\sum_{i=1}^n a_i}\right) \leq \frac{\sum_{i=1}^n a_i f(x_i)}{\sum_{i=1}^n a_i}.$$

Proof. We commence the proof by induction, assuming first that the sum of a_i equals 1. Let λ_1, λ_2 be non-negative real numbers such that $\lambda_1 + \lambda_2 = 1$. By convexity of f, we have

$$f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2).$$

Make then a hypothesis that

$$f\left(\sum_{i=1}^n \lambda_i x_i\right) \leq \sum_{i=1}^n \lambda_i f(x_i) \tag{1.1}$$

for an $n > 2$, where $\sum_{i=1}^n \lambda_i = 1$. Then, if $\sum_{i=1}^{n+1} \lambda_i = \sum_{i=1}^n \lambda_i + \lambda_{n+1} = 1$, by convexity and the hypothesis, we have

$$f\left(\sum_{i=1}^{n+1} \lambda_i x_i\right) = f\left(\sum_{i=1}^n \lambda_i x_i + \lambda_{n+1} x_{n+1}\right) \leq \sum_{i=1}^n \lambda_i f(x_i) + \lambda_{n+1} f(x_{n+1}).$$

Therefore, by induction, (1.1) holds for an arbitrary n. To lift the requirement of a_i summing to 1, note that if $a_1, \ldots, a_n$ are strictly positive, replacing λ_i with $\dfrac{a_i}{\sum_{i=1}^n a_i}$ gives the desired inequality. $\square$

We can now prove Lemma 1.2.

Proof of Lemma 1.2. Let X and Y be random variables with distributions μ and ν, respectively, and let η be the distribution of their joint random variable.

(i.) Immediately from the definition, we obtain

$$\begin{aligned} H(X|Y) &= \sum_{y_i} \nu(y_i) H(X|Y = y_i) \\ &= -\sum_{y_i} \sum_{x_i} \nu(y_i) \frac{\eta(x_i, y_i)}{\nu(y_i)} \log_2 \frac{\eta(x_i, y_i)}{\nu(y_i)} \\ &= \sum_{x_i, y_j} \eta(x_i, y_j) \log_2 \frac{\nu(y_j)}{\eta(x_i, y_j)}. \end{aligned}$$

(ii.) Using Jensen's inequality for $\log_2$,

$$\begin{aligned} H(X|Y) &= \sum_{x_i, y_j} \eta(x_i, y_j) \log_2 \frac{\nu(y_j)}{\eta(x_i, y_j)} \\ &= \sum_{x_i, y_j} \mu(x_i) \frac{\eta(x_i, y_j)}{\mu(x_i)} \log_2 \frac{\nu(y_j)}{\eta(x_i, y_j)} \\ &= \sum_{x_i} \mu(x_i) \sum_{y_j} \frac{\eta(x_i, y_j)}{\mu(x_i)} \log_2 \frac{\nu(y_j)}{\eta(x_i, y_j)} \\ &\leq \sum_{x_i} \mu(x_i) \log_2 \left(\sum_{y_j} \frac{\eta(x_i, y_j)}{\mu(x_i)} \frac{\nu(y_j)}{\eta(x_i, y_j)} \right) \\ &= H(X). \end{aligned}$$

If X and Y are independent,

$$H(X|Y) = -\sum_{x_i, y_j} \mu(x_i)\nu(y_j) \log_2 \mu(x_i) = H(X).$$

(iii.) Assuming $H(X)$ and $H(Y)$ are finite,

$$\begin{aligned} H(X|Y) &= \sum_{x_i, y_j} \eta(x_i, y_j) \log_2 \frac{\nu(y_j)}{\eta(x_i, y_j)} \\ &= -\sum_{x_i, y_j} \eta(x_i, y_j) \log_2 \eta(x_i, y_j) + \sum_{x_i, y_j} \eta(x_i, y_j) \log_2 \nu(y_j) \\ &= H(X, Y) - H(Y). \end{aligned}$$

This completes the proof. □

We refer to the entropy with respect to the partition $\mathcal{D}_n$ as the scale-n entropy. The normalized scale-n entropy for a probability measure is defined as

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

Below are listed some basic properties of the entropy of a probability measure that we will be using constantly.

Lemma 1.4. *Let α be a probability vector, let μ, ν be probability measures on $\mathcal{X}$ and let $\mathcal{E}, \mathcal{F}$ be countable, measurable partitions of $\mathcal{X}$. Entropy has the following properties:*

(i.) *$H(\mu, \mathcal{E}) \geq 0$, where equality holds if and only if the support of μ is contained in a single atom of $\mathcal{E}$.*

(ii.) *If μ is supported on k atoms of $\mathcal{E}$, then $H(\mu, \mathcal{E}) \leq \log_2 k$ with equality if and only if μ is uniform.*

(iii.) *If $\mathcal{F}$ refines $\mathcal{E}$, then $H(\mu, \mathcal{F}) \geq H(\mu, \mathcal{E})$.*

(iv.) *If $\mathcal{E} \vee \mathcal{F} = \{E \cap F \mid E \in \mathcal{E},\ F \in \mathcal{F}\}$ is the join of $\mathcal{E}$ and $\mathcal{F}$, then $H(\mu, \mathcal{E} \vee \mathcal{F}) = H(\mu, \mathcal{F}) + H(\mu, \mathcal{E}|\mathcal{F})$.*

(v.) *The functions $\mu \mapsto H(\mu, \mathcal{E})$ and $\mu \mapsto H(\mu, \mathcal{E}|\mathcal{F})$ are concave.*

(vi.) *The function $\mu \mapsto H(\mu, \mathcal{E})$ obeys the "convexity" bound*

$$H\left(\sum_{i=1}^{k} \alpha_i \mu_i, \mathcal{E}\right) \leq \sum_{i=1}^{k} \alpha_i H(\mu_i, \mathcal{E}) + H(\alpha).$$

Proof. Write $\mathcal{E} = \{E_i\}_{i \in \mathbb{N}}$ and $\mathcal{F} = \{F_j\}_{j \in \mathbb{N}}$.

(i.) Clearly

$$H(\mu, \mathcal{E}) = -\sum_{i \in \mathbb{N}} \mu(E_i) \log_2 \mu(E_i) \geq 0,$$

since $0 \leq \mu(E_i) \leq 1$ for all E_i and therefore $-\log_2 \mu(E_i) \geq 0$ for all E_i. If $H(\mu, \mathcal{E}) = 0$, it means that every term in the sum equals 0, which in turn implies that $\mu(E_i) = 1$ for exactly one $i \in \mathbb{N}$ and $\mu(E_j) = 0$ otherwise.

(ii.) Assume $\mu(E_i) > 0$ for $i \in \{1, 2, \ldots, k\}$ and $\mu(E_i) = 0$ otherwise. Then, since the function $x \mapsto -\log_2 x$ is convex and $\sum_{i=1}^k \mu(E_i) = 1$, we have from Jensen's inequality that

$$-\log_2 k = -\log_2 \left(\sum_{n=1}^{k} \frac{1}{\mu(E_i)} \mu(E_i) \right) \leq \sum_{i=1}^{k} -\log_2 \left(\frac{1}{\mu(E_i)} \right) \mu(E_i)$$
$$= \sum_{i=1}^{k} \log_2 (\mu(E_i)) \mu(E_i) = -H(\mu, \mathcal{E}).$$

If μ is uniform, i.e. $\mu(E_i) = 1/k$ for all $1 \leq i \leq k$,

$$\begin{aligned} H(\mu, \mathcal{E}) &= -\sum_{i=1}^{k} \frac{1}{k} \log_2 \frac{1}{k} \\ &= \sum_{i=1}^{k} \frac{\log_2 k}{k} \\ &= \log_2 k. \end{aligned}$$

and by differentiation we can see this is the only maximum point of the function $(x_1, \ldots, x_k) \mapsto \sum_{i=1}^k -x_i \log_2 x_i$ in the domain satisfying $\sum_{i=1}^k x_i = 1$.

(iii.) Assume $\mathcal{F}$ refines $\mathcal{E}$. Then, for every $E_i \in \mathcal{E}$ there is a countable $J_i \subset \mathbb{N}$ such that $E_i = \bigcup_{j \in J_i} F_j$. Therefore

$$\begin{aligned} H(\mu, \mathcal{E}) &= -\sum_{i \in \mathbb{N}} \mu(E_i) \log_2 \mu(E_i) = -\sum_{i \in \mathbb{N}} \mu \left(\bigcup_{j \in J_i} F_j \right) \log_2 \mu \left(\bigcup_{k \in J_i} F_k \right) \\ &= -\sum_{i \in \mathbb{N}} \sum_{j \in J_i} \mu(F_j) \log_2 \left(\sum_{k \in J_i} \mu(F_k) \right) \\ &\leq -\sum_{i \in \mathbb{N}} \sum_{j \in J_i} \mu(F_j) \log_2 \mu(F_j) = H(\mu, \mathcal{F}), \end{aligned}$$

where the inequality follows from the fact that $-\log_2$ is decreasing.

(iv.) Let $\mathcal{E} \vee \mathcal{F}$ be the join of $\mathcal{E}$ and $\mathcal{F}$. Then,

$$\begin{aligned} H(\mu, \mathcal{E}|\mathcal{F}) &= \sum_{j \in \mathbb{N}} \mu(F_j) \cdot H(\mu_{F_j}, \mathcal{E}) = -\sum_{j \in \mathbb{N}} \sum_{i \in \mathbb{N}} \mu(F_j) \mu_{F_j}(E_i) \log_2 \mu_{F_j}(E_i) \\ &= \sum_{j \in \mathbb{N}} \sum_{i \in \mathbb{N}} \mu(E_i \cap F_j) \log_2 \mu(F_j) - \sum_{j \in \mathbb{N}} \sum_{i \in \mathbb{N}} \mu(E_i \cap F_j) \log_2 \mu(E_i \cap F_j) \\ &= -H(\mu, \mathcal{F}) + H(\mu, \mathcal{E} \vee \mathcal{F}). \end{aligned}$$

(v.) Since $\frac{d^2}{dx^2}(-x\log_2 x) = \frac{d^2}{dx^2}\frac{-x\log x}{\log 2} = \frac{-1}{x\log 2} < 0$ for all $x > 0$, the function $x \mapsto -x\log_2 x$ is concave on every subinterval of $(0,\infty)$. Hence $\mu \mapsto H(\mu, \mathcal{E})$ is concave on $\mathcal{P}(\mathcal{X})$ as a sum of concave functions. To prove the concavity of conditional entropy, we observe the following: if μ and ν are probability measures on $\mathbb{R}^d$, $0 < t < 1$ and $\tau = t\mu + (1-t)\nu$, for any $F \in \mathcal{F}$ and Borel E,

$$\begin{aligned}\tau_F(E) = \frac{1}{\tau(F)}\tau_{|F}(E) &= \frac{t\mu_{|F}(E) + (1-t)\nu_{|F}(E)}{t\mu(F) + (1-t)\nu(F)} \\ &= \frac{t\mu(F)}{\tau(F)}\mu_F(E) + \frac{(1-t)\nu(F)}{\tau(F)}\nu_F(E),\end{aligned}$$

so τ_F is a convex combination of μ_F and ν_F. Therefore, by concavity of entropy,

$$\begin{aligned}H(t\mu + (1-t)\nu, \mathcal{E}|\mathcal{F}) &= \sum_{F\in\mathcal{F}} \tau(F)H(\tau_F, \mathcal{E}) \\ &\geq \sum_{F\in\mathcal{F}} \tau(F)\left(\frac{t\mu(F)}{\tau(F)}H(\mu_F, \mathcal{E}) + \frac{(1-t)\nu(F)}{\tau(F)}H(\nu_F, \mathcal{E})\right) \\ &= t\sum_{F\in\mathcal{F}} \mu(F)H(\mu_F, \mathcal{E}) + (1-t)\sum_{F\in\mathcal{F}} \nu(F)H(\nu_F, \mathcal{E}) \\ &= tH(\mu, \mathcal{E}|\mathcal{F}) + (1-t)H(\nu, \mathcal{E}|\mathcal{F}).\end{aligned}$$

vi. Assume α is a probability vector. Then,

$$\begin{aligned}H\left(\sum_{i=1}^k \alpha_i\mu_i, \mathcal{E}\right) &= -\sum_{l\in\mathbb{N}}\sum_{i=1}^k \alpha_i\mu_i(E_l)\log_2\left(\sum_{j=1}^k \alpha_j\mu_j(E_l)\right) \\ &\leq -\sum_{l\in\mathbb{N}}\sum_{i=1}^k \alpha_i\mu_i(E_l)\log_2(\alpha_i\mu_i(E_l)) \\ &= \sum_{i=1}^k \alpha_i\left(-\sum_{l\in\mathbb{N}} \mu_i(E_l)(\log_2\alpha_i + \log_2\mu_i(E_l))\right) \\ &= \sum_{i=1}^k \alpha_i H(\mu_i, \mathcal{E}) - \sum_{i=1}^k \alpha_i \sum_{l\in\mathbb{N}} \mu_i(E_l)\log_2\alpha_i \\ &= \sum_{i=1}^k \alpha_i H(\mu_i, \mathcal{E}) - \sum_{i=1}^k \alpha_i\log_2\alpha_i,\end{aligned}$$

which completes the proof.

$\square$

From the lemma we obtain some immediate bounds for a probability measure on the unit interval (or cube, in $\mathbb{R}^d$); following from the second statement, for $\mu \in \mathcal{P}([0,1]^d)$ we have

$$H(\mu, \mathcal{D}_m) \leq md \tag{1.2}$$

since there are $(2^m)^d$ atoms of $\mathcal{D}_m$ inside the cube $[0,1]^d$. For the corresponding bound for conditional entropy, observe that any atom of $\mathcal{D}_n$ contains $(2^{-n}/2^{-m-n})^d = 2^{md}$ atoms of the partition $\mathcal{D}_{n+m}$; hence

$$H(\mu, \mathcal{D}_{n+m}|\mathcal{D}_n) = \sum_{D \in \mathcal{D}_n} \mu(D) \cdot H(\mu_D, \mathcal{D}_{n+m}) \leq md.$$

The next lemma presents some continuity properties for the entropy.

Lemma 1.5. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and let $\mathcal{E}, \mathcal{F}$ be partitions of $\mathbb{R}^d$.*

(i.) *If $m \in \mathbb{N}$, $K \subset \mathbb{R}^d$ is compact and $\mu \in \mathcal{P}(K)$, for every $\varepsilon > 0$ there is a neighbourhood $U \subset \mathcal{P}(K)$ of μ such that $|H(\nu, \mathcal{D}_m) - H(\mu, \mathcal{D}_m)| < \varepsilon$ for $\nu \in U$.*

(ii.) *If each $E \in \mathcal{E}$ intersects at most k elements of $\mathcal{F}$ and vice versa, then $|H(\mu, \mathcal{E}) - H(\mu, \mathcal{F})| = O(\log_2 k)$.*

(iii.) *If $f, g : \mathbb{R}^d \to \mathbb{R}^k$ are Borel functions and $\|f(x) - g(x)\| \leq C \cdot 2^{-m}$ for all $x \in \mathbb{R}^d$, then $|H(f\mu, \mathcal{D}_m) - H(g\mu, \mathcal{D}_m)| \leq O_k(C)$.*

(iv.) *If $\nu(A) = \mu(A + x_0)$ for all $A \in \mathbb{R}^d$, then $|H(\mu, \mathcal{D}_m) - H(\nu, \mathcal{D}_m)| = O_d(1)$.*

(v.) *If $C^{-1} \leq m'/m \leq C$, then $|H(\mu, \mathcal{D}_m) - H(\mu, D_{m'})| \leq O_{C,d}(1)$.*

Proof. (i.) Let $\mu \in \mathcal{P}(K)$ and let $\varepsilon > 0$. Since K is bounded, we may define

$$M_1 = \max\{|\log_2 \mu(D)| \mid D \in \mathcal{D}_m\},$$
$$M_2 = \max\{\nu(D) \mid D \in \mathcal{D}_m\}.$$

Choose $\delta' > 0$ so that $|x \log_2 x| < \varepsilon/2|\mathcal{D}_m|$ whenever $0 \leq x < \delta'$ and, using again the boundedness of K, let $0 < \delta < \min\left\{\frac{1}{4|\mathcal{D}_m|M_1}\varepsilon,\ \delta'\right\}$ be such that

$$|\log_2 \nu(D) - \log_2 \mu(D)| \leq \frac{1}{4|\mathcal{D}_m|M_2}\varepsilon$$

whenever $\|\mu - \nu\| < \delta$ and when D has positive μ-measure. Let $\nu \in B(\mu, \delta)$. Denoting by $\mathcal{D}_n^0$ the set of atoms in the partition $\mathcal{D}_n$ that have μ measure 0, we have

$$|H(\nu, \mathcal{D}_m) - H(\mu, \mathcal{D}_m)| = |\sum_{D\in\mathcal{D}_m} (\nu(D)\log_2 \nu(D) - \mu(D)\log_2 \mu(D))\,|$$

$$\leq \sum_{D\in\mathcal{D}_m\setminus\mathcal{D}_m^0} |\nu(D)(\log_2 \nu(D) - \log_2 \mu(D)) + (\nu(D) - \mu(D))\log_2 \mu(D)|$$

$$+ \sum_{D\in\mathcal{D}_m^0} |\nu(D)\log_2 \nu(D)|$$

$$\leq \sum_{D\in\mathcal{D}_m\setminus\mathcal{D}_m^0} \left(\nu(D)\frac{1}{4|\mathcal{D}_m|M_2}\varepsilon + |\log_2 \mu(D)|\frac{1}{4|\mathcal{D}_m|M_1}\varepsilon\right) + \sum_{D\in\mathcal{D}_m^0} \frac{\varepsilon}{|\mathcal{D}_m|}$$

$$\leq \sum_{D\in\mathcal{D}_m\setminus\mathcal{D}_m^0} \frac{\varepsilon}{2|\mathcal{D}_m|} + \frac{\varepsilon}{2} \leq \varepsilon.$$

(ii.) Let $\mathcal{E}$ and $\mathcal{F}$ be partitions fulfilling the hypothesis. Assuming $H(\mu,\mathcal{E}) \geq H(\mu,\mathcal{F})$, by Lemma 1.4 (iv.) we have

$$\begin{aligned}|H(\mu,\mathcal{E}) - H(\mu,\mathcal{F})| &= H(\mu,\mathcal{E}) - H(\mu,\mathcal{F}) \leq H(\mu,\mathcal{E}\vee\mathcal{F}) - H(\mu,\mathcal{F}) \\ &= H(\mu,\mathcal{E}|\mathcal{F}) = \sum_{F\in\mathcal{F}} \mu(F)H(\mu_F,\mathcal{E})\end{aligned}$$

and since μ_F is supported on at most k atoms of $\mathcal{E}$, we have

$$\sum_{F\in\mathcal{F}} \mu(F)H(\mu_F,\mathcal{E}) \leq \sum_{F\in\mathcal{F}} \mu(F)\log_2 k = \log_2 k.$$

(iii.) Let $f, g : \mathbb{R}^d \to \mathbb{R}^k$ be such that $\|f(x) - g(x)\| \leq C2^{-m}$ for all $x \in \mathbb{R}^d$. Write $\mathcal{F} = \{f^{-1}(D) \mid D \in \mathcal{D}_m^k\}$ and $\mathcal{G} = \{g^{-1}(D) \mid D \in \mathcal{D}_m^k\}$. Note that these are partitions of $\mathbb{R}^d$. Now, for any $D \in \mathcal{D}_m^k$, there are $O_k(C^k)$ atoms of the same partition within distance $C2^{-m}$ from D. So, any $f^{-1}(D) \in \mathcal{F}$ can intersect at most $O_k(C^k)$ atoms of $\mathcal{G}$ (the preimages of the atoms closest to D in the partition $\mathcal{D}_m^k$) since f and g cannot map any point to atoms with distance greater than $C2^{-m}$. Hence by (ii.),

$$|H(f\mu,\mathcal{D}_m) - H(g\mu,\mathcal{D}_m)| = |H(\mu,\mathcal{F}) - H(\mu,\mathcal{G})| \leq \log_2 O_k(C^k) \leq O_k(C).$$

(iv.) Let ν be a translate of μ by x_0, i.e. $\nu(A) = \mu(A + x_0)$ for all A. We may assume that $\|x_0\| \leq O_d(1)2^{-m}$. Define the affine translation $f : \mathbb{R}^d \to \mathbb{R}^d$, $f(x) = x - x_0$ so that we have $\|f(x) - x\| \leq O_d(1)2^{-m}$ and $\nu = f\mu$. The statement holds by (iii.).

(v.) Assume that $C^{-1} \leq m'/m \leq C$. We may assume that $m' \geq m$ and write $m' = m + k(C)$ for some integer $k(C)$ depending on C. Since any atom of $\mathcal{D}_m$ then intersects $2^{-dm}/2^{-dm-dk(C)} = 2^{dk(C)}$ atoms of $\mathcal{D}_{m'}$, by (ii.) we have

$$|H(\mu, \mathcal{D}_m) - H(\mu, \mathcal{D}_{m'})| \leq d \cdot k(C) = O_{C,d}(1).$$

□

The following lemma shows that the normalized entropy $\mu \mapsto H_n(\mu)$ is uniformly continuous in $\mathcal{P}([0,1]^d)$ and n.

Lemma 1.6. *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,*

$$|H(\mu, \mathcal{A}) - H(\nu, \mathcal{A})| < \varepsilon \log_2 k + H((1-\varepsilon), \varepsilon).$$

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

$$|H_m(\mu) - H_m(\nu)| < d\varepsilon + \frac{H((1-\varepsilon), \varepsilon)}{m}.$$

Proof. Let $\varepsilon > 0$ be given, let $\delta < \varepsilon^2/18$ and let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ be such that $\|\mu - \nu\| < \delta$. Since both μ and ν are absolutely continuous with respect to $\mu + \nu$, we may define f and g to be the densities of μ and ν with respect to $\mu + \nu$. Define the measure τ by setting $\tau(A) = \int_A \min\{f, g\}\ d(\mu + \nu)$ for all Borel sets A. Clearly $\tau \leq \mu$ and $\tau \leq \nu$ everywhere.

Let A be the subset of $\mathbb{R}^d$ such that $f(x) - g(x) > \varepsilon/3$ whenever $x \in A$. Since

$$\mu(A) = \int_A f\ d(\mu + \nu) \geq \int_A g\ d(\mu + \nu) + \frac{\varepsilon}{3}(\mu + \nu)(A) = \nu(A) + \frac{\varepsilon}{3}(\mu + \nu)(A),$$

we see that $(\mu + \nu)(A) \leq 3\delta/\varepsilon < \varepsilon/6$. Similarly, if B is the set of points of $\mathbb{R}^d$ such that $g(x) - f(x) > \varepsilon/3$ for every $x \in B$, we find that $(\mu + \nu)(B) \leq \varepsilon/6$. Set $R = \mathbb{R}^d \setminus (A \cup B)$ and observe that

$$\mu(R) = 1 - \mu(A \cup B) \geq 1 - (\mu + \nu)(A) - (\mu + \nu)(B) \geq 1 - \varepsilon/3$$

and $(\mu + \nu)(R) \leq 2$. Therefore,

$$\begin{aligned} \tau(\mathbb{R}^d) &= \int_R \min\{f, g\}\ d(\mu + \nu) + \int_{A \cup B} \min\{f, g\}\ d(\mu + \nu) \\ &\geq \int_R (f - \varepsilon/3)\ d(\mu + \nu) \\ &= \mu(R) - \varepsilon(\mu + \nu)(R)/3 \\ &\geq 1 - \varepsilon. \end{aligned}$$

Define $\tilde{\tau} = \dfrac{1}{\tau(\mathbb{R}^d)}\tau$ and note that $\tilde{\tau}$ is a probability measure that satisfies

$$(1-\varepsilon)\tilde{\tau} \le \tau \le \min\{\mu, \nu\}.$$

Hence we can define probability measures

$$\begin{aligned} \mu' &= \frac{1}{\varepsilon}\left(\mu - (1-\varepsilon)\tilde{\tau}\right) \\ \nu' &= \frac{1}{\varepsilon}\left(\nu - (1-\varepsilon)\tilde{\tau}\right) \end{aligned}$$

so that we have

$$\begin{aligned} \mu &= (1-\varepsilon)\tilde{\tau} + \varepsilon\mu', \\ \nu &= (1-\varepsilon)\tilde{\tau} + \varepsilon\nu'. \end{aligned}$$

By Lemma 1.4 (v.) and (vi.), assuming $H(\mu, \mathcal{A}) \ge H(\nu, \mathcal{A})$, we have

$$\begin{aligned} &|H(\mu, \mathcal{A}) - H(\nu, \mathcal{A})| = H(\mu, \mathcal{A}) - H(\nu, \mathcal{A}) \\ \le&(1-\varepsilon)H(\tilde{\tau}, \mathcal{A}) + \varepsilon H(\mu', \mathcal{A}) + H((1-\varepsilon), \varepsilon) - (1-\varepsilon)H(\tilde{\tau}, \mathcal{A}) \\ \le&\varepsilon \log_2 k + H((1-\varepsilon), \varepsilon). \end{aligned}$$

If $\mu, \nu \in \mathcal{P}([0,1]^d)$, replacing $\mathcal{A}$ with $\mathcal{D}_m$, using the approximation (1.2) in the place of $\log_2 k$ and dividing by m yields the second inequality.

□

If μ is a probability measure on the unit interval, by (1.2) and (ii.) of Lemma 1.4, $H_n(\mu) \le 1$ with equality if and only if μ is uniform. Also, if μ is atomic on $[0,1]$, $H_n(\mu)$ tends to 0 as n tends to infinity; we may restrict μ on a finite set with mass arbitrarily close to 1 by writing its support as a countable, increasing union of finite sets,

$$\operatorname{supp}(\mu) = \bigcup_{n\in\mathbb{N}} A_n, \;\; A_n = \{x \mid \mu(\{x\}) > 1/n\}.$$

For any δ and large enough N, the normalized restriction of μ on A_N is then a probability measure on the unit interval with $\|\mu - \mu_{A_N}\| < \delta$. By the above lemma, $|H_n(\mu) - H_n(\mu_{A_N})| < \varepsilon$ for all n, and since $H_n(\mu_{A_N}) \le \log_2 |A_N|/n$ which tends to 0 as n grows, we have $H_n(\mu) < \varepsilon$ for all large enough n.

Our main motivation throughout the section is to find out what is required from two measures $\mu, \nu \in \mathcal{P}([0,1])$ for their convolution not to have substantially greater randomness when compared to that of μ or ν; in terms of entropy, we want to find out when their (normalized) entropies are roughly the same, or when

$$|H_n(\mu * \nu) - H_n(\mu)| < \delta$$

holds with a small δ and large n. The direction $H_n(\mu * \nu) > H_n(\mu) - \delta$ holds trivially for any measures when n is sufficiently large; see Lemma 1.25. For the more interesting inequality

$$H_n(\mu * \nu) < H_n(\mu) + \delta, \tag{1.3}$$

if either μ is uniform or ν is atomic, the inequality turns out to hold trivially. To see this, observe that if $\mu = u$, the upper bound of (1.3) is just the trivial bound of normalized entropy. On the other hand, if $\nu = \sum_{i=1}^{\infty} a_i \delta_{x_i}$ is atomic and A_k are the sets defined above, for every k there is a finite set $J_k \subset \mathbb{N}$ such that $\nu_{A_k} = \sum_{i \in J_k} a_i \delta_{x_i}$. Therefore, by Lemma 1.6, Lemma 1.5 (iv.) and the convexity bound of Lemma 1.4,

$$\begin{aligned}
H_n(\mu * \nu) &= H_n\left(\lim_{k\to\infty} \sum_{i \in J_k} a_i(\mu * \delta_{x_i})\right) \\
&\leq \lim_{k\to\infty} \sum_{i \in J_k} a_i H_n(\mu * \delta_i) - \frac{1}{n} \lim_{k\to\infty} \sum_{i \in J_k} a_i \log_2 a_i \\
&\leq H_n(\mu) + O(1/n) + H_n(\nu) + \varepsilon \\
&< H_n(\mu) + \delta
\end{aligned}$$

when ε is small and n is large enough, since $H_n(\nu)$ tends to 0. A more interesting result is, however, that these two trivial conditions turn out to be the only possible ones, in a local and statistical sense. We will see this in Theorem 1.26, given originally by Hochman in [7].

1.1 Global and local properties of a measure

As we saw when introducing the component measures, taking the expected value over all raw components of a fixed scale yields the original measure. It turns out we can also approximate the entropy of a measure using the entropies of its components.

We make an observation regarding the connection between the entropies of raw and rescaled components. Let μ be a probability measure on $\mathbb{R}^d$ and fix a point x. Observe that since the homothety $T_{\mathcal{D}_n(x)}$ maps $\mathcal{D}_n(x)$ into $[0,1)^d$, the preimages of the atoms of the level-m dyadic partition of $[0,1)^d$ are the atoms of the level-$(n+m)$ dyadic partition of $\mathcal{D}_n(x)$. Hence

$$\begin{aligned}
H(\mu^{x,n}, \mathcal{D}_m) &= -\sum_{D\in\mathcal{D}_m} \frac{1}{\mu(\mathcal{D}_n(x))}\mu(T^{-1}_{D_n(x)}(D))\log_2\left(\frac{1}{\mu(\mathcal{D}_n(x))}\mu(T^{-1}_{D_n(x)}(D))\right) \\
&= -\sum_{D\in\mathcal{D}_{n+m}} \frac{1}{\mu(\mathcal{D}_n(x))}\mu\left(D\cap\mathcal{D}_n(x)\right)\log_2\left(\frac{1}{\mu(\mathcal{D}_n(x))}\mu\left(D\cap\mathcal{D}_n(x)\right)\right) \\
&= H(\mu_{x,n}, \mathcal{D}_{n+m}). \qquad (1.4)
\end{aligned}$$

The next lemma is a very useful tool in moving between the scales of a measure.

Lemma 1.7. *For any $r\geq 1$, $\mu\in\mathcal{P}([-r,r]^d)$ and all integers $m<n$,*

$$H_n(\mu) = \mathbb{E}_{0\leq i<n}\left(H_m(\mu^{x,i})\right) + O_d\left(\frac{m}{n} + \frac{\log_2(2r)}{n}\right).$$

Proof ([7], p. 17). We begin by inspecting $\mathbb{E}_{0\leq i<n}\left(H_m(\mu^{x,i})\right)$. For every n, we may rewrite the expected value using conditional entropy,

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

Hence, by the law of total expectation, we have

$$\mathbb{E}_{0\leq i<n}\left(H_m(\mu^{x,i})\right) = \frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{m}H(\mu, \mathcal{D}_{i+m}|\mathcal{D}_i). \qquad (1.5)$$

By Lemma 1.5 (ii.),

$$\begin{aligned}
H(\mu, \mathcal{D}_{i+m}|\mathcal{D}_i) &= H(\mu, \mathcal{D}_{i+m}\vee\mathcal{D}_i) - H(\mu, \mathcal{D}_i) = H(\mu, \mathcal{D}_{i+m}) - H(\mu, \mathcal{D}_i) \\
&= O(\log_2 2^{dm}) = O_d(m) \qquad (1.6)
\end{aligned}$$

for every i. Therefore, by introducing the error term $O_d(m/n)$ we may delete up to m terms from the sum (1.5). This allows us to assume, without loss of generality, that $n/m\in\mathbb{N}$; if k is such that $km\leq n$ and $(k+1)m>n$, we forget the last $n - km < m$ terms in the sum. When $m = 1$, using Lemma 1.4 (iv.) and the fact that $H(\mu, \mathcal{D}_0) \leq \log_2(2r)^d$ we have

$$\begin{aligned}
\sum_{i=0}^{n-1} H(\mu, \mathcal{D}_{i+1} | \mathcal{D}_i) &= \sum_{i=0}^{n-1} \left(H(\mu, \mathcal{D}_{i+1} \vee \mathcal{D}_i) - H(\mu, \mathcal{D}_i) \right) \\
&= \sum_{i=0}^{n-1} \left(H(\mu, \mathcal{D}_{i+1}) - H(\mu, \mathcal{D}_i) \right) \\
&= H(\mu, \mathcal{D}_n) - H(\mu, \mathcal{D}_0) \\
&= H(\mu, \mathcal{D}_n | \mathcal{D}_0) \\
&= H(\mu, \mathcal{D}_n) - O_d(\log_2 2r).
\end{aligned}$$

For a general $m \in \mathbb{N}$, using the equality above and the fact that m divides n, we decompose the sum and obtain

$$\begin{aligned}
\sum_{i=0}^{n-1} \frac{1}{m} H(\mu, \mathcal{D}_{i+m} | \mathcal{D}_i) &= \frac{1}{m} \sum_{p=0}^{m-1} \left(\sum_{k=0}^{n/m-1} H(\mu, \mathcal{D}_{(k+1)m+p} | \mathcal{D}_{km+p}) \right) \\
&= \frac{1}{m} \sum_{p=0}^{m-1} \left(\sum_{k=0}^{n/m-1} H(\mu, \mathcal{D}_{(k+1)m+p}) - H(\mu, \mathcal{D}_{km+p}) \right) \\
&= \frac{1}{m} \sum_{p=0}^{m-1} H(\mu, \mathcal{D}_{n+p} | \mathcal{D}_p). \qquad (1.7)
\end{aligned}$$

We note that

$$\begin{aligned}
H(\mu, \mathcal{D}_n) + H(\mu, \mathcal{D}_{n+p} | \mathcal{D}_n) &= H(\mu, \mathcal{D}_{n+p} \vee \mathcal{D}_n) \\
&= H(\mu, \mathcal{D}_{n+p}) \\
&= H(\mu, \mathcal{D}_p) + H(\mu, \mathcal{D}_{n+p} | \mathcal{D}_p). \qquad (1.8)
\end{aligned}$$

By Lemma 1.5, we have the bound $H(\mu, \mathcal{D}_p) \leq \log_2 \left(2r \cdot 2^p\right)^d = dp + d\log_2(2r)$ and by (1.6), $H(\mu, \mathcal{D}_{rm+p} | \mathcal{D}_{rm}) = O_d(p)$. Using the equality (1.8) and the fact that $0 \leq p < m$, we get

$$\begin{aligned}
\left| \frac{1}{n} H(\mu, \mathcal{D}_{n+p} | \mathcal{D}_p) - H_n(\mu) \right| &= \left| \frac{1}{n} H(\mu, \mathcal{D}_{n+p} | \mathcal{D}_n) - \frac{1}{n} H(\mu, \mathcal{D}_p) \right| \\
&\leq \frac{1}{n} H(\mu, \mathcal{D}_{n+p} | \mathcal{D}_n) + \frac{1}{n} H(\mu, \mathcal{D}_p) \\
&\leq O_d \left(\frac{m + \log_2(2r)}{n} \right).
\end{aligned}$$

for all p. Summing the above inequality m times as $p = 0, \dots, m-1$ yields

$$m \cdot H_n(\mu) \le \sum_{p=0}^{m-1} \frac{1}{n} H(\mu, \mathcal{D}_{n+p}|\mathcal{D}_p) + m \cdot O_d\left(\frac{m}{n} + \frac{\log_2(2r)}{n}\right).$$

Dividing by m gives us the sum from (1.7). Using the equality (1.5) then finishes the proof. $\square$

We can derive a similar approximation for the entropy of the convolution. In this case, we only receive a lower bound for the entropy of μ. This is due to the fact that since the components of the convolution are not equal to the convolution of two components, we have only concavity to use in obtaining the connection between the average small-scale entropy and the entropy of the convolution measure.

Lemma 1.8. *Let $r > 0$ and $\mu, \nu \in \mathcal{P}([-r, r]^d)$. Then for $m < n \in \mathbb{N}$,*

$$\begin{aligned} H_n(\mu * \nu) &\ge \mathbb{E}_{0 \le i \le n}\left(\frac{1}{m} H(\mu_{x,i} * \nu_{y,i}, \mathcal{D}_{i+m}|\mathcal{D}_i)\right) + O_d\left(\frac{m + \log_2(4r)}{n}\right) \\ &\ge \mathbb{E}_{0 \le i \le n}\left(H_m(\mu^{x,i} * \nu^{y,i})\right) + O_d\left(\frac{1}{m} + \frac{m}{n} + \frac{\log_2(4r)}{m}\right). \end{aligned}$$

Proof ([7], p. 18). Note that for all $p = 0, \dots, m-1$, using the chain rule for conditional entropy, we have

$$\begin{aligned} H(\mu * \nu, \mathcal{D}_{n+p}) &= H(\mu * \nu, \mathcal{D}_n) + H(\mu * \nu, \mathcal{D}_{n+p}|\mathcal{D}_n) \\ &= H(\mu * \nu, \mathcal{D}_n) + \sum_{D \in \mathcal{D}_n} (\mu * \nu)(D) H((\mu * \nu)_D, \mathcal{D}_{n+p}) \\ &\le H(\mu * \nu, \mathcal{D}_n) + \log_2 2^{dp} \\ &= H(\mu * \nu, \mathcal{D}_n) + dp. \end{aligned}$$

Since $\mu * \nu$ has its support contained in $[-2r, 2r]^d$ and $H(\mu * \nu, \mathcal{D}_p) \le d \log_2(4r2^p) \le dm + d\log_2(4r)$, using the above we can write

$$\begin{aligned} H_n(\mu * \nu) &\ge \frac{1}{n} H(\mu * \nu, \mathcal{D}_{n+p}) - \frac{dp}{n} \\ &\ge \frac{1}{n} H(\mu * \nu, \mathcal{D}_{n+p}|\mathcal{D}_p) + \frac{1}{n} H(\mu * \nu, \mathcal{D}_p) - \frac{dm}{n} \\ &= \frac{1}{n} H(\mu * \nu, \mathcal{D}_{n+p}|\mathcal{D}_p) + O_d\left(\frac{m}{n} + \frac{\log_2(4r)}{n}\right). \end{aligned}$$

Again by the chain rule, we have

$$H(\mu * \nu, \mathcal{D}_{n+p}|\mathcal{D}_p) = \sum_{k=0}^{\lfloor n/m \rfloor -1} H(\mu * \nu, \mathcal{D}_{p+(k+1)m}|\mathcal{D}_{p+km}) + H(\mu * \nu, \mathcal{D}_{n+p}|\mathcal{D}_{p+\lfloor n/m \rfloor m})$$

and since $H(\mu * \nu, \mathcal{D}_{n+p}|\mathcal{D}_{p+\lfloor n/m \rfloor m}) \leq O_d(m)$, by adding to the error term $O_d(m/n)$ we may again assume that m divides n.

The convolution of random components $\mu_{x,i} * \nu_{y,i}$ is a random variable $\mathbb{R}^d \times \mathbb{R}^d \to \mathcal{P}(\mathbb{R}^d)$ distributed according to $\mu \times \nu$ and admits a constant value at each $D \times E \in \mathcal{D}_n \times \mathcal{D}_n$. Hence for any k, by bilinearity of convolution,

$$\begin{aligned}
\mathbb{E}_{i=k}(\mu_{x,i} * \nu_{y,i}) &= \int (\mu_{x,k} * \nu_{y,k}) \; d(\mu \times \nu) \\
&= \sum_{D,E \in \mathcal{D}_k} (\mu_D * \nu_E) \mu(D) \nu(E) \\
&= \sum_{D,E \in \mathcal{D}_k} \mu_{|D} * \nu_{|E} = \mu * \nu.
\end{aligned}$$

Using the concavity of conditional entropy given by Lemma 1.4, we get

$$\begin{aligned}
H(\mu * \nu, \mathcal{D}_{n+p}|\mathcal{D}_p) &= \sum_{k=0}^{n/m-1} H(\mu * \nu, \mathcal{D}_{p+(k+1)m}|\mathcal{D}_{p+km}) + O_d(m/n) \\
&= \sum_{k=0}^{n/m-1} H\left(\mathbb{E}_{i=p+km}(\mu_{x,i} * \nu_{y,i}), \mathcal{D}_{p+(k+1)m}|\mathcal{D}_{p+km}\right) + O_d(m/n) \\
&\geq \sum_{k=0}^{n/m-1} \mathbb{E}_{i=p+km}\left(H(\mu_{x,i} * \nu_{y,i}, \mathcal{D}_{i+m}|\mathcal{D}_i)\right) + O_d(m/n).
\end{aligned}$$

Combining this with the discussion above, we obtain

$$\begin{aligned}
H_n(\mu * \nu) &= \frac{1}{m}\sum_{p=0}^{m-1}\frac{1}{n}H(\mu*\nu,\mathcal{D}_n) \\
&= \frac{1}{m}\sum_{p=0}^{m-1}\left(\frac{1}{n}H(\mu*\nu,\mathcal{D}_{n+p}|\mathcal{D}_p) + O_d\left(\frac{m}{n}+\frac{\log_2(4r)}{n}\right)\right) \\
&\geq \frac{1}{mn}\sum_{p=0}^{m-1}\left(\sum_{k=0}^{n/m-1}\mathbb{E}_{i=p+km}\left(H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_{i+m}|\mathcal{D}_i)\right)\right) + O_d\left(\frac{m}{n}+\frac{\log_2(4r)}{n}\right) \\
&= \sum_{j=0}^{n-1}\frac{1}{n}\mathbb{E}_{i=j}\left(\frac{1}{m}H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_{i+m}|\mathcal{D}_i)\right) + O_d\left(\frac{m}{n}+\frac{\log_2(4r)}{n}\right) \\
&= \mathbb{E}_{0\leq i<n}\left(\frac{1}{m}H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_{i+m}|\mathcal{D}_i)\right) + O_d\left(\frac{m}{n}+\frac{\log_2(4r)}{n}\right),
\end{aligned}$$

which proves the first inequality. To derive the second one, we need an identity similar to (1.4). If $\pi : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is the addition map, if $T_{\mathcal{D}_n(x,y)}$ is the homothety of $\mathbb{R}^{2d}$ mapping $\mathcal{D}_n(x,y)$ to $[0,1)^{2d}$, $(t_1,t_2) \in \mathbb{Z}^d \times \mathbb{Z}^d$ is the vector by which it translates and if $D \in \mathcal{D}_n$,

$$\begin{aligned}
T^{-1}_{\mathcal{D}_i(x,y)}(\pi^{-1}(D)) &= \{T^{-1}_{\mathcal{D}_i(x,y)}(a,b) \mid a+b \in D \in \mathcal{D}_n\} \\
&= \{(u,v) \mid T_{\mathcal{D}_i(x)}(u) + T_{\mathcal{D}_i(y)}(v) \in D\} \\
&= \{(u,v) \mid 2^i(u+v)+t_1+t_2 \in D\} \\
&= \{(u,v) \mid u+v \in 2^{-i}(D-t_1-t_2)\} \\
&= \pi^{-1}(2^{-i}(D-t_1-t_2)).
\end{aligned}$$

Note that $2^{-i}(D-t_1-t_2)$ is an atom of $\mathcal{D}_{n+i}$; denote it by $\tilde{D}$. We have

$$\begin{aligned}
(\mu^{x,i} * \nu^{y,i})(D) &= (\mu\times\nu)^{(x,y),i}(\pi^{-1}(D)) \\
&= \frac{1}{(\mu\times\nu)(\mathcal{D}_i(x,y))}(\mu\times\nu)(\mathcal{D}_i(x,y)\cap T^{-1}_{\mathcal{D}_i(x,y)}(\pi^{-1}(D))) \\
&= \frac{1}{(\mu\times\nu)(\mathcal{D}_i(x,y))}(\mu\times\nu)(\mathcal{D}_i(x,y)\cap \pi^{-1}(\tilde{D})) \\
&= (\mu_{x,i} * \nu_{y,i})(\tilde{D})
\end{aligned}$$

and therefore

$$\begin{aligned}
H(\mu^{x,i} * \nu^{y,i}, \mathcal{D}_n) &= -\sum_{D \in \mathcal{D}_n} (\mu^{x,i} * \nu^{y,i})(D) \log_2 (\mu^{x,i} * \nu^{y,i})(D) \\
&= -\sum_{\tilde{D} \in \mathcal{D}_{n+i}} (\mu_{x,i} * \nu_{y,i})(\tilde{D}) \log_2 (\mu_{x,i} * \nu_{y,i})(\tilde{D}) \\
&= H(\mu_{x,i} * \nu_{y,i}, \mathcal{D}_{n+i}).
\end{aligned}$$

With this identity, we calculate

$$\begin{aligned}
&\mathbb{E}_{i=j} \left(H(\mu_{x,i} * \nu_{y,i}, \mathcal{D}_{i+m} | \mathcal{D}_i) \right) \\
&= \mathbb{E}_{i=j} \left(H(\mu_{x,i} * \nu_{y,i}, \mathcal{D}_{i+m}) - H(\mu_{x,i} * \nu_{y,i}, \mathcal{D}_i) \right) \\
&= \mathbb{E}_{i=j} \big(H(\mu^{x,i} * \nu^{y,i}, \mathcal{D}_m) - H(\mu^{x,i} * \nu^{y,i}, \mathcal{D}_0) \big) \\
&= \mathbb{E}_{i=j} \big(H(\mu^{x,i} * \nu^{y,i}, \mathcal{D}_m) \big) + O_d(1),
\end{aligned}$$

since $\mu^{x,i} * \nu^{y,i}$ is supported on $[0,2)^d$, an area that intersects 2^{d+1} atoms of $\mathcal{D}_0$. Applying this to the first inequality in the statement of the lemma, we get

$$\begin{aligned}
&\mathbb{E}_{0 \le i < n} \left(\frac{1}{m} H(\mu_{x,i} * \nu_{y,i}, \mathcal{D}_{i+m} | \mathcal{D}_i) \right) + O_d \left(\frac{m}{n} + \frac{\log_2(4r)}{n} \right) \\
=&\mathbb{E}_{0 \le i < n} \left(\frac{1}{m} H(\mu^{x,i} * \nu^{y,i}, \mathcal{D}_m) \right) + O_d \left(\frac{1}{m} \right) + O_d \left(\frac{m}{n} + \frac{\log_2(4r)}{n} \right) \\
=&\mathbb{E}_{0 \le i < n} \big(H_m(\mu^{x,i} * \nu^{y,i}) \big) + O_d \left(\frac{1}{m} + \frac{m}{n} + \frac{\log_2(4r)}{n} \right)
\end{aligned}$$

which is what we wanted to show. □

In our earlier discussion, we saw that the approximate equality (1.3) holds trivially when either μ is uniform or ν is atomic. We stated that these conditions turn out to be the only possible ones in a local and statistical sense; more precisely, it means that either μ has to be almost uniform in its small-scale components or the components of ν have to be close to atomic. We now define what it means for a measure to be "almost uniform" or "close to atomic".

Definition 1.9. A probability measure $\mu \in \mathcal{P}([0,1])$ is (ε, m)-*atomic* if $H_m(\mu) < \varepsilon$.

Recall that normalized scale-n entropy of an atomic measure approaches 0 as n tends to infinity; hence the definition for approximate atomicity is quite intuitive. In a similar manner, we define almost-uniformity by comparing the normalized entropy of a measure to that of a uniform measure.

Definition 1.10. A probability measure $\mu \in \mathcal{P}([0,1])$ is (ε, m)*-uniform* if $H_m(\mu) > 1 - \varepsilon$.

Almost-atomicity and almost-uniformity are properties that pass to component measures, given a small enough scale.

Lemma 1.11. *If $\mu \in \mathcal{P}([0,1])$ is (ε, m)-atomic, then for $k < m$,*

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

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

Proof. By Lemma 1.7,

$$\mathbb{E}_{0 \leq i < m}\left(H_k(\mu^{x,i})\right) = H_m(\mu) + O\left(\frac{k}{m} + \frac{\log_2(1)}{m}\right) < \varepsilon + O\left(\frac{k}{m}\right).$$

By Markov's inequality, we have

$$\begin{aligned}
\mathbb{P}_{0 \leq i < m}\left(\mu^{x,i} \text{ is } (\varepsilon', k)\text{-atomic}\right) &= \mathbb{P}_{0 \leq i < m}\left(H_k(\mu^{x,i}) < \varepsilon'\right) \\
&= 1 - \mathbb{P}_{0 \leq i < m}\left(H_k(\mu^{x,i}) \geq \varepsilon'\right) \\
&\geq 1 - \frac{\mathbb{E}_{0 \leq i < m}\left(H_k(\mu^{x,i})\right)}{\varepsilon'} \\
&> 1 - \frac{\varepsilon + O\left(k/m\right)}{\varepsilon'} \\
&= 1 - \varepsilon',
\end{aligned}$$

when $\varepsilon' = \sqrt{\varepsilon + O\left(\frac{k}{m}\right)}$. □

Lemma 1.12. *If $\mu \in \mathcal{P}([0,1])$ is (ε, n)-uniform, then for every $1 \leq m < n$,*

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

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

Proof. As in the previous proof, Lemma 1.7 gives us

$$\mathbb{E}_{0 \leq i < n}\left(H_m(\mu^{x,i})\right) = H_n(\mu) - O\left(\frac{m}{n} + \frac{\log_2(1)}{n}\right) > 1 - \varepsilon - O\left(\frac{m}{n}\right).$$

Since normalized entropy on the line is never greater than 1 for a measure on $[0,1]$, Markov's inequality gives us

$$\begin{aligned}
\mathbb{P}_{0\le i<n}\left(\mu^{x,i} \text{ is } (\varepsilon', m)\text{-uniform}\right) &= \mathbb{P}_{0\le i<n}\left(H_m(\mu^{x,i}) > 1-\varepsilon'\right) \\
&= 1 - \mathbb{P}_{0\le i<n}\left(1 - H_m(\mu^{x,i}) \ge \varepsilon'\right) \\
&\ge 1 - \frac{1 - \mathbb{E}_{0\le i<n}\left(H_m(\mu^{x,i})\right)}{\varepsilon'} \\
&> 1 - \frac{\varepsilon + O(m/n)}{\varepsilon'} \\
&= 1 - \varepsilon',
\end{aligned}$$

when $\varepsilon' = \sqrt{\varepsilon + O\left(\frac{m}{n}\right)}$. □

The following result is also a consequence of Markov's inequality and will prove useful in some occasions.

Lemma 1.13. *Suppose that $\mathcal{A} \subset \mathcal{P}([0,1])$ and that*

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

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

$$\mathbb{P}_{i=q}\left(\mu^{x,i} \in \mathcal{A}\right) > 1 - \sqrt{\varepsilon}$$

for all $q \in I$.

Proof. Define a function $f : \{0,\dots,n-1\} \to [0,1]$, by setting $f(q) = \mathbb{P}_{i=q}(\mu^{x,i} \in \mathcal{A})$ for all q. Then

$$\mathbb{E}_{0\le q<n}(f(q)) = \frac{1}{n}\sum_{q=0}^{n-1} \mathbb{P}_{i=q}(\mu^{x,i} \in \mathcal{A}) = \mathbb{P}_{0\le i<n}(\mu^{x,i} \in \mathcal{A}) > 1 - \varepsilon.$$

By Markov's inequality,

$$\begin{aligned}
\mathbb{P}_{0\le q<n}\left(f(q) > 1 - \sqrt{\varepsilon}\right) &= 1 - \mathbb{P}_{0\le q<n}\left(1 - f(q) \ge \sqrt{\varepsilon}\right) \\
&\ge 1 - \frac{1 - \mathbb{E}_{0\le q<n}(f(q))}{\sqrt{\varepsilon}} \\
&> \frac{\sqrt{\varepsilon} - 1 + 1 - \varepsilon}{\sqrt{\varepsilon}} = 1 - \sqrt{\varepsilon}.
\end{aligned}$$

Let I be the subset of $\{0,\dots,n-1\}$ such that $f(q) = \mathbb{P}_{i=q}\left(\mu^{x,i} \in \mathcal{A}\right) > 1 - \sqrt{\varepsilon}$ for every $q \in I$. Since the probability above depends only on the uniform selection of $0 \le q < n$, the inequality states that $|I| > (1-\sqrt{\varepsilon})n$. □

1.2 Covering lemmas

We collect here some basic covering lemmas that we require later on, when we need to classify scales of a measure based on the average behaviour of its components on that scale.

Lemma 1.14. *Let $I \subset \{0, \ldots, n\}$ and $m \in \mathbb{N}$ be given. Then there is a subset I' of I such that*

$$I \subset I' + [0, m] = \bigcup_{i \in I'} [i, i+m]$$

and $[i, i+m] \cap [j, j+m] = \emptyset$ for distinct $i, j \in I'$.

Proof. We define I' inductively. If I is empty, $I' = \emptyset$ is the subset of the lemma. Assuming I is nonempty, first set $I' = \{\min I\}$. Then, if the set $I \setminus \bigcup_{i \in I'} [i, i+m]$ is nonempty, add the smallest of its elements to I'. Do this until $I \subset I' + [0, m]$; because I is bounded, this occurs at some point. $\square$

Lemma 1.15. *Let $I, J \subset \{0, \ldots, n\}$ and $m \in \mathbb{N}, \delta > 0$. Suppose that*

$$|[i, i+m] \cap J| \geq \left(1 - \frac{\delta}{2}\right) |[i, i+m]| = \left(1 - \frac{\delta}{2}\right)(m+1)$$

for all $i \in I$. Then there is a subset $J' \subset J$ such that

$$|J' \cap (J' - l)| \geq \left(1 - \delta - \frac{l}{m+1}\right) |I|$$

for all $0 \leq l \leq m$.

Proof. Let $I' \subset I$ be the subset given by Lemma 1.14. Define

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

Then

$$\begin{aligned} J' \cap (J' - l) &= \left(\bigcup_{i \in I'} (J \cap [i, i+m]) \right) \cap \left(\bigcup_{i \in I'} (J \cap [i, i+m]) - l \right) \\ &\supset \bigcup_{i \in I'} \left((J \cap [i, i+m]) \cap ((J \cap [i, i+m]) - l) \right). \end{aligned}$$

Note that since

$$|J \cap [i, i+m]| \geq (1 - \delta/2) |[i, i+m]|$$

and

$$|\,((J \cap [i,i+m]) - l) \cap [i,i+m]| \geq \left(1 - \frac{\delta}{2} - \frac{l}{|[i,i+m]|}\right)|[i,i+m]|,$$

the intersection of $J \cap [i,i+m]$ and $(J \cap [i,i+m]) - l$ differs from the interval $[i,i+m]$ by at most

$$\left(\frac{\delta}{2} + \frac{\delta}{2} - \frac{l}{m+1}\right)|[i,i+m]|$$

elements for all $i \in I'$. Since $I \subset U_{i \in I'}[i,i+m]$,

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

□

Lemma 1.16. *Let m, δ be given, and let I_1, J_1 and I_2, J_2 be two pairs of subsets of $\{0, \ldots, n\}$ satisfying the assumptions of Lemma 1.15. Suppose also that $I_1 \cap I_2 = \emptyset$. Then there exist $J_1' \subset J_1$ and $J_2' \subset J_2$ with $J_1' \cap J_2' = \emptyset$ and $|J_1' \cup J_2'| \geq (1-\delta)^2|I_1 \cup I_2|$.*

Proof. Let $I_1' \subset I_1$ be the subset given by Lemma 1.14 and

$$J_1' = J_1 \cap \left(\bigcup_{i \in I_1'}[i,i+m]\right).$$

By Lemma 1.15, choosing $l = 0$, we have $|J_1'| \geq (1-\delta)|I_1|$. Let

$$U = \bigcup_{i \in I_1'}[i,i+m].$$

As in the proof of Lemma 1.15, we see that $|J_1'| = |U \cap J_1| \geq (1-\delta)|U|$. Since $I_1 \subset U$ and $I_1 \cap I_2 = \emptyset$, we have

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

Using the fact that $|J_1'| \geq (1-\delta)|I_1|$, we get

$$\begin{aligned}
|J_1' \cup I_2| &= |J_1'| + |I_2| - |J_1' \cap I_2| \\
&\geq |J_1'| + |I_2| - \left(\frac{1}{1-\delta}|J_1'| - |I_1|\right) \\
&\geq |I_2| - \frac{\delta}{1-\delta}|J_1'| + |I_1| \\
&\geq (1-\delta)(|I_1| + |I_2|). \qquad (1.9)
\end{aligned}$$

Now replace I_1 with $I_2 \setminus J_1'$ and J_1 with J_2. We see that the pair $I_2 \setminus J_1', J_2$ satisfies the conditions of Lemma 1.15; $|[i, i+m] \cap J_2| \geq (1-\delta/2)(m+1)$ for all $i \in I_2 \setminus J_1' \subset I_2$. If $I_2' \subset I_2$ is the subset given by Lemma 1.14, define

$$\tilde{U} = \bigcup_{i \in I_2' \setminus J_1'} [i, i+m]$$

and set $J_2' = J_2 \cap \tilde{U}$. By Lemma 1.15, $|J_2'| \geq (1-\delta)|\tilde{U}| \geq (1-\delta)|I_2 \setminus J_1'|$. We approximate

$$|J_2' \cap J_1'| \leq |\tilde{U}| - |I_2 \setminus J_1'| \leq \frac{1}{1-\delta}|J_2'| - |I_2 \setminus J_1'|$$

and then, using $|J_2'| \geq (1-\delta)|I_2 \setminus J_1'|$, we get

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

Combining this with (1.9) gives $|J_1' \cup J_2'| \geq (1-\delta)^2|I_1 \cup I_2|$, so that it only remains to show that J_1' and J_2' are disjoint. However, if this is not the case, we can replace J_1' with $J_1' \setminus J_2'$ without affecting the lower bound derived above. □

1.3 Local properties of the convolution measure

We now begin investigating the small-scale behaviour of a convolution measure as the number of measures in the convolution increases. The central limit theorem tells us that if we have a large number of probability measures, their convolution may be rescaled so that it approaches the Gaussian measure of same mean and variance. Since the Gaussian has a continuous density function, with small enough scale we may observe it to be near uniform. As a consequence, we state that the components of a convolution of multiple probability measures are typically almost uniform.

Proposition 1.17. *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 $Var(\mu) \geq \sigma k$. Then*

$$\mathbb{P}_{i=p-\lfloor \log_2 \sqrt{k} \rfloor}\left(\mu^{x,i} \textit{ is } (\delta, m)\textit{-uniform}\right) > 1-\delta.$$

Note that $p - \lfloor \log \sqrt{k} \rfloor$ can and generally will be negative; hence the conclusion of the proposition primarily concerns the components supported on intervals of length $O_p(\sqrt{k}) = O_{\sigma,\delta,m}(\sqrt{k})$ (note that μ is a probability measure on the interval $[0,k]$).

Proof of Proposition 1.17 ([7], p. 22). Let γ be a probability measure on $\mathbb{R}$ with continuous density function f, that is, $\gamma(I) = \int_I f(x)\,dx$ for all $I \subset \mathbb{R}$. Let I be an interval and assume $x \in I$ is such that $f(x) \neq 0$. Applying the mean value theorem to the function $y \mapsto \int_{-\infty}^{y} f(t)\,dt$ yields

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

Since f is continuous, the right-hand side tends to 0 as the endpoints of I approach x. Hence, for any $D \in \mathcal{D}_n$ with positive γ-mass,

$$\left| \frac{\gamma(D)}{\mathrm{diam}(D)} - f(x) \right| = o(1)$$

for some $x \in D \in \mathcal{D}_n$ with $f(x) \neq 0$, as $n \to \infty$. For any given m and assuming n is large enough,

$$\frac{\gamma(D)}{\gamma(E)} = \frac{\gamma(D)/\mathrm{diam}(D)}{\gamma(E)/\mathrm{diam}(E)} \approx \frac{f(x)}{f(y)} \approx 1,$$

when $D, E \in \mathcal{D}_{n+m}$ have positive mass and $D, E \subset D_n(x)$. In other words, for every x the distribution of $\gamma^{x,n}$ on the level-m dyadic subintervals of $[0,1)$ approaches the uniform distribution as n tends to infinity. Moreover, the rate of this convergence depends only on the modulus of continuity of the density function f. Let now $\mathcal{G}$ denote the family of all Gaussian measures restricted to the interval $I = [-R, R]$, with mean $\langle \gamma \rangle = 0$ and variance $\mathrm{Var}(\gamma) \in [\sigma, 4]$. For any $\varepsilon > 0$, by choosing R large enough we may assume that $\gamma(I) > 1 - \varepsilon$ for every $\gamma \in \mathcal{G}$ since now the mass of I depends only on $\mathrm{Var}(\gamma)$. We point out that the family of all densities of measures in $\mathcal{G}$ is an equicontinuous family; recall that these functions are of the form $x \mapsto \frac{1}{\sqrt{2\pi}\mathrm{Var}(\gamma)} e^{-\frac{x^2}{2\mathrm{Var}(\gamma)^2}}$, so the modulus of continuity of any function depends only on $\mathrm{Var}(\gamma)$. Restricting the variance on a compact interval with 0 excluded bounds the modulus uniformly.

Since the function $\mu \mapsto H(\mu, \mathcal{D}_m)$ is continuous on $\mathcal{P}([0,1])$, writing u for the uniform distribution on $[0,1)$ we conclude that $\lim_{n\to\infty} H_m(\gamma^{x,n}) = H_m(u) = 1$ for every Gaussian γ. Note that by equicontinuity this convergence is uniform in the set $\mathcal{G}$. Using Fatou's lemma, we obtain

$$\liminf_{p\to\infty} \mathbb{E}_{i=p}(H_m(\gamma^{x,i})) \geq \int \liminf_{n\to\infty} H_m(\gamma^{x,p})\,d\gamma = 1 - \varepsilon \tag{1.10}$$

for all $\gamma \in \mathcal{G}$. Let $p_0(\delta, m)$ be such that $\mathbb{E}_{i=p}(H_m(\gamma^{x,i})) > 1 - (\delta/2)^2$ for all $p \geq p_0$ and $\gamma \in \mathcal{G}$. By Markov's inequality, we have

$$\begin{aligned}
\mathbb{P}_{i=p}(H_m(\gamma^{x,i}) > 1 - \delta/2) &= 1 - \mathbb{P}_{i=p}(1 - H_m(\gamma^{x,i}) > \delta/2) \\
&\geq \frac{\delta/2 - 1 + \mathbb{E}_{i=p}(H_m(\gamma^{x,i}))}{\delta/2} \\
&\geq \frac{\delta/2 - 1 + 1 - (\delta/2)^2}{\delta/2} = 1 - \delta/2. \qquad (1.11)
\end{aligned}$$

Now, let $\mu_1, \ldots, \mu_k \in \mathcal{P}([0,1])$, set $\mu = \mu_1 * \cdots * \mu_k$ and fix numbers $\sigma, \delta \in (0,1)$. Define a mapping $F : [0,1] \to [0, 2^{-\lfloor \log_2 \sqrt{k} \rfloor}]$,

$$F(x) = 2^{-\lfloor \log_2 \sqrt{k} \rfloor} x$$

and set $\mu' = F\mu$; the measure μ' is μ scaled by a factor of $2^{-\lfloor \log_2 \sqrt{k} \rfloor}$. We want to apply the Berry-Esseen theorem (Theorem 0.21) to μ'. Since μ_i are probability measures on $[0,1]$, their third moments

$$\rho_i = \int_0^1 x^3 \; d\mu_i \leq \int_0^1 1 \; d\mu_i = 1$$

for all i, and by assumption, $\sigma k \leq \sum_{i=1}^k \mathrm{Var}(\mu_i) = \mathrm{Var}(\mu)$. Using these bounds, we receive the corresponding bounds for μ'; since $1/\sqrt{k} \leq 2^{-\lfloor \log_2 \sqrt{k} \rfloor} \leq 2/\sqrt{k}$, for the mean we have

$$\langle \mu' \rangle = \langle F\mu \rangle = \int x \; dF\mu = \int 2^{-\lfloor \log_2 \sqrt{k} \rfloor} x \; d\mu = 2^{-\lfloor \log_2 \sqrt{k} \rfloor} \langle \mu \rangle$$

and for the variance

$$\mathrm{Var}(\mu') = \int (x - \langle F\mu \rangle)^2 \; dF\mu = 2^{-2\lfloor \log_2 \sqrt{k} \rfloor} \mathrm{Var}(\mu) \in \left[\frac{1}{k} k\sigma, \frac{4}{k} k \right] = [\sigma, 4].$$

If $\pi : \mathbb{R}^k \to \mathbb{R}$ is the addition map, for all $A \subset \mathbb{R}$,

$$\begin{aligned}
\pi^{-1}(F^{-1}(A)) &= \left\{ (x_1, \ldots, x_k) \;\middle|\; \sum_{i=1}^k 2^{-\lfloor \log_2 \sqrt{k} \rfloor} x_i \in A \right\} \\
&= \left\{ 2^{\lfloor \log_2 \sqrt{k} \rfloor} (x_1, \ldots, x_k) \;\middle|\; \sum_{i=1}^k x_i \in A \right\} \\
&= F^{-1}(\pi^{-1}(A))
\end{aligned}$$

and therefore

$$\begin{aligned}\mu' = F(\mu_1 * \cdots * \mu_k) &= (\mu_1 \times \cdots \times \mu_k) \circ \pi^{-1} \circ F^{-1} \\ &= (\mu_1 \times \cdots \times \mu_k) \circ F^{-1} \circ \pi^{-1} \\ &= \mu_1' * \cdots * \mu_k'.\end{aligned}$$

For the third moments of $\mu_i' = F\mu_i$ we have the bounds

$$\rho_i' = \int x^3 \; dF\mu_i = 2^{-3\lfloor \log_2 \sqrt{k} \rfloor} \int x^3 \; d\mu \leq O(1)k^{-3/2}.$$

Applying the first inequality in the Berry-Esseen theorem to μ', if γ is a Gaussian with the same mean and variance as μ', we have

$$|\mu'(I) - \gamma(I)| \leq O(1) \cdot \frac{k \cdot k^{-3/2}}{\sigma^{3/2}} = O_\sigma(k^{-1/2})$$

for all intervals I. In particular, μ' agrees with γ on intervals of length 2^{-m-p} to a degree that can be made arbitrarily small by taking k large in a manner depending on σ and p. By Lemma 1.6, for any ε' and large enough k,

$$|H_m((\mu')^{x,i}) - H_m(\gamma^{x,i})| < \varepsilon'.$$

Choosing $p \geq p_0(\sigma, \delta, m)$, $k \geq k_0(\sigma, \delta, m)$ and combining the above with (1.10) and (1.11), if ε' is small with respect to δ we have

$$\mathbb{P}_{i=p}\left(H_m((\mu')^{x,i}) > 1 - \delta\right) > 1 - \delta, \tag{1.12}$$

since the rate of convergence in (1.10) does not depend on k. For every $A \subset \mathbb{R}$,

$$\begin{aligned}(\mu')^{x,p}(A) &= \frac{1}{\mu\left(2^{\lfloor \log_2 \sqrt{k} \rfloor}\mathcal{D}_p(x)\right)}\mu\left(2^{\lfloor \log_2 \sqrt{k} \rfloor}(\mathcal{D}_p(x) \cap T^{-1}_{\mathcal{D}_p(x)}(A))\right) \\ &= \frac{1}{\mu(\mathcal{D}_{p-\lfloor \log_2 \sqrt{k} \rfloor}(y))}\mu(\mathcal{D}_{p-\lfloor \log_2 \sqrt{k} \rfloor}(y) \cap T^{-1}_{\mathcal{D}_{p-\lfloor \log_2 \sqrt{k} \rfloor}(y))}(A)) \\ &= \mu^{y,p-\lfloor \log_2 \sqrt{k} \rfloor}(A),\end{aligned}$$

where $y = 2^{\lfloor \log_2 \sqrt{k} \rfloor}x$. Hence, the inequality (1.12) scaled back to μ is

$$\mathbb{P}_{i=p-\lfloor \log_2 \sqrt{k} \rfloor}(H_m(\mu^{x,i}) > 1 - \delta) > 1 - \delta,$$

which is what we wanted to show. $\square$

The following proposition specializes this results for a k-fold self convolution. We now require the average variance of components of any fixed scale to be positive.

Proposition 1.18. *Let $\sigma, \delta > 0$ and let $m \geq 2$ be an integer. Then there exists a $p = p_1(\sigma, \delta, m)$ such that for all 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*

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

If $\lambda > \sigma$, then for $j_0 = i_0 - \lfloor \log_2 \sqrt{k} \rfloor + p$ and $\nu = \mu^{*k}$, we have

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

Proof ([7], p. 23). Let μ, λ and m be given. Fix integers k and p. We will later see how large they should be. Let $i_0 \geq 0$ and define $j_0 = i_0 - \lfloor \log_2 \sqrt{k} \rfloor + p$ as in the statement. Denote by $\tilde{\mu}$ the k-fold self-product $\tilde{\mu} = \mu \times \cdots \times \mu$ and let $\pi : \mathbb{R}^k \to \mathbb{R}$ be the addition map

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

Then $\nu = \mu^{*k} = \pi\tilde{\mu}$ and, since $\tilde{\mu} = \mathbb{E}_{i=i_0}(\tilde{\mu}_{x,i})$,

$$\begin{aligned}
\nu &= \pi\left(\mathbb{E}_{i=i_0}(\tilde{\mu}_{x,i})\right) \\
&= \pi\left(\sum_{D \in \mathcal{D}_{i_0}} \tilde{\mu}(D)\tilde{\mu}_D\right) \\
&= \sum_{D \in \mathcal{D}_{i_0}} \tilde{\mu}(D)\pi\tilde{\mu}_D \\
&= \mathbb{E}_{i=i_0}(\pi\tilde{\mu}_{x,i}).
\end{aligned}$$

Denoting $\pi\tilde{\mu}_{x,i_0}$ by η, we will first show that for a $\delta_1 > 0$ depending on δ, the measure η satisfies

$$\mathbb{P}_{j=j_0}\left(\eta^{y,j} \text{ is } (\delta_1, m)\text{-uniform}\right) > 1 - \delta_1. \tag{1.13}$$

Note that the random component $\tilde{\mu}_{x,i_0}$ is a product of components of μ, $\tilde{\mu}_{x,i_0} = \mu_{x_1,i_0} \times \cdots \times \mu_{x_k,i_0}$ and the marginal measures μ_{x_j,i_0} are independently distributed according to μ. Note also that the raw components differ from the rescaled ones only by a scaling factor 2^{i_0} and a translation; for all A, if $D = [k2^{-i_0}, (k+1)2^{-i_0})$ is the level-i_0 atom containing x_j,

$$\begin{aligned}\mu_{x_j,i_0}(A) &= \frac{1}{\mu(D)}\mu(D \cap A) \\ &= \frac{1}{\mu(D)}\mu\left(D \cap (2^{-i_0}((2^{i_0}A - k) + k))\right) = \mu^{x_j,i_0}(2^{i_0}A - k).\end{aligned}$$

Hence the expected variance of the raw components is $2^{-2i_0}\lambda$. Recall that $\mathrm{Var}(\pi\tilde{\mu}_{x,i_0}) = \sum_{j=1}^{k}\mathrm{Var}(\mu_{x_j,i_0})$. For any $\delta_2 > 0$, the weak law of large numbers tells us there is such a k that

$$\mathbb{P}_{i=i_0}\left(\left|\frac{1}{k}\mathrm{Var}(\pi\tilde{\mu}_{x,i}) - 2^{-2i}\lambda\right| < 2^{-2i}\delta_2\right) > 1 - \delta_2.$$

Note that since the rescaled components are measures on $[0,1)$, λ is finite. Since $\lambda > \sigma$ by assumption, if we choose δ_2 small enough in a manner depending on σ, the inequality in the event above implies

$$\mathrm{Var}(\pi\tilde{\mu}_{x,i_0}) \geq 2^{-2i_0}\cdot k\sigma > 2^{-2i_0}\cdot k\sigma/2 \tag{1.14}$$

with probability greater than $1 - \delta_2$ over the choice of $\tilde{\mu}_{x,i_0}$.

By the inequality above, $\mathrm{Var}(\eta) \geq \sigma_1(\sigma)k$, so by Proposition 1.17 there exists an integer $p = p_0'(\sigma_1(\sigma), \delta_1, m) = p_0(\sigma, \delta_1, m)$ such that for all $k \geq k_0(\sigma, \delta_1, m)$, we have

$$\mathbb{P}_{j=i_0+p-\lfloor\log_2\sqrt{k}\rfloor}\left(\eta^{y,j} \text{ is } (\delta_1, m)\text{-uniform}\right) > 1 - \delta_1.$$

Here the scale is increased by i_0 because the components μ_{x_j,i_0} are probability measures on $D_{i_0}(x_j) \in \mathcal{D}_{i_0}$ instead of on $[0,1]$ as in the statement of the proposition; replacing $[0,1]$ with $D_{i_0}(x_j)$ and increasing the scaling of F in the proof by 2^{i_0} we arrive at the same result with scale $i_0+p-\lfloor\log_2\sqrt{k}\rfloor = j_0$.

In the discussion above, we assumed the inequality (1.14) to hold. Since it holds with probability greater than $1 - \delta_2$, what we have calculated is

$$\mathbb{P}_{i=i_0}\left(\mathbb{P}_{j=j_0}\left(\eta^{y,j} \text{ is } (\delta_1, m)\text{-uniform}\right) > 1 - \delta_1\right) > 1 - \delta_2.$$

Using concavity of entropy and Markov's inequality, we can estimate

$$
\begin{aligned}
\mathbb{E}_{j=j_0}\left(H_m(\nu^{x,j})\right) &= \frac{1}{m}H(\nu, \mathcal{D}_{j_0+m}|\mathcal{D}_{j_0}) \\
&= \frac{1}{m}H\left(\mathbb{E}_{i=i_0}(\eta), \mathcal{D}_{j_0+m}|\mathcal{D}_{j_0})\right) \\
&\geq \mathbb{E}_{i=i_0}\left(\frac{1}{m}H\left(\mathbb{E}_{j=j_0}(\eta_{y,j}), \mathcal{D}_{j_0+m}|\mathcal{D}_{j_0})\right)\right) \\
&\geq \mathbb{E}_{i=i_0}\left(\mathbb{E}_{j=j_0}\left(\frac{1}{m}H(\eta_{y,j}, \mathcal{D}_{j+m}|\mathcal{D}_j)\right)\right) \\
&= \mathbb{E}_{i=i_0}\left(\mathbb{E}_{j=j_0}\left(H_m(\eta^{y,j})\right)\right) \\
&\geq (1-\delta_1)\mathbb{E}_{i=i_0}\left(\mathbb{P}_{j=j_0}\left(H_m(\eta^{y,j}) > 1-\delta_1\right)\right) \\
&\geq (1-\delta_1)^2\mathbb{P}_{i=i_0}\left(\mathbb{P}_{j=j_0}\left(H_m(\eta^{y,j}) > 1-\delta_1\right) > 1-\delta_1\right) \\
&> (1-\delta_1)^3
\end{aligned}
$$

by choosing $\delta_2 < \delta_1$. Hence,

$$
\begin{aligned}
\mathbb{P}_{j=j_0}\left(\nu^{x,j} \text{ is } (\delta, m)\text{-uniform}\right) &\geq 1 - \frac{1 - \mathbb{E}_{j=j_0}\left(H_m(\nu^{x,j})\right)}{\delta} \\
&> 1 - \frac{1-(1-\delta_1)^3}{\delta} \\
&> 1-\delta,
\end{aligned}
$$

when we choose $\delta_1 < 1-(1-\delta^2)^{1/3}$.

□

We point out a quite intuitive and useful connection between the entropy and variance of a measure.

Lemma 1.19. *Let m be an integer and let $\varepsilon > 0$ be small. For all $\mu \in \mathcal{P}([0,1])$, the following holds: if $\operatorname{Var}(\mu) < \varepsilon$, then $H_m(\mu) \leq 1/m + O_m(\varepsilon)$ and if $H_m(\mu) < \varepsilon$, then $\operatorname{Var}(\mu) \leq 2^{-m} + O_m(\varepsilon)$,*

Proof. Let $\mu \in \mathcal{P}([0,1])$ and let D' be the union of two atoms of $\mathcal{D}_m$ closest to $\langle\mu\rangle$. Let $k2^{-m}$ be the endpoint of this interval closest to $\langle\mu\rangle$. Assume that $\operatorname{Var}(\mu) < \varepsilon$. Then

$$
\begin{aligned}
&\varepsilon > \operatorname{Var}(\mu) = \int_{[0,1]} (x - \langle\mu\rangle)^2 \, d\mu \geq \int_{[0,1]\setminus D'} (x-\langle\mu\rangle)^2 \, d\mu \\
&\geq (k2^{-m} - \langle\mu\rangle)^2\mu([0,1]\setminus D') \\
&\geq 2^{-4m}\mu([0,1]\setminus D'),
\end{aligned}
$$

since $(x - \langle\mu\rangle)^2 \geq (k2^{-m} - \langle\mu\rangle)^2$ whenever x is outside D' and by the choice of D', $(k2^{-m} - \langle\mu\rangle)^2 \geq (2^{-2m})^2$.

Hence $\mu(D) \leq 2^{4m}\varepsilon$ for all $D \in \mathcal{D}_m \backslash D'$ and $-\sum_{D\in\mathcal{D}_m\backslash D'} \mu(D)\log_2 \mu(D) \leq O_m(\varepsilon)$, since the function $x \mapsto x\log_2 x$ is continuous and the sum is finite. Again, since D' contains two atoms of D_m, $-\sum_{D\in D'} \mu(D)\log_2 \mu(D) \leq \log_2(2) = 1$. Combining these upper bounds, we have $H_m(\mu) \leq O_m(\varepsilon) + 1/m$.

Assume then that $H_m(\mu) < \varepsilon$. Then there is exactly one atom $D'' \in \mathcal{D}_m$ with μ-measure close to 1 and $\mu(D) \leq O_m(\varepsilon)$ for every other $D \in \mathcal{D}_m \setminus D''$. Therefore,

$$\mathrm{Var}(\mu) = \int_{[0,1]} (x - \langle\mu\rangle)^2 \, d\mu \leq \sum_{D\in\mathcal{D}_m\backslash D''} 1 \cdot \mu(D) + 2^{-2m}\mu(D'') \leq O_m(\varepsilon) + 2^{-m}.$$

□

We conclude that a probability measure with small variance is locally almost atomic.

Corollary 1.20. *Let $m \in \mathbb{N}$ and $\varepsilon > 0$. For $N > N(m,\varepsilon)$ and $0 < \delta < \delta(m,\varepsilon,N)$, if $\mu \in \mathcal{P}([0,1])$ and Var$(\mu) < \delta$, then*

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

Proof. Let N be an integer. Using Lemma 1.19, choose an integer $m' = m'(\varepsilon,m) > m$ and a positive $\varepsilon' = \varepsilon'(\varepsilon,m') < \varepsilon$ such that the inequality $H_{m'}(\mu^{x,i}) < \varepsilon'$ implies $\mathrm{Var}(\mu^{x,i}) \leq 2^{-m'} + O_{m'}(\varepsilon') < \varepsilon$. Then it suffices to find such N and δ that $\mathrm{Var}(\mu) < \delta$ implies

$$\mathbb{P}_{0\leq i<N}\left(H_{m'}(\mu^{x,i}) < \varepsilon' \text{ and } H_m(\mu^{x,i}) < \varepsilon\right) > 1 - \varepsilon. \tag{1.15}$$

Indeed, choose $N(m,\varepsilon)$ and $\varepsilon'' = \varepsilon''(m,\varepsilon)$ so that by Lemma 1.11, the inequality $H_N(\mu) < \varepsilon''$ implies

$$\mathbb{P}_{0\leq i<N}\left(H_{m'}(\mu^{x,i}) < \varepsilon'\right) > 1 - \varepsilon'. \tag{1.16}$$

Then, since $m' > m$,

$$H_m(\mu^{x,i}) \leq \frac{m'}{m}\frac{1}{m'} \cdot H(\mu^{x,i}, D_{m'}) = \frac{m'}{m} H_{m'}(\mu^{x,i})$$

and when $\varepsilon' < m/m' \cdot \varepsilon$,

$$\mathbb{P}_{0\leq i<N}\left(H_m(\mu^{x,i}) < \varepsilon\right) \geq \mathbb{P}_{0\leq i<N}\left(H_{m'}(\mu^{x,i}) < \varepsilon'\right) > 1 - \varepsilon' > 1 - \varepsilon. \tag{1.17}$$

Note that the inequality 1.16 and therefore also the inequality above hold with all $N > N(m,\varepsilon)$, since $\varepsilon' \geq \sqrt{\varepsilon'' + O(m/N(m,\varepsilon))} \geq \sqrt{\varepsilon'' + O(m/N)}$. Finally, increase N and choose $\delta = \delta(m,\varepsilon,N)$ so that $\mathrm{Var}(\mu) < \delta$ implies $H_N(\mu) \leq 1/N + O_N(\delta) < \varepsilon''$ by Lemma 1.19. Now (1.16) and (1.17) imply (1.15), which in turn implies the statement of the corollary. $\square$

As a conclusion of Propositions 1.17 and 1.18, we show that for any probability measure on the unit interval, the typical components of the k-fold self convolution are almost uniform, unless the typical components of the measure itself are almost atomic.

Theorem 1.21. *Let $0 < \delta < 1$ 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 $I, J \subset \{1, \ldots, n\}$ with $|I \cup J| > (1-\delta)n$ such that, writing $\nu = \mu^{*k}$,*

$$\mathbb{P}_{i=q}\left(\nu^{x,i} \text{ is } (\delta,m)\text{-uniform}\right) \geq 1 - \delta \text{ for } q \in I, \tag{1.18}$$
$$\mathbb{P}_{i=q}\left(\mu^{x,i} \text{ is } (\delta,m)\text{-atomic}\right) \geq 1 - \delta \text{ for } q \in J. \tag{1.19}$$

Proof ([7], p. 25). Let $0 < \delta < 1$ and $m \geq 2$ be given. In this proof, $k_1(\cdot)$ and $p_1(\cdot)$ are the functions from Proposition 1.18. Let $\tilde{\rho} : (0,1] \to (0,1]$ be a function with the following requirements that become more clear over the course of the proof: for all $\sigma \in (0,1]$,

1) $\tilde{\rho}(\sigma) < \dfrac{\sigma}{2}$

2) $\tilde{\rho}(\sigma) < (\rho')^2$, where $\rho' = \rho'(\delta, \sigma)$ is a small number specified later

3) $\tilde{\rho}(\sigma) < \delta$.

Define a decreasing sequence $\sigma_0 > \sigma_1 > \ldots$ recursively by setting $\sigma_0 = 1$ and $\sigma_i = \tilde{\rho}(\sigma_{i-1})$ for all $i \geq 1$. Fix a probability measure μ on $[0,1]$ and a large integer n; we shall later see how large an n is desirable. For $0 \leq q < n$, 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 $\mathbb{P}_{0 \leq q < n}\left(\lambda_q \in (\sigma_s, \sigma_{s-1}]\right) < \frac{\delta}{2}$; if this was not the case, we would have $\mathbb{P}_{0 \leq q < n}\left(\lambda_q \in (\sigma_s, \sigma_{s-1}]\right) \geq \frac{\delta}{2}$ for all $s \in [1, 1 + 2/\delta]$ and

$$\mathbb{P}_{0 \leq q < n}\left(\lambda_q \in (\sigma_{\left\lfloor 1 + \frac{2}{\delta} \right\rfloor}, 1]\right) \geq \left\lfloor \frac{2}{\delta} + 1 \right\rfloor \frac{\delta}{2} > 1.$$

For this s, define

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

and set

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

Then by the choice of s, we have

$$\begin{aligned}\mathbb{P}_{0\leq q<n}\left(\lambda_q \notin (\sigma_s, \sigma_{s-1}]\right) &= \mathbb{P}_{0\leq q<n}\left(\lambda_q \leq \rho\right) + \mathbb{P}_{0\leq q<n}\left(\lambda_q > \sigma\right) \\ &= \frac{|I' \cup J'|}{n} > 1 - \frac{\delta}{2}.\end{aligned} \tag{1.20}$$

Let $k \geq k_1(\sigma, \delta, m)$ and let $l \geq 0$ be the integer

$$l = \lfloor \log_2 \sqrt{k} \rfloor - p_1(\sigma, \delta, m).$$

Since n may be taken arbitrarily large with respect to l, by deleting at most l elements of I' we can assume that $I' \subset [l, n]$ and that the size bound of (1.20) remains valid. Let

$$I = I' - l \subset [0, n - l].$$

Since $k \geq k_1(\sigma, \delta, m)$, by Proposition 1.18,

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

which is (1.18).

By definition of J',

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

Using this and Markov's inequality, we get

$$\begin{aligned}\mathbb{P}_{i=q}\left(\mathrm{Var}(\mu^{x,i}) < \sqrt{\rho}\right) &= 1 - \mathbb{P}_{i=q}\left(\mathrm{Var}(\mu^{x,i}) \geq \sqrt{\rho}\right) \\ &\geq 1 - \frac{\lambda_q}{\sqrt{\rho}} \geq 1 - \sqrt{\rho} \text{ for } q \in J'.\end{aligned} \tag{1.21}$$

Fix a small $\rho' = \rho'(\delta, \sigma)$ and a large integer $N = N(l, \delta, \rho')$ for which we add more specific requirements over the remainder of the proof. By taking n large relative to N and l, we can assume that $I', J' \subset \{l, \ldots, n - N\}$ while having the size bound of (1.20) still valid. By the definition of $\tilde{\rho}$, we have $\rho < \min\{\delta, (\rho')^2, \sigma/2\}$ and hence by Corollary 1.20 (with $\sqrt{\rho}$ in the place of the δ in the lemma and $\min\{\rho', \sigma/2, \delta\}$ in the place of ε), any measure $\theta \in \mathcal{P}([0,1])$ satisfying $\mathrm{Var}(\theta) < \sqrt{\rho}$ also satisfies

$$\mathbb{P}_{0\leq i<N}\left(\mathrm{Var}(\theta^{y,i}) < \sigma/2 \text{ and } \theta^{y,i} \text{ is } (\delta, m)\text{-atomic}\right) > 1 - \rho'.$$

We point out the following, quite intuitive identity:

$$\begin{aligned}
(\mu^{x,q})^{T_{\mathcal{D}_q(x)}(x),i} &= \frac{1}{\mu^{x,q}(\mathcal{D}_i(T_{\mathcal{D}_q(x)}(x))}\mu^{x,q}_{|\mathcal{D}_i(T_{\mathcal{D}_q(x)}(x))} \circ T^{-1}_{\mathcal{D}_i(T_{\mathcal{D}_q(x)}(x))} \\
&= \frac{1}{\mu(\mathcal{D}_{q+i}(x))}(\mu_{|\mathcal{D}_q(x)} \circ T^{-1}_{\mathcal{D}_q(x)})_{|D_i(T_{\mathcal{D}_q(x)}(x))} \circ T^{-1}_{\mathcal{D}_i(T_{\mathcal{D}_q(x)}(x))} \\
&= \frac{1}{\mu(\mathcal{D}_{q+i}(x))}\mu_{|\mathcal{D}_{q+i}(x)} \circ (T_{\mathcal{D}_i(T_{\mathcal{D}_q(x)}(x))} \circ T_{\mathcal{D}_q(x)})^{-1} \\
&= \frac{1}{\mu(\mathcal{D}_{q+i}(x))}\mu_{|\mathcal{D}_{q+i}(x)} \circ T^{-1}_{\mathcal{D}_{q+i}(x)} \\
&= \mu^{x,q+i}.
\end{aligned}$$

Since $\sqrt{\rho} < \rho'$, combining the above with (1.21) gives us

$$\begin{aligned}
&\mathbb{P}_{q\leq i<q+N}\left(\mathrm{Var}(\mu^{x,i}) < \sigma/2 \text{ and } \mu^{x,i} \text{ is } (\delta,m)\text{-atomic}\right) \\
&\geq \mathbb{P}_{0\leq i<N}\left(\mathrm{Var}(\mu^{x,q+i}) < \sigma/2 \text{ and } \mu^{x,q+i} \text{ is } (\delta,m)\text{-atomic, and } \mathrm{Var}(\mu^{x,q}) < \sqrt{\rho}\right) \\
&= \mathbb{P}_{0\leq i<N}\left(\mathrm{Var}\left(\mu^{x,q+i}\right) < \sigma/2 \text{ and } \mu^{x,q+i} \text{ is } (\delta,m)\text{-atomic} \;\middle|\; \mathrm{Var}(\mu^{x,q}) < \sqrt{\rho}\right) \\
&\quad \cdot \mathbb{P}_{i=q}\left(\mathrm{Var}(\mu^{x,q}) < \sqrt{\rho}\right) \\
&> (1-\rho')(1-\sqrt{\rho}) > 1 - 2\rho' \text{ for } q \in J'.
\end{aligned} \tag{1.22}$$

Define the set

$$U = \left\{h \in \mathbb{N} \mid \mathbb{P}_{i=h}\left(\mathrm{Var}(\mu^{x,i}) < \sigma/2 \text{ and } \mu^{x,i} \text{ is } (\delta,m)\text{-atomic}\right) > 1 - \sqrt{2\rho'}\right\}.$$

By Lemma 1.13 and the inequality (1.22),

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

Applying Lemma 1.15 to J' and U, we obtain $U' \subset U$ for which $|U'| > (1 - 2\sqrt{2\rho'})|J'|$ and $|U' \cap (U' - l)| > (1 - 2\sqrt{2\rho'} - \frac{l}{N})|J'|$. Defining

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

and choosing N, l so that $\frac{l}{N} < \sqrt{2\rho'}$, we obtain for the size of J that

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

Next we will show that I and J are disjoint. Suppose that this is not the case and let $q \in I \cap J$. Since $I = I' - l$, then $q + l \in I'$. By the definition of I', $\lambda_{q+l} > \sigma$. On the other hand, $q \in J \subset U' - l$ implies $q + l \subset U' \subset U$, so using the definition of U and assuming that ρ' satisfies $\sqrt{2\rho'} < \sigma/2$, we have

$$\begin{aligned}
\lambda_{q+l} &= \mathbb{E}_{i=q+l}\left(\mathrm{Var}(\mu^{x,i})\right) \\
&= \int_0^1 \mathrm{Var}(\mu^{x,q+l})\, d\mu \\
&\leq \frac{\sigma}{2} \cdot \mu(\{x \mid \mathrm{Var}(\mu^{x,q+l}) < \sigma/2\}) + 1 \cdot \mu(\{x \mid \mathrm{Var}(\mu^{x,q+l}) > \sigma/2)\}) \\
&= \frac{\sigma}{2} \cdot \mathbb{P}_{i=q+l}\left(\mathrm{Var}(\mu^{x,i}) < \sigma/2\right) + 1 \cdot \mathbb{P}_{i=q+l}\left(\mathrm{Var}(\mu^{x,i}) \geq \sigma/2\right) \\
&< \frac{\sigma}{2} + \sqrt{2\rho'} \\
&< \sigma,
\end{aligned}$$

which is a contradiction. Hence $I \cap J = \emptyset$.

It remains to prove the lower bound for $|I \cup J|$. Since $I' \cap J' = \emptyset$ and $|I' \cup J'| > (1 - \frac{\delta}{2})n$ by (1.20), adding to our assumptions of ρ' that it satisfies $3\sqrt{3\rho'} < \delta/2$, we have

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

which completes the proof. □

The following lemma shows that for probability measures μ and ν on a countable commutative group, the entropy of the convolution of μ and ν^{*k} does not increase as k grows. We will afterwards generalize the result to $\mathbb{R}$, using a discretization argument.

Lemma 1.22. *Let $\Gamma \subset \mathbb{R}$ 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 non-increasing in k. In particular,

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

Proof ([7], p. 27). Let X_0 and $Z_1, \ldots, Z_n$ be independent, bijective random variables taking values in Γ. Let μ be the distribution of X_0 and ν the distribution of Z_i for all i. Define the random variable $X_n = X_0 + Z_1 + \cdots + Z_n$, the distribution of which is then $\mu * \nu^{*k}$. Since Γ is commutative, given $Z_1 = g$, the distribution of X_n is the same as the distribution of $X_{n-1} + g$. Hence

$$\begin{aligned} H(X_n|Z_1) &= \sum_{z_i \in \Gamma} \nu(z_i) H(X_n|Z_1 = z_i) = \sum_{z_i \in \Gamma} \nu(z_i) H(X_{n-1} + z_i) \\ &= \sum_{z_i \in \Gamma} \nu(z_i) H(X_{n-1}) = H(X_{n-1}). \end{aligned}$$

Furthermore, applying the chain rule for conditional entropy,

$$\begin{aligned} H(Z_1|X_n) &= H(Z_1, X_n) - H(X_n) \\ &= H(Z_1) + H(X_n|Z_1) - H(X_n) \\ &= H(Z_1) + H(X_{n-1}) - H(X_n) \\ &= H(\nu) + H(\mu * \nu^{*(n-1)}) - H(\mu * \nu^{*n}). \end{aligned} \tag{1.23}$$

The random variable X_n is a Markov process, that is, given $X_{n-1} = a$, the value of X_n depends only on a and not on any of the sums $X_0, \ldots, X_{n-2}$. Therefore, given a fixed value for X_n, the random variables $Z_1 = X_1 - X_0$ and X_{n+1} are independent. We have

$$\begin{aligned} H(Z_1|X_{n+1}) &= \sum_{x_i \in \Gamma} (\mu * \nu^{*(n+1)})(x_i) H(Z_1|X_{n+1} = x_i) \\ &\geq \sum_{x_i \in \Gamma} (\mu * \nu^{*(n+1)})(x_i) H((Z_1|X_{n+1} = x_i)|X_n) \\ &= \sum_{x_i \in \Gamma} (\mu * \nu^{*(n+1)})(x_i) \sum_{y_j \in \Gamma} (\mu * \nu^{*n})(y_j) H((Z_1|X_{n+1} = x_i)|X_n = y_j) \\ &= \sum_{x_i \in \Gamma} (\mu * \nu^{*(n+1)})(x_i) \sum_{y_j \in \Gamma} (\mu * \nu^{*n})(y_j) H(Z_1|X_n = y_j) \\ &= H(Z_1|X_n). \end{aligned}$$

Applying (1.23) to both sides of this inequality we get

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

which proves that the δ_k defined in the lemma is non-increasing. In particular, we see that

$$\begin{aligned} & H(\mu * \nu^{*k}) - H(\mu) \\ &= H(\mu * \nu^{*k}) - H(\mu * \nu^{*(k-1)}) + H(\mu * \nu^{*(k-1)}) - \ldots + H(\mu * \nu) - H(\mu) \\ &= \delta_k + \delta_{k-1} + \ldots \delta_0 \\ &\leq k \cdot \delta_0, \end{aligned}$$

which is the second statement. $\square$

To get the analogous statement for measures on $\mathbb{R}$, we use a discretization argument; for $m \in \mathbb{N}$, define

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

Then each atom of $\mathcal{D}_m$, as an interval of length 2^{-m}, contains exactly one element of M_m. Define the *m-discretization map* $\sigma_m : \mathbb{R} \to M_m$ by setting $\sigma_m(x) = v$ 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 its support is contained in M_m. For any μ, define its m-discretization $\sigma_m\mu$ as the push-forward through σ_m;

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

Observe that the scale-m entropy of a discretization is equal to that of the original measure,

$$\begin{aligned} H_m(\mu) &= \frac{-1}{m} \sum_{D \in \mathcal{D}_m} \mu(D) \log_2 \mu(D) \\ &= \frac{-1}{m} \sum_{v \in M_m^d} \mu(\mathcal{D}_m(v)) \log_2 \mu(\mathcal{D}_m(v)) \\ &= H_m(\sigma_m\mu). \end{aligned}$$

Similar result holds for the convolution of discretizations, up to an error term that diminishes as the scale decreases.

Lemma 1.23. *Given $\mu_1, \dots, \mu_k \in \mathcal{P}(\mathbb{R})$ with $H(\mu_i) < \infty$ and $m \in \mathbb{N}$,*

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

Proof. Denote by $\pi : \mathbb{R}^k \to \mathbb{R}$ the addition map $(x_1, \dots, x_k) \mapsto \sum_{i=1}^k x_i$. Then $\mu_1 * \cdots * \mu_k = \pi(\mu_1 \times \cdots \times \mu_k)$ and $\sigma_m\mu_1 * \cdots * \sigma_m\mu_k = \pi \circ \sigma_m^k(\mu_1 \times \cdots \times \mu_k)$, where $\sigma_m^k(x_1, \dots, x_k) = (\sigma_m x_1, \dots, \sigma_m x_k)$. Since

$$|\pi(x_1, \dots, x_k) - \pi \circ \sigma_m^k(x_1, \dots, x_k)| = \left| \sum_{i=1}^k (x_i - \sigma_m(x_i)) \right| \leq k \cdot 2^{-m},$$

the result follows from Lemma 1.5 (iii.). $\square$

We may now generalize Lemma 1.22 to probability measures on the line.

Proposition 1.24. *Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ with $H_n(\mu), H_n(\nu) < \infty$. Then*

$$H_n(\mu * \nu^{*k}) \leq H_n(\mu) + k \cdot (H_n(\mu * \nu) - H_n(\mu)) + O\left(k/m\right).$$

Proof. Write $\tilde{\mu} = \sigma_n \mu$ and $\tilde{\nu} = \sigma_n \nu$ for the n-discretizations of μ and ν. Since $(M_m, +)$ is a commutative group, we have by Lemma 1.22 that

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

For an n-discrete measure, the entropy with respect to partition into its atoms is equal to the scale-n entropy. Since $\tilde{\mu} * \tilde{\nu}^{*k}$ is also discrete, having its support contained in $\sum_{i=1}^{k} M_m$, dividing the inequality above by n we obtain

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

We obtain the result by applying Lemma 1.23 to $H_n(\tilde{\mu} * \tilde{\nu}^{*k})$ and $H_n(\tilde{\mu} * \tilde{\nu})$. $\square$

Using discretizations, we may easily prove the trivial inequality $H_n(\mu * \nu) \leq H_n(\mu) - O(1/m)$.

Lemma 1.25. *For $m \in \mathbb{N}$ and $\mu, \nu \in \mathcal{P}([-r, r]^d)$ with $H_n(\mu) < \infty$ and $H_n(\nu) < \infty$,*

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

Proof. Denote the m-discretizations of μ and ν by $\tilde{\mu}$ and $\tilde{\nu}$ as in the previous proof. We can write the convolution of $\tilde{\mu}$ and $\tilde{\nu}$ as an integral of a translation of $\tilde{\mu}$ in the following way:

$$\tilde{\mu} * \tilde{\nu} = \sum_{y \in M_m} \tilde{\mu}(\cdot - y)\tilde{\nu}(y) = \sum_{y \in M_m} \tilde{\mu}(\cdot - y)\delta_y(y)\tilde{\nu}(y) = \int_{\mathbb{R}} (\tilde{\mu} * \delta_y) \, d\tilde{\nu}(y).$$

Applying concavity of entropy and Lemma 1.5 (iv.), we get

$$\begin{aligned}
H_m(\tilde{\mu} * \tilde{\nu}) &= H_m\left(\int (\tilde{\mu} * \delta_y) \, d\tilde{\nu}(y)\right) \\
&\geq \int H_m(\tilde{\mu} * \delta_y) \, d\tilde{\nu}(y) \\
&= \int (H_m(\tilde{\mu}) - O(1/m)) \, d\tilde{\nu}(y) \\
&= H_m(\tilde{\mu}) - O(1/m).
\end{aligned}$$

Applying Lemma 1.23 to this inequality, we obtain

$$H_m(\mu * \nu) \geq H_m(\tilde{\mu}) - O(1/m) = H_m(\mu) - O(1/m).$$

The inequality $H_m(\mu * \nu) \geq H_m(\nu) - O(1/m)$ is obtained identically. $\square$

1.4 The inverse theorem

We now have all the tools required to finish the discussion on the conditions in which

$$H_m(\mu * \nu) \leq H_m(\nu) + \delta$$

holds. The section focused on entropy is concluded in the proof of the following theorem, given by Hochman in [7]. It will be of use to us later, when small-scale analysis of entropy turns out to be a powerful tool in analysing some other properties of a measure.

Theorem 1.26. *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 $I, J \subset \{0, \dots, n-1\}$ with $|I \cup J| > (1-\varepsilon)n$, such that

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

Instead of commencing to prove the theorem as it is, we formulate another result, of which Theorem 1.26 is a formal consequence.

Theorem 1.27. *For every $0 < \varepsilon_1, \varepsilon_2 < 1$ and integers $m_1, m_2 \geq 2$, 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 $H_n(\mu * \nu) \geq H_n(\mu) + \delta$, or there exist disjoint subsets $I, J \subset \{0, \dots, n-1\}$ with $|I \cup J| \geq (1-\varepsilon)n$ and*

$$\mathbb{P}_{i=k}\left(\mu^{x,i} \textit{ is } (\varepsilon_1, m_1)\textit{-uniform}\right) > 1 - \varepsilon_1 \textit{ for } k \in I,$$
$$\mathbb{P}_{i=k}\left(\nu^{x,i} \textit{ is } (\varepsilon_2, m_2)\textit{-atomic}\right) > 1 - \varepsilon_2 \textit{ for } k \in J.$$

Observe that choosing $\varepsilon_1 = \varepsilon_2 = \varepsilon'$, $m_1 = m_2 = m$ and assuming that $H_n(\mu * \nu) < H_n(\mu) + \delta$, this theorem implies Theorem 1.26.

Proof of Theorem 1.27 ([7], p. 29). Assume first that $\varepsilon_1 = \varepsilon_2$ and $m_1 = m_2 = m$; we shall later remove these restrictions. Let $k = k(\varepsilon, m)$ be as in Theorem 1.21 with $\delta = \varepsilon/2$ and let $\mu, \nu \in \mathcal{P}([0,1])$. Define

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

Assuming n is large enough, Theorem 1.21 gives us disjoint subsets $I, J \subset \{0, \dots, n-1\}$ such that $|I \cup J| > (1 - \varepsilon/2)n$ and

$$\mathbb{P}_{i=k}\left(\tau^{x,i} \text{ is } \left(\frac{\varepsilon}{2}, m\right)\text{-uniform}\right) > 1 - \frac{\varepsilon}{2} \text{ for } k \in I \tag{1.24}$$

and

$$\mathbb{P}_{i=k}\left(\nu^{x,i} \text{ is } (\varepsilon, m)\text{-atomic}\right) \geq 1 - \frac{\varepsilon}{2} \text{ for } k \in J. \tag{1.25}$$

Denote by I_0 (which may be empty) the subset of I for which

$$\mathbb{P}_{i=k}\left(\mu^{x,i} \text{ is } (\varepsilon, m)\text{-uniform}\right) > 1 - \varepsilon \text{ for } k \in I_0. \tag{1.26}$$

If $|I_0| > (1-\varepsilon/2)|I|$, we have $|I_0 \cup J| \geq (1-\varepsilon/2)|I| + |J| \geq (1-\varepsilon/2)|I \cup J| > (1-\varepsilon/2)^2 n > (1-\varepsilon)n$. By inequalities (1.25) and (1.26) the pair J, I_0 is then of the form presented in the second alternate statement of the theorem.

If this is not the case, we define $I_1 = I \backslash I_0$. Then $|I_1| = |I| - |I_0| \geq |I|\varepsilon/2 \geq n\varepsilon/4$, since if we had $|I| \leq n\varepsilon/2$, then $|J| \geq (1-\varepsilon/2)n - n\varepsilon/2 = (1-\varepsilon)n$ and the theorem holds with $I = \emptyset$. Now, for all $k \in I_1$, we have by independence of $\mu^{y,i}$ and $\tau^{x,i}$ that

$$\begin{aligned}\mathbb{P}_{i=k}\left(\tau^{x,i} \text{ is } \left(\frac{\varepsilon}{2}, m\right)\text{-uniform and } \mu^{y,i} \text{ is not } (\varepsilon, m)\text{-uniform}\right) &> \left(1 - \frac{\varepsilon}{2}\right)\varepsilon \\ &\geq \frac{\varepsilon}{2}.\end{aligned}$$

Expressed in terms of entropy, for $k \in I_1$, this event implies that

$$H_m(\tau^{y,k}) > 1 - \varepsilon + \varepsilon/2 > H_m(\mu^{x,k}) + \varepsilon/2$$

and hence by Lemma 1.25, we have

$$H_m(\mu^{x,k} * \tau^{y,k}) \geq H_m(\mu^{x,k}) + \varepsilon/2 - O(1/m) \tag{1.27}$$

for all $k \in I_1$ with $\mu \times \tau$-probability greater than $\varepsilon/2$. For all other scales i, we have the deterministic bound $H_m(\mu^{x,i} * \tau^{y,i}) \geq H_m(\mu^{x,i}) - O(1/m)$. Let A be the subset of $\mathbb{R}^2$ in which we have the bound 1.27. Using Lemmas 1.7, 1.8 and the inequalities above, and assuming m is sufficiently large with respect to ε,

$$\begin{aligned}
H_n(\mu * \tau) \geq& \mathbb{E}_{0\leq i<n}(H_m(\mu^{x,i} * \tau^{y,i})) + O\left(\frac{m}{n}\right) \\
=& \frac{|I_1|}{n}\mathbb{E}_{i\in I_1}(H_m(\mu^{x,i} * \tau^{y,i})) \\
&+ \frac{n-|I_1|}{n}\mathbb{E}_{i\in I\setminus I_1}(H_m(\mu^{x,i} * \tau^{y,i})) + O\left(\frac{m}{n}\right) \\
=& \frac{|I_1|}{n}\sum_{i\in I_1}\frac{1}{|I_1|}\int_{\mathbb{R}^2} H_m(\mu^{x,i} * \tau^{y,i})\ d\mu\times\nu \\
&+ \frac{n-|I_1|}{n}\mathbb{E}_{i\in I\setminus I_1}(H_m(\mu^{x,i} * \tau^{y,i})) + O\left(\frac{m}{n}\right) \\
>& \frac{|I_1|}{n}\sum_{i\in I_1}\frac{1}{|I_1|}\int_{\mathbb{R}^2\setminus A} H_m(\mu^{x,i})\ d\mu\times\nu \\
&+ \frac{|I_1|}{n}\sum_{i\in I_1}\frac{1}{|I_1|}\left(\int_A H_m(\mu^{x,i})\ d\mu\times\nu + \left(\frac{\varepsilon}{2}\right)^2\right) \\
&+ \frac{n-|I_1|}{n}\mathbb{E}_{i\in I\setminus I_1}(H_m(\mu^{x,i})) + O\left(\frac{1}{m}+\frac{m}{n}\right) \\
=& \frac{|I_1|}{n}\left(\mathbb{E}_{i\in I_1}(H_m(\mu^{x,i})) + \left(\frac{\varepsilon}{2}\right)^2\right) \\
&+ \frac{n-|I_1|}{n}\mathbb{E}_{i\in I\setminus I_1}(H_m(\mu^{x,i})) + O\left(\frac{1}{m}+\frac{m}{n}\right) \\
\geq& \mathbb{E}_{0\leq i<n}(H_m(\mu^{x,i})) + \frac{n\varepsilon}{4n}\left(\frac{\varepsilon}{2}\right)^2 + O\left(\frac{1}{m}+\frac{m}{n}\right) \\
=& H_n(\mu) + \left(\frac{\varepsilon}{2}\right)^3 + O\left(\frac{1}{m}+\frac{m}{n}\right).
\end{aligned}$$

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

$$H_n(\mu * \tau) > H_n(\mu) + \frac{\varepsilon^3}{10}. \tag{1.28}$$

By Proposition 1.24,

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

Combining this with the inequality (1.28), we obtain

$$H_n(\mu * \nu) \geq \frac{H_n(\mu * \tau) - H_n(\mu)}{k} + H_n(\mu) - O\left(\frac{1}{n}\right) \geq H_n(\mu) + \frac{\varepsilon^3}{100k}$$

which is the first of the two alternate statements of the theorem, with $\delta = \varepsilon^3/100k$.

So far, we have required that m is large, $m = m_1 = m_2$ and $\varepsilon_1 = \varepsilon_2$. We will now generalize the result for arbitrary integers m_1, m_2 and positive real numbers $\varepsilon_1, \varepsilon_2$. Indeed, let $\varepsilon_1, \varepsilon_2, m_1$ and m_2 be given, let $0 < \varepsilon' \leq \min\{\varepsilon_1, \varepsilon_2\}$ be small and choose m' large with respect to ε', m_1, m_2. Applying the previous discussion for a large enough m', either $H_n(\mu * \nu) > H_n(\mu) + (\varepsilon')^3/100k$ or we obtain $I', J' \subset \{0, \ldots, n-1\}$ such that

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

and

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

Assuming these events, applying Lemma 1.12 to $\mu^{x,i} \in \mathcal{P}([0,1])$ with m_1/m' small enough, we get

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

Assuming that $\sqrt{2\varepsilon'} < \varepsilon$ and that n is large, and setting

$$V_1 = \{0 \leq j < n \mid \mathbb{P}_{u=j}(\mu^{x,u} \text{ is } (\varepsilon, m_1)\text{-uniform}) > 1 - \varepsilon\},$$

by Lemma 1.13 we have $|[i, i+m') \cap V_1| > (1 - (2\varepsilon')^{1/4})m'$ for all $i \in I'$, since $[i, i+m') \cap V_1$ contains the subset of $\{i, \ldots, i+m'-1\}$ given by the lemma. Similarly, defining

$$V_2 = \{0 \leq j < n \mid \mathbb{P}_{u=j}(\nu^{x,u} \text{ is } (\varepsilon, m_2)\text{-atomic}) > 1 - \varepsilon\}$$

and applying Lemma 1.11 with m_2/m' small, we obtain

$$\mathbb{P}_{j \leq u < j+m'}\left(\nu^{x,u} \text{ is } (\sqrt{2\varepsilon'}, m_2)\text{-atomic}\right) > 1 - \sqrt{2\varepsilon'}$$

for all $j \in J'$. Hence, by Lemma 1.13, $|[j, j+m') \cap V_2| > (1 - (2\varepsilon')^{1/4})m'$ for all $j \in J'$. Then, applying Lemma 1.16 to pairs I', V_1 and J', V_2, we obtain disjoint sets $I'' \subset I'$ and $J'' \subset J'$ with

$$|I'' \cup J''| \geq (1 - 2(2\varepsilon')^{1/4})^2 |I' \cup J'| \geq (1 - \varepsilon')n$$

when ε' is small enough. These are the subsets of $\{0, \ldots, n-1\}$ presented in the second alternate statement of the theorem.

□

We conclude this section by bringing up an immediate application of the inverse theorem.

Theorem 1.28. *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*

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

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

$$H_n(\nu) > 3\sqrt{\varepsilon} \Longrightarrow H_n(\mu * \nu) \ge H_n(\mu) + \delta.$$

Proof. Let ε be given, $\mu, \nu \in \mathcal{P}([0,1])$ and let n be large with respect to ε. Let $\delta = \delta(\varepsilon, m)$ be the number of the same name given by Theorem 1.27 and assume that $H_n(\mu * \nu) \le H_n(\mu) + \delta$ for all large enough n. Since $\mathbb{P}_{0\le i<n}\left(H_m(\mu^{x,i}) < 1-\varepsilon\right) > 1-\varepsilon$, by Lemma 1.13 there is a subset I of $\{0, \ldots, n-1\}$ with cardinality $|I'| > (1-\sqrt{\varepsilon})n$, such that

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

for all $k \in I'$. Hence, if I is the set of integers such that

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

it has cardinality $|I| < \sqrt{\varepsilon}/n$. Since $H_n(\mu * \nu) \le H_n(\mu) + \delta$, Theorem 1.27 then asserts that if J is the subset of $\{0, \ldots, n-1\}$ for which

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

it has cardinality $|J| > (1-\varepsilon)n - |I| > (1-2\sqrt{\varepsilon})n$. By Lemma 1.7, this in turn implies that

$$\begin{aligned}
H_n(\nu) &= \mathbb{E}_{0\le i<n}\left(H_m(\nu^{x,i})\right) + O\left(\frac{m+1}{n}\right) \\
&= \frac{1}{n}\sum_{k=0}^{n-1}\int_{\mathbb{R}} H_m(\nu^{x,k})\ d\nu + O\left(\frac{m+1}{n}\right) \\
&\le \frac{(1-2\sqrt{\varepsilon})n}{n}\left((\varepsilon(1-\varepsilon)+\varepsilon\right) + \frac{2\sqrt{\varepsilon}n}{n} + O\left(\frac{m+1}{n}\right) \\
&= 2\sqrt{\varepsilon} + 2\varepsilon - 2\varepsilon^{3/2} - \varepsilon^2 - 2\varepsilon^{5/2} + 2\varepsilon^{3/2} + O\left(\frac{m+1}{n}\right) \\
&\le 3\sqrt{\varepsilon}.
\end{aligned}$$

$\square$

2 Fractal dimensions

The latter half of this thesis concerns the dimensions of sets and measures. We begin by introducing some definitions of dimension for both sets and measures separately, although it turns out that the dimension of a measure is strongly connected to the dimensions of sets included in its support. The converse also holds; we may gain information on the structure of a set by inspecting what kind of measures it can support.

2.1 Dimension of a set

We quote K. Falconer ([4]) to give an intuition behind the notion of fractional dimension:

"*Fundamental to most definitions of dimension is 'measurement at scale δ'. For each δ, we measure a set in a way that ignores irregularities of size less than δ, and we see how these measurements behave as $\delta \to 0$.*"

The "measurement at scale δ" can be done from outside, in some sense, by covering the set using sets of some fixed form (e.g. balls or cubes) or arbitrary sets of diameter at most δ. Or, we may approach the measurement from inside by estimating how many disjoint sets of a fixed form we may fit inside the set. Of the three definitions we introduce, Hausdorff dimension is the most common one and is adapted as the main notion in this thesis. The other notions we mention are the box dimension and the packing dimension. For us, the latter two possess no substantial interest of their own, but in some occasions will be used in deriving bounds for the Hausdorff dimension.

Let A be a bounded subset of a metric space $(\mathcal{X}, d)$. For any $\varepsilon > 0$, denote by $N_\varepsilon(A)$ the number of closed balls of radius ε required to cover A. Since A is bounded, this number is finite for every ε.

Definition 2.1. The *upper box dimension* of A is defined by

$$\overline{\dim}_B(A) = \limsup_{\varepsilon \to 0} \frac{\log N_\varepsilon(A)}{\log(1/\varepsilon)}.$$

Similarly, we define the *lower box dimension* as

$$\underline{\dim}_B(A) = \liminf_{\varepsilon \to 0} \frac{\log N_\varepsilon(A)}{\log(1/\varepsilon)}.$$

If these limits both equal a, we call a the *box dimension* of A and denote it by $\dim_B A$.

The box dimension is often the easiest of the three notions to calculate or numerically estimate. However, it is quite rough when it comes to comparing the "size" of two sets; for example, it is easy to see from the definition that the box dimension assigns any dense subset of the space, even if countable, the full dimension, i.e. dimension equal to that of an open ball. This is due to the lack of countable stability, a property desirable from a notion of dimension and possessed by the two other definitions.

The definition of Hausdorff dimension takes a more measure-theoretical approach. This allows it to retain many of the well-behaving properties of a measure.

Definition 2.2. Let A be a subset of a separable metric space $(\mathcal{X}, d)$. For every $0 \leq s < \infty$ and $0 < \delta \leq \infty$, define

$$\mathcal{H}^s_\delta(A) = \inf\{\sum_{i=1}^{\infty} \operatorname{diam}(A_i)^s |\ A_i \subset \mathcal{X},\ A \subset \bigcup_{i=1}^{\infty} A_i,\ \operatorname{diam}(A_i) \leq \delta\}.$$

The *s-dimensional Hausdorff measure* of the set A is then defined by

$$\mathcal{H}^s(A) = \lim_{\delta \to 0} \mathcal{H}^s_\delta(A).$$

The s-dimensional Hausdorff measure is a Borel measure on $\mathcal{X}$. We collect some of its well-known properties in the following proposition. Although we will rarely be interested in the Hausdorff measure of a set, the properties of the measure pass, in some way, to the dimension.

Proposition 2.3. *In the following statements, F is a subset of $(\mathcal{X}, d)$, a is an element of $\mathcal{X}$ and λ is a positive real number.*

(i.) *The Hausdorff measure is translation invariant, that is,* $\mathcal{H}^s(F + a) = \mathcal{H}^s(F)$

(ii.) $\mathcal{H}^s(\lambda F) = \lambda^s \mathcal{H}^s(F)$

(iii.) *If* $\mathcal{H}^s(F) < \infty$ *for some* s*, then* $\mathcal{H}^t(F) = 0$ *for all* $t > s$*. Also, if* $\mathcal{H}^s(F) > 0$ *for some* s*, then* $\mathcal{H}^t(F) = \infty$ *for all* $t < s$

(iv.) *If* F *is a Borel set of* $\mathbb{R}^n$*, then* $\mathcal{H}^n(F) = c(n) \cdot \mathcal{L}^n(F)$*, where* $\mathcal{L}^n$ *is the* n*-dimensional Lebesgue measure and* $c(n) \in \mathbb{R}$*.*

The Hausdorff dimension of a set A is defined as the "jumping point" of s, where the values of $\mathcal{H}^s(A)$ change from infinity to zero. The existence and uniqueness of this value follows from the third statement in the proposition above. More explicitly, the definition is given as follows:

Definition 2.4. Let $A \subset \mathcal{X}$. The Hausdorff dimension of A is defined as

$$\dim_H A = \inf\{0 \le s < \infty |\ \mathcal{H}^s(A) = 0\} = \sup\{0 \le s < \infty |\ \mathcal{H}^s(A) = \infty\}.$$

Following the translation invariance of the Hausdorff measure, the Hausdorff dimension is also translation invariant. Also, by Proposition 2.3 (ii.), scaling a set by a positive scalar does not change its dimension. The following proposition lists some additional properties quite immediate from the definition.

Proposition 2.5. *Let $(\mathcal{X}, d)$ and $(\mathcal{Y}, d')$ be separable metric spaces.*

(i.) Hausdorff dimension satisfies monotonicity: if $E \subset F \subset \mathcal{X}$, then $\dim_H E \le \dim_H F$

(ii.) If $f : \mathcal{X} \to \mathcal{Y}$ is a Lipschitz mapping, that is, there is a $0 < c < \infty$ such that $d'(f(x), f(y)) \le c \cdot d(x, y)$ for all $x, y \in \mathcal{X}$, then $\dim_H f(E) \le \dim_H E$ for all $E \in \mathcal{X}$.

(iii.) Hausdorff dimension is countably stable, that is, if $\{F_i\}_{i\in\mathbb{N}}$ is a countable family of subsets of $\mathcal{X}$, then $\dim_H \bigcup_{i=1}^{\infty} F_i = \sup_{i\in\mathbb{N}}(\dim_H F_i)$

Proof. (i.) Since any cover $\{F_i\}_{i\in\mathbb{N}}$ of F also covers E, for each δ,

$$\begin{aligned}
\mathcal{H}^s_\delta(E) &= \inf\left\{\sum_{i=1}^{\infty} \operatorname{diam}(E_i)^s \mid E \subset \bigcup_{i=1}^{\infty} E_i,\ \operatorname{diam}(E_i) \le \delta\right\} \\
&\le \inf\left\{\sum_{i=1}^{\infty} \operatorname{diam}(F_i)^s \mid F \subset \bigcup_{i=1}^{\infty} F_i,\ \operatorname{diam}(F_i) \le \delta\right\} \\
&= \mathcal{H}^s_\delta(F).
\end{aligned}$$

Taking limit as $\delta \to 0$ yields $\mathcal{H}^s(E) \le \mathcal{H}^s(F)$ and

$$\begin{aligned}
\dim_H E &= \inf\{0 < s < \infty |\ \mathcal{H}^s(E) = 0\} \\
&\le \inf\{0 < s < \infty |\ \mathcal{H}^s(F) = 0\} \\
&= \dim_H F.
\end{aligned}$$

(ii.) Let $f : \mathcal{X} \to \mathcal{Y}$ be a Lipschitz mapping and let $\{U_i\}_{i\in\mathbb{N}}$ be a cover of E with diameter of each U_i at most δ. Then $\{f(U_i)\}_{i\in\mathbb{N}}$ is a cover of $f(E)$ with diameter of each $f(U_i)$ at most $c\delta$. Hence, if $\delta_i = \operatorname{diam}(U_i)$,

$$\mathcal{H}^s_{c\delta}(f(E)) \le \sum_{i=1}^{\infty} (c\delta_i)^s = c^s \sum_{i=1}^{\infty} \delta_i^s.$$

Taking infimum over all covers $\{U_i\}_{i\in\mathbb{N}}$ gives $\mathcal{H}^s_{c\delta}(f(E)) \leq c^s\mathcal{H}^s_\delta(E)$ and taking limit as $\delta \to 0$ yields

$$\mathcal{H}^s(f(E)) \leq c^s\mathcal{H}^s(E).$$

Since $0 < c^s < \infty$ for all s, we get

$$\begin{aligned}\dim_H f(E) &= \inf\{0 < s < \infty |\ \mathcal{H}^s(f(E)) = 0\}\\ &\leq \inf\{0 < s < \infty |\ c^s\mathcal{H}^s(E) = 0\}\\ &= \inf\{0 < s < \infty |\ \mathcal{H}^s(E) = 0\} = \dim_H E,\end{aligned}$$

proving the second statement.

(iii.) Let $\{F_i\}_{i\in\mathbb{N}}$ be a countable family of subsets of $\mathcal{X}$. We have by monotonicity that $\dim_H F_i \leq \dim_H \left(\bigcup_{i=1}^\infty F_i\right)$ for all i and hence $\sup(\dim_H F_i) \leq \dim_H \left(\bigcup_{i=1}^\infty F_i\right)$. On the other hand, since $\mathcal{H}^s\left(\bigcup_{i=1}^\infty F_i\right) \leq \sum_{i=1}^\infty \mathcal{H}^s(F_i)$, we have

$$\begin{aligned}\dim_H\left(\bigcup_{i=1}^\infty F_i\right) &= \inf\{0 < s < \infty \mid \mathcal{H}^s\left(\bigcup_{i=1}^\infty F_i\right) = 0\}\\ &\leq \inf\{0 < s < \infty \mid \sum_{i=1}^\infty \mathcal{H}^s(F_i) = 0\}\\ &= \inf\{0 < s < \infty \mid \mathcal{H}^s(F_i) = 0 \text{ for every } i\}\\ &= \sup(\dim_H F_i).\end{aligned}$$

□

When we introduced the definition of box dimension, we remarked that it is not countably stable. While this is true, we could easily modify the definition to obtain the property. This transforms the definition to what is often called the upper and lower modified box dimensions.

Definition 2.6. Let A be a subset of a metric space $(\mathcal{X}, d)$. The *upper and lower modified box dimensions* of A are defined by

$$\begin{aligned}\overline{\dim}_{MB}A &= \inf\{\sup_n \overline{\dim}_B A_n \mid A \subset \bigcup_{n=1}^\infty A_n\},\\ \underline{\dim}_{MB}(A) &= \inf\{\sup_n \underline{\dim}_B A_n \mid A \subset \bigcup_{n=1}^\infty A_n\}\end{aligned}$$

We omitted this remark in the beginning because it turns out that the upper modified box dimension actually coincides with our next definition, the packing dimension. It is defined by first constructing the packing measure, and this construction proceeds much like the one of Hausdorff measure; instead of finding covers for the set, we are now packing balls inside it. Due to this change, we also have to take an additional step in the definition.

Definition 2.7. Let A be a subset of a metric space $(\mathcal{X}, d)$. For every $0 \leq s < \infty$ and $\delta > 0$, define

$$P^s_\delta(A) = \sup \left\{ \sum_{i \in I} \operatorname{diam}(B_i)^s \;\middle|\; \begin{array}{l} \{B_i\}_{i \in I} \text{ is a countable family of pairwise} \\ \text{disjoint closed balls with diameter at} \\ \text{most } \delta \text{ and centers in } A \end{array} \right\}.$$

We define the *s-dimensional packing pre-measure* of A as

$$P^s_0(A) = \lim_{\delta \to 0} P^s_\delta(A).$$

The *s-dimensional packing measure* is then defined by

$$P^s(A) = \inf\{\sum_{i=1}^{\infty} P^s_0(A_i) \mid A \subset \bigcup_{i=1}^{\infty} A_i\}.$$

As when constructing the Hausdorff measure, we would like to define the packing measure as P^s_0. However, this set function is not yet a measure; it is missing countable additivity. This can be seen by considering countable dense sets as in the case of box dimension. We remark that the packing measure is also a Borel measure.

The packing dimension is then defined similarly to Hausdorff dimension; the existence of such a critical value s follows from taking the limit as δ tends to 0 in the definition of P^s_0.

Definition 2.8. The *packing dimension* of a set A is

$$\dim_P(A) = \sup\{0 \leq s < \infty \mid P^s(A) = \infty\} = \inf\{0 \leq s < \infty \mid P^s(A) = 0\}.$$

It is quite interesting that this measure-theoretical definition leads to the modification of the box dimension that is built in a completely different way, without using any measure on the background. We only present the connection in $\mathbb{R}^n$; in a general metric space it is not necessarily true (or would require the use of radii instead of diameters in the definition of P^s), since the proof makes use of the equality $\operatorname{diam}(B(x, r)) = 2r$ which is not true in general case.

Theorem 2.9. *If $F \subset \mathbb{R}^n$, then*

$$\dim_P(A) = \overline{\dim}_{MB}(A).$$

Proof. See [4], Proposition 3.8. $\square$

We will adapt the Hausdorff dimension as the main notion and in the future will refer to it just as the dimension of A. The subscript from the notation is also omitted, the dimension of A being denoted by $\dim A$.

2.2 Estimating the dimension

One of the disadvantages of Hausdorff dimension is the fact that it can be, and often is, difficult to calculate. While an upper, finite bound for the s-dimensional Hausdorff measure of a set is sometimes found with relative ease, it is the lower bound that may prove very non-trivial. An exception to this is brought by so-called *self-similar* sets that inhibit extremely regular structure, provided the self-similar parts that the set consists of, referred to as cylinders, are sufficiently separated from each other. We return to discuss self-similar sets in Section 3.

The following lemma may aid us in finding a countable cover for a set, using an arbitrarily large family of balls of bounded radii.

Lemma 2.10. *Let $\mathcal{A}$ be a family of open balls contained in a bounded region of a separable metric space $\mathcal{X}$. There exists a countable disjoint subcollection $\mathcal{B}$ of $\mathcal{A}$ such that*

$$\bigcup_{B\in\mathcal{A}} B \subset \bigcup_{B\in\mathcal{B}} \tilde{B}, \tag{2.1}$$

where $\tilde{B}$ is a ball concentric with B and of five times the radius.

Proof. Let $\varepsilon > 0$ and $R = \sup\{\mathrm{rad}(B_i) \mid B_i \in \mathcal{A}\}$. Since the balls in $\mathcal{A}$ are contained in a bounded region, R is finite. Denote by $\mathcal{A}_n$ the subcollection of $\mathcal{A}$ consisting of balls with radius in the interval $(2^{-n-1}R, 2^{-n}R]$. We construct the collection $\mathcal{B}$ inductively.

First, set $\mathcal{C}_0 = \mathcal{A}_0$ and define $\mathcal{B}_0$ as the maximal disjoint subcollection of $\mathcal{C}_0$. Suppose then that the collections $\mathcal{B}_0, \ldots, \mathcal{B}_{k-1}$ have been chosen. Define

$$\mathcal{C}_k = \left\{ B \in \mathcal{A}_k \mid B \cap C = \emptyset \text{ for all } C \in \bigcup_{i=0}^{k-1} \mathcal{B}_i \right\}$$

and let $\mathcal{B}_k$ be the maximal disjoint subcollection of $\mathcal{C}_k$. Finally, define

$$\mathcal{B} = \bigcup_{n=0}^{\infty} \mathcal{B}_n.$$

By definition, $\mathcal{B}$ is a disjoint collection and since every $\mathcal{B}_k$ is finite, $\mathcal{B}$ is countable. Let B be a ball in $\mathcal{A}$ and n be such that $B \in \mathcal{A}_n$. Then, either $B \in \mathcal{B}_n$, or B intersects a ball either in $\mathcal{B}_n$ or in $\bigcup_{i=0}^{n-1} \mathcal{B}_i$. In any case, B intersects a ball $B' \in \bigcup_{i=0}^{n} \mathcal{B}_i$, and such a ball has radius $\mathrm{rad}(B') > 2^{-n-1}R$. Since $\mathrm{rad}(B) \leq 2^{-n}R$, it is less than two times that of B' and therefore $B \subset \tilde{B}'$. □

Sometimes we may receive useful upper bounds for the dimension of a set using the following inequalities.

Proposition 2.11. *For any subset E of $\mathbb{R}^n$,*

$$\dim E \leq \dim_P E \leq \overline{\dim}_B E.$$

Proof. To prove the first inequality, we show that $\mathcal{H}^s(E) \leq P_0^s(E)$ for every $0 \leq s < \infty$. Then, if $\{E_i\}_{i\in\mathbb{N}}$ is an arbitrary countable cover of E, we have $\mathcal{H}^s(E) \leq \sum_{i=1}^{\infty} \mathcal{H}^s(E_i) \leq \sum_{i=1}^{\infty} P_0^s(E_i)$ and taking infimum over all such covers yields $\mathcal{H}^s(E) \leq P^s(E)$. By the definitions of $\dim_H$ and $\dim_P$, this gives then the first inequality.

Indeed, let $\varepsilon > 0$ and let $\delta > 0$ be such that $P_\delta^s(E) \leq P_0^s(E) + \varepsilon$. We may assume that $P_0^s(E)$ is finite, since otherwise the inequality is trivial. Let $\{B_i\}_{i\in\mathbb{N}}$ be a collection of balls with centers in E, $\mathrm{diam}(B_i) < \delta$ for each i and

$$\sum_{i=1}^{\infty} (\mathrm{diam}(B_i))^s \leq P_\delta^s(E) \leq \sum_{i=1}^{\infty} (\mathrm{diam}(B_i))^s + \varepsilon.$$

Since we assumed that $P_0^s(E) < \infty$, there exists an integer k such that

$$\sum_{i=k+1}^{\infty} (\mathrm{diam}(B_i))^s \leq \varepsilon.$$

Now we apply Lemma 2.10 to the collection

$$\mathcal{A} = \{B(x,r) \mid x \in E, r < \delta/10, B(x,r) \cap \bigcup_{i=1}^{k} B_i = \emptyset\}$$

to obtain a disjoint, countable subcollection $\{B_i'\}_{i\in\mathbb{N}}$ such that $E \setminus \bigcup_{i=1}^{k} B_i \subset \bigcup_{B\in\mathcal{A}} B \subset \bigcup_{i\in\mathbb{N}} 5B_i'$. By definition of P_δ^s we have

$$\sum_{i=1}^{k} (\mathrm{diam}(B_i))^s + \sum_{i=1}^{\infty} (\mathrm{diam}(B_i'))^s \leq P_\delta^s(E) \leq \sum_{i=1}^{\infty} (\mathrm{diam}(B_i))^s + \varepsilon$$

and therefore $\sum_{i=1}^{\infty}(\mathrm{diam}(B_i'))^s \leq 2\varepsilon$. Summing this up, since $\bigcup_{i=1}^{k} B_i \cup \bigcup_{i=1}^{\infty} 5B_i'$ covers E, we have

$$\begin{aligned}\mathcal{H}_\delta^s(E) &\leq \sum_{i=1}^{k}(\mathrm{diam}(B_i))^s + 5^s\sum_{i=1}^{\infty}(\mathrm{diam}(B_i'))^s\\ &\leq P_\delta^s(E) + 5^s 2\varepsilon\\ &\leq P_0^s(E) + (1+5^s 2)\varepsilon.\end{aligned}$$

Letting $\delta \to 0$ and $\varepsilon \to 0$ finishes the proof of $\dim_H E \leq \dim_P E$.

For the second inequality, let $0 < t < s < \dim_P E$. Then $P_\delta^s(E) = \infty$ for every $\delta > 0$ ($P_\delta^s(E)$ decreases as δ tends to 0) and so for every δ, we find disjoint balls B_i with centers in E and diameters at most δ such that $\sum_{i=1}^{\infty}(\mathrm{diam}(B_i))^s > 1$. Let n_k denote the number of balls for which $2^{-k-1} < \mathrm{diam}(B_i) \leq 2^k$ for every k. Then, using these bounds, we may estimate diameters of the balls in the sum above to obtain $\sum_{k=1}^{\infty} n_k 2^{-ks} > 1$.

Now, the number n_k cannot be too small for every k; namely, there exists a k such that $n_k > 2^{kt}(1-2^{t-s})$. This can be seen by assuming otherwise and then estimating the sum above to arrive at a contradiction with the strict lower bound of 1. By the definition of n_k, each of these balls contains another ball of diameter 2^{-k-1}, so $N_{2^{-k-1}}(E) \geq n_k$. Therefore

$$\begin{aligned}\overline{\dim}_B E &= \limsup_{\varepsilon\to 0}\frac{\log N_\varepsilon(E)}{\log(1/\varepsilon)}\\ &\geq \limsup_{k\to\infty}\frac{\log n_k}{\log 2^{k+1}}\\ &\geq \limsup_{k\to\infty}\frac{kt\log 2 + \log(1-2^{t-s})}{k+1}\\ &= t.\end{aligned}$$

Since $t < s < \dim_P E$ were arbitrary, this completes the proof of the second inequality. □

If we are able to find a mass distribution such that the mass it gives to subsets of small diameter is not too large, this mass distribution may be used to bound the Hausdorff measure of any set from below.

Theorem 2.12 (Mass distribution principle)**.** *Let F be a subset of $\mathcal{X}$. Suppose μ^* is an outer measure on $\mathcal{X}$ with $\mu^*(\mathcal{X}) < \infty$ and suppose there exist positive and finite real numbers s, δ and c such that for all $U \subset \mathcal{X}$ with* $\mathrm{diam}(U) \leq \delta$,

$$\mu^*(U) \leq c \cdot \mathrm{diam}(U)^s.$$

Then $\mathcal{H}^s(F) \geq \dfrac{\mu^*(F)}{c}$.

Proof. Let $\{U_i\}_{i\in\mathbb{N}}$ be a cover of F with $\operatorname{diam}(U_{i\in\mathbb{N}}) \leq \delta$ for every i and let μ^* be an outer measure satisfying the hypothesis. Then

$$\mu^*(F) \leq \mu^*\left(\bigcup_{i\in\mathbb{N}} U_i\right) \leq \sum_{i\in\mathbb{N}} \mu^*(U_i) \leq c \sum_{i\in\mathbb{N}} \operatorname{diam}(U_i)^s.$$

For any $\varepsilon > 0$, choose a cover $\{U_i\}_{i\in\mathbb{N}}$ of F such that $\operatorname{diam}(U_i) \leq \delta$ and

$$\mathcal{H}^s_\delta(F) \geq \sum_{i\in\mathbb{N}} \operatorname{diam}(U_i)^s - \varepsilon \geq \frac{\mu^*(F)}{c} - \varepsilon.$$

Since this holds for any small $\delta > 0$, we obtain the statement by letting $\varepsilon \to 0$. $\square$

Remark 2.13. We may also apply the mass distribution principle to a Borel mass distribution in the place of μ^*, since all the sets U can be replaced by their closures without affecting their diameters.

Using the following construction, we may sometimes obtain an outer measure to use in applying the mass distribution principle.

Theorem 2.14. *Let $\mathcal{X}$ be a set and $\mathcal{E}_0 = \{\mathcal{X}\}$. For each $n \in \mathbb{N}$, let $\mathcal{E}_n$ be a finite collection of disjoint subsets of $\mathcal{X}$ and let $\mathcal{E} = \bigcup_{n\in\mathbb{N}} \mathcal{E}_n$. Let $\zeta : \mathcal{E} \to [0,\infty]$ be a set mapping and assume the following conditions hold:*

(i.) $\bigcup_{E\in\mathcal{E}_n} = \mathcal{X}$ *for every* $n \in \mathbb{N}$*,*

(ii.) Each $E \in \mathcal{E}_n$ contains finitely many sets of $\mathcal{E}_{n+1}$ and there is exactly one $F \in \mathcal{E}_{n-1}$ such that $E \subset F$,

(iii.)

$$\sum_{E\in\mathcal{E}_1} \zeta(E) = \zeta(\mathcal{X}).$$

Moreover, if $E \in \mathcal{E}_n$ and $E = \bigcup_{n=1}^k E_n$ for $E_1, \ldots, E_k \in \mathcal{E}_{n+1}$, then

$$\sum_{n=1}^{k} \zeta(E_k) = \zeta(E).$$

(iv.) If $E_n \in \mathcal{E}_n$ and $E_n \subset E_{n-1}$ for every n, the set $\bigcap_{n\in\mathbb{N}} E_n$ contains exactly one point of the set $\mathcal{X}$.

Then there is an outer measure μ^* on $\mathcal{X}$ such that $\mu^*(E) = \zeta(E)$ for all $E \in \mathcal{E}$.

Proof. See [13], Theorem 5.6. $\square$

2.3 Dimension of a measure

In the case of measures, we use the notion of dimension to measure how discrete the behaviour of the measure is. An atomic measure, for example, has no "body" and emphasizes only individual points; therefore we would like to say it has no dimension anywhere. On the other hand, the Lebesgue measure on $\mathbb{R}^n$ somehow fills the entire space and we want it to have the "full dimension" n. With this motivation, we define the local dimension in a point x by comparing how the measure of a small ball centered in x relates to the radius of the ball.

Definition 2.15. Let μ be a finite non-zero measure on a locally compact metric space $\mathcal{X}$ and let $x \in \mathcal{X}$. The *upper and lower local dimensions* of μ are defined by

$$\overline{D}(\mu, x) = \limsup_{r\to 0} \frac{\log \mu(B(x,r))}{\log r},$$

$$\underline{D}(\mu, x) = \liminf_{r\to 0} \frac{\log \mu(B(x,r))}{\log r}.$$

If $\overline{D}(\mu, x) = \underline{D}(\mu, x)$ in some point x, we say that μ has *local dimension* in x and denote it by $D(\mu, x) = \overline{D}(\mu, x) = \underline{D}(\mu, x)$. If the local dimension exists and takes the same value α in μ-almost every point, we say that μ is *exact dimensional* and α is its (exact) dimension, denoted by $\dim \mu = \alpha$.

If μ has local dimension s in x, then by the definition $\mu(B(x,r))$ behaves like cr^s for some constant c and all small r. It turns out that if we have this kind of knowledge of the behaviour of μ in a Borel set E of positive measure, we may use it to estimate the s-dimensional Hausdorff- and packing measures, and therefore the respective dimensions, of E. We restrict ourselves to $\mathbb{R}^n$ because in the proof we require the fact that the diameter of a ball is two times the radius.

Lemma 2.16. *Let $E \subset \mathbb{R}^n$ be a Borel set, μ a finite measure and $0 < c < \infty$.*

(i.) *If* $\limsup_{r\to 0} \dfrac{\mu(B(x,r))}{r^s} < c$ *for every* $x \in E$, *then* $\mathcal{H}^s(E) \geq \dfrac{\mu(E)}{c}$

(ii.) *If* $\limsup_{r\to 0} \dfrac{\mu(B(x,r))}{r^s} > c$ *for every* $x \in E$, *then* $\mathcal{H}^s(E) \leq 10^s \dfrac{\mu(\mathbb{R}^n)}{c}$

If μ is a Radon measure, in addition, we also have the following.

(iii.) *If* $\liminf_{r\to 0} \dfrac{\mu(B(x,r))}{r^s} > c$ *for every* $x \in E$, *then* $P^s(E) \geq \dfrac{\mu(E)}{c}$

(iv.) *If* $\liminf_{r\to 0} \dfrac{\mu(B(x,r))}{r^s} < c$ *for every* $x \in E$*, then* $P^s(E) \leq 2^s \dfrac{\mu(E)}{c}$.

The last two statements are included for the sake of completeness; their proofs are omitted because they proceed in same manner as the first two. The proof of the third statement also requires the Vitali covering theorem.

Proof. Let $n \in \mathbb{N}$ and define the set

$$E_n = \{x \in E \mid \mu(B(x,r)) < cr^s \text{ for all } 0 < r \leq 1/n\}.$$

Clearly $(E_n)_{n\in\mathbb{N}}$ forms an increasing sequence of sets. By the hypothesis of the first statement, for every x there exists such an n that $\mu(B(x,r)) < cr^s$ whenever $0 < r \leq 1/n$; hence $E = \bigcup_{n\in\mathbb{N}} E_n$ and $\mu(E) = \lim_{n\to\infty} \mu(E_n)$.

Fix n, let $0 < \delta < 1/n$ and let $\{U_i\}_{i\in\mathbb{N}}$ be a cover of E such that $\operatorname{diam}(U_i) \leq \delta$ for all i. Since $E_n \subset E$, $\{U_i\}_{i\in\mathbb{N}}$ is also a cover of E_n. For any U_i that intersects E_n, if x_i is a point in this intersection,

$$\mu(U_i) \leq \mu(B(x_i, \operatorname{diam}(U_i))) < c \cdot \operatorname{diam}(U_i)^s$$

and hence

$$\mu(E_n) \leq \sum_{i\in\mathbb{N} \,\mid\, U_i\cap E_n \neq \emptyset} \mu(U_i) \leq c \sum_{i\in\mathbb{N}} \operatorname{diam}(U_i)^s.$$

Since $\{U_i\}_{i\in\mathbb{N}}$ is an arbitrary cover of E with $\operatorname{diam}(U_i) \leq \delta$, we have $\mu(E_n) \leq c\mathcal{H}^s_\delta(E)$. Letting first $\delta \to 0$ and then $n \to \infty$, we obtain $\mu(E) \leq c\mathcal{H}^s(E)$ which is the first statement.

Moving on to the second one, assume first that E is bounded. Let $\delta > 0$ and define the family of open balls

$$B(\delta) = \{B(x,r) \mid x \in E, \mu(B(x,r)) > cr^s \text{and } 0 < r \leq \delta\}.$$

By the hypothesis, for every x there exists an $r \leq \delta$ such that $\mu(B(x,r)) > cr^s$. Hence $E \subset \bigcup_{B\in B(\delta)} B$. Applying Lemma 2.10 to the collection $B(\delta)$, we obtain a disjoint, countable subcollection $B(\delta)' \subset B(\delta)$ such that $\bigcup_{B\in B(\delta)} B \subset \bigcup_{B\in B(\delta)'} 5B$, where $5B$ is the ball concentric with B and of five times the radius. Thus, $\{5B\}_{B\in B(\delta)'}$ is a cover of E with the diameter of each ball at most 10δ and

$$\begin{aligned}
\mathcal{H}^s_{10\delta}(E) &\leq \sum_{B\in B(\delta)'} \operatorname{diam}(5B)^s = 5^s \sum_{B\in B(\delta)'} \operatorname{diam}(B)^s \\
&\leq \frac{10^s}{c} \sum_{B\in B(\delta)'} \mu(B) = \frac{10^s}{c}\mu\left(\bigcup_{B\in B(\delta)'} B\right) \leq \frac{10^s}{c}\mu(\mathbb{R}^n).
\end{aligned}$$

By letting $\delta \to 0$, we obtain the second statement for a bounded E. If E is unbounded and $\mathcal{H}^s(E) > \frac{10^s}{c}\mu(\mathbb{R}^n)$, we have $\mathcal{H}^s(E \cap \overline{B}(0,r)) > \frac{10^s}{c}\mu(\mathbb{R}^n)$ for a large enough r. Since the set $E \cap \overline{B}(0,r)$ is bounded, this is a contradiction. □

Corollary 2.17. *Let E be a Borel set of $\mathbb{R}^n$, $F \subset E$ and let μ be a finite measure.*

(i.) *If $\underline{D}(\mu,x) \geq s$ for all $x \in F$ and $\mu(F) > 0$, then $\dim E \geq s$*

(ii.) *If $\underline{D}(\mu,x) \leq s$ for all $x \in E$, then $\dim E \leq s$*

Additionally, if μ is a Radon measure,

(iii.) *If $\overline{D}(\mu,x) \geq s$ for all $x \in F$ and $\mu(F) > 0$, then $\dim_P E \geq s$*

(iv.) *If $\overline{D}(\mu,x) \leq s$ for all $x \in E$, then $\dim_P E \leq s$.*

Proof. By the hypothesis of the first statement,

$$\underline{D}(\mu,x) = \liminf_{r\to 0} \frac{\log \mu(B(x,r))}{\log r} \geq s$$

for every $x \in F$. Let $t < s$. Then for all small enough r, $\log \mu(B(x,r))/\log r > t$ and so for all $x \in F$,

$$\begin{aligned}\limsup_{r\to 0} \frac{\mu(B(x,r))}{r^t} &\leq \limsup_{r\to 0} \mu(B(x,r)) r^{\frac{-\log \mu(B(x,r))}{\log r}} \\ &= 1.\end{aligned}$$

By Lemma 2.16 (i.), $\mathcal{H}^t(F) \geq \mu(F) > 0$. Since $t < s$ was arbitrary, $\dim E \geq \dim F \geq s$.

Assume then that $\underline{D}(\mu,x) \leq s$ for all $x \in E$. Let $t > s$. By the hypothesis, there exists a sequence $(r_k)_k$ that tends to 0 such that $\log \mu(B(x,r_k))/\log r_k < t$ for all $k \in \mathbb{N}$. Therefore,

$$\begin{aligned}\limsup_{r\to 0} \frac{\mu(B(x,r))}{r^t} &\geq \limsup_{k\to\infty} \frac{\mu(B(x,r_k))}{r_k^t} \\ &\geq \limsup_{k\to\infty} \mu(B(x,r_k)) r_k^{\frac{-\log \mu(B(x,r_k))}{\log r_k}} \\ &= 1\end{aligned}$$

By Lemma 2.16 (ii.), $\mathcal{H}^t(E) < 10^s \mu(\mathbb{R}^n) < \infty$. Since $t > s$ was arbitrary, we have $\dim E \leq s$.

The proofs of statements (iii.) and (iv.) proceed identically, using Lemma 2.16 (iii.) and (iv.) and are therefore omitted. □

An immediate conclusion is that if a measure on a Euclidean space is exact dimensional, its dimension equals the dimension of its support.

Corollary 2.18. *Let μ be an exact dimensional measure on $\mathbb{R}^n$ and let E be its support. Then*

$$\dim \mu = \dim E.$$

Proof. Since $\dim \mu = \underline{D}(\mu, x)$ for all $x \in E$, the result is given by Corollary 2.17. $\square$

Since we now have seen that the lower local dimension on a set E somehow corresponds to the Hausdorff dimension of E and the upper local dimension corresponds to the packing dimension E, it is natural to introduce Hausdorff- and packing dimensions for measures as well.

Definition 2.19. Let μ be a finite measure on $\mathcal{X}$. The *upper and lower Hausdorff dimensions* of μ are defined by

$$\overline{\dim}_H \mu = \operatorname*{ess\,sup}_{x \sim \mu} \underline{D}(\mu, x),$$

$$\underline{\dim}_H \mu = \operatorname*{ess\,inf}_{x \sim \mu} \underline{D}(\mu, x),$$

and the *upper and lower packing dimensions* are defined by

$$\overline{\dim}_P \mu = \operatorname*{ess\,sup}_{x \sim \mu} \overline{D}(\mu, x),$$

$$\underline{\dim}_P \mu = \operatorname*{ess\,inf}_{x \sim \mu} \overline{D}(\mu, x).$$

The notation $\operatorname{ess\,sup}_{x \sim \mu}$ stands for essential supremum. For a real-valued Borel function f defined on the set $\mathcal{X}$, it is defined by

$$\operatorname*{ess\,sup}_{x \sim \mu} f = \inf\{a \in \mathbb{R} \mid \mu(f^{-1}(a, \infty)) = 0\},$$

i.e. function f can exceed $\operatorname{ess\,sup}_{x \sim \mu} f$ only on a null set with respect to μ. Since $x \mapsto \log \mu(B(x, r)) / \log r$ is a Borel function for every r, so are its pointwise limits $\underline{D}(\mu, x)$ and $\overline{D}(\mu, x)$. Similarly, essential infimum is defined by

$$\operatorname*{ess\,inf}_{x \sim \mu} f = \sup\{a \in \mathbb{R} \mid \mu(f^{-1}(-\infty, a)) = 0\}.$$

We point out the following trivial inequalities regarding the definitions:

$$\begin{aligned} \underline{\dim}_H \mu &\leq \overline{\dim}_H \mu, \qquad (2.2)\\ \underline{\dim}_P \mu &\leq \overline{\dim}_P \mu,\\ \underline{\dim}_H \mu &\leq \underline{\dim}_P \mu,\\ \overline{\dim}_H \mu &\leq \overline{\dim}_P \mu. \end{aligned}$$

The lower Hausdorff- and packing dimensions are completely characterized by the dimensions of the Borel sets the measure has mass on.

Theorem 2.20. *For a finite measure μ,*

$$\underline{\dim}_H \mu = \inf\{\dim E \mid E \textit{ is Borel and } \mu(E) > 0\}.$$

Additionally, if μ is a Radon measure,

$$\underline{\dim}_P \mu = \inf\{\dim_P E \mid E \textit{ is Borel and } \mu(E) > 0\}.$$

Proof. Let E be a Borel set with $\mu(E) > 0$ and let $s < \underline{\dim}_H \mu$. By definition, $\operatorname{ess\,inf}_{x \sim \mu} \underline{D}(\mu, x) = \underline{\dim}_H \mu$ and therefore there exists $F \subset E$ such that $\underline{D}(\mu, x) \geq s$ for every x in F and $\mu(F) = \mu(E) > 0$. Corollary 2.17 now asserts that $\dim E \geq s$ and hence

$$\underline{\dim}_H \mu \leq \inf\{\dim E \mid E \text{ is Borel and } \mu(E) > 0\}.$$

On the other hand, for every $\varepsilon > 0$ there is a set F with positive μ-measure such that $\underline{D}(\mu, x) < \underline{\dim}_H \mu + \varepsilon$ for every $x \in F$. By Corollary 2.17, $\dim F \leq \underline{\dim}_H \mu + \varepsilon$ and therefore, since ε is arbitrary,

$$\underline{\dim}_H \mu \geq \inf\{\dim E \mid E \text{ is Borel and } \mu(E) > 0\}.$$

The proofs of the statements concerning $\underline{\dim}_P \mu$ proceed identically. □

Using the characterization given by Theorem 2.20, we can derive properties for $\underline{\dim}_H$ that correspond to those of the Hausdorff dimension of a set, proposed in Proposition 2.5.

Proposition 2.21. *Let μ, ν, μ_i be finite measures on $\mathcal{X}$ for all $i \in \mathbb{N}$. The lower Hausdorff dimension satisfies the following:*

(i.) If μ is absolutely continuous with respect to ν, then $\underline{\dim}_H \mu \geq \underline{\dim}_H \nu$

(ii.) If $f : \mathcal{X} \to \mathcal{Y}$ is a Lipschitz mapping, then $\underline{\dim}_H f\mu \leq \underline{\dim}_H \mu$

(iii.) If $\sum_{i=1}^{\infty} \mu_i$ is a finite measure, then $\underline{\dim}_H \sum_{i=1}^{\infty} \mu_i = \inf_{i \in \mathbb{N}}\{\underline{\dim}_H \mu_i\}$

Proof. Let μ, ν, μ_i be finite measures on $\mathcal{X}$.

(i.) Since μ is absolutely continuous with respect to ν, $\mu(E) = 0$ for every E for which $\nu(E) = 0$. Hence $\{E \subset \mathcal{X} \mid \mu(E) > 0\} \subset \{E \subset \mathcal{X} \mid \nu(E) > 0\}$ and

$$\begin{aligned}\underline{\dim}_H \mu &= \inf\{\dim E \mid E \text{ is Borel and } \mu(E) > 0\} \\ &\geq \inf\{\dim E \mid E \text{ is Borel and } \nu(E) > 0\} = \underline{\dim}_H \nu.\end{aligned}$$

(ii.) Let $f : \mathcal{X} \to \mathcal{Y}$ be a Lipschitz mapping. Observe that due to continuity of such mappings, f is a Borel function, and therefore $f\mu$ is a finite Borel measure on $\mathcal{Y}$. By Proposition 2.5 (ii.), dimension of a set is never less than the dimension of its preimage through a Lipschitz mapping. Hence

$$\begin{aligned}\underline{\dim}_H f\mu &= \inf\{\dim F \mid F \subset \mathcal{Y} \text{ is Borel and } \mu(f^{-1}(F)) > 0\}\\ &\geq \inf\{\dim f^{-1}(F) \mid F \subset \mathcal{Y} \text{ is Borel and } \mu(f^{-1}(F)) > 0\}\\ &\geq \inf\{\dim E \mid E \subset \mathcal{X} \text{ is Borel and } \mu(E) > 0\}\\ &= \underline{\dim}_H \mu.\end{aligned}$$

(iii.) Suppose that $\sum_{i=1}^{\infty} \mu_i$ is finite. Since $\mu_i(E)$ is always non-negative,

$$\begin{aligned}\{E \subset \mathcal{X} \mid \sum_{i=1}^{\infty} \mu_i(E) > 0\} &= \{E \subset \mathcal{X} \mid \mu_i(E) > 0 \text{ for some } i \in \mathbb{N}\}\\ &= \bigcup_{i=1}^{\infty} \{E \subset \mathcal{X} \mid \mu_i(E) > 0\ \}\end{aligned}$$

and hence

$$\begin{aligned}\underline{\dim}_H \sum_{i=1}^{\infty} \mu_i &= \inf\{\dim E \mid E \text{ is Borel and } \sum_{i=1}^{\infty} \mu_i(E) > 0\}\\ &= \inf \bigcup_{i=1}^{\infty} \{\dim E \mid E \text{ is Borel and } \mu_i(E) > 0\}\\ &= \inf_{i \in \mathbb{N}} \{\underline{\dim}_H \mu_i\}.\end{aligned}$$

□

The last definition for dimension we introduce is given as the limit of the normalized entropy of the measure.

Definition 2.22. Let μ be a probability measure on $\mathbb{R}^d$. The *upper and lower entropy dimensions* of μ are defined as

$$\begin{aligned}\overline{\dim}_e \mu &= \limsup_{n\to\infty} \frac{1}{n} H(\mu, \mathcal{D}_n),\\ \underline{\dim}_e \mu &= \liminf_{n\to\infty} \frac{1}{n} H(\mu, \mathcal{D}_n),\end{aligned}$$

If $\overline{\dim}_e \mu = \underline{\dim}_e \mu = a$, we call a the *entropy dimension* of μ and denote it by $\dim_e \mu$.

A useful result is that for an exact dimensional measure, all the definitions for its dimension coincide. When dealing with such measures, we may always pick the definition that is the most convenient.

Theorem 2.23. *A measure μ is exact dimensional if and only if $\underline{\dim}_H\mu = \overline{\dim}_P\mu$.*

Proof. By definition, μ is exact dimensional if and only if $\underline{D}(\mu, x) = \overline{D}(\mu, x) = \alpha$ for μ-almost all x. Assuming this holds, we have

$$\underline{\dim}_H\mu = \operatorname*{ess\,inf}_{x\sim\mu} \underline{D}(\mu, x) = \alpha = \operatorname*{ess\,sup}_{x\sim\mu} \overline{D}(\mu, x) = \overline{\dim}_P\mu.$$

On the other hand, if $\underline{\dim}_H\mu = \overline{\dim}_P\mu$, all the inequalities in (2.2) hold with equality and

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

Hence $\underline{D}(\mu, x)$ is constant μ-almost everywhere. Similarly,

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

Therefore $\overline{D}(\mu, x)$ is also constant and, since $\operatorname{ess\,inf}_{x\sim\mu} \underline{D}(\mu, x) = \operatorname{ess\,inf}_{x\sim\mu} \overline{D}(\mu, x)$, equals $\underline{D}(\mu, x)$ μ-almost everywhere. $\square$

Lemma 2.24. *If μ is an exact dimensional probability measure on $\mathbb{R}^d$ with bounded support, the entropy dimension exists and $\dim\mu = \dim_e\mu$.*

Proof ([5], p. 193) . By Theorem 2.23, for an exact dimensional μ, $\underline{\dim}_H\mu = \overline{\dim}_P\mu = \alpha$. Denote by $\mathcal{D}_{n-\log_2\sqrt{d}}(x)$ the unique atom of $\mathcal{D}_{n-\log_2\sqrt{d}}$ that contains x and by $B_n(x)$ the co-centric ball of radius $\operatorname{diam}(\mathcal{D}_{n-\log_2\sqrt{d}}(x)) = 2^{-n}$. Since $x \mapsto \mu(\mathcal{D}_n(x))$ is constant inside each atom of $\mathcal{D}_n$ and $\mathcal{D}_n(x) \subset$

$B_n(x)$, we have by Fatou's lemma that

$$\begin{aligned}
\underline{\dim}_e \mu &= \liminf_{n\to\infty} \frac{H(\mu, \mathcal{D}_n)}{n} \\
&= \liminf_{n\to\infty} \frac{H(\mu, \mathcal{D}_{n-\log_2\sqrt{d}}) + O_d(\log_2\sqrt{d})}{n} \\
&\geq \liminf_{n\to\infty} \int_{\mathbb{R}^d} \frac{-\log \mu(\mathcal{D}_{n-\log_2\sqrt{d}}(x))}{\log 2^n}\ d\mu(x) \\
&\geq \int_{\mathbb{R}^d} \liminf_{n\to\infty} \frac{-\log \mu(\mathcal{D}_{n-\log_2\sqrt{d}}(x))}{\log 2^n}\ d\mu(x) \\
&\geq \int_{\mathbb{R}^d} \liminf_{n\to\infty} \frac{-\log \mu(B_n(x))}{\log 2^n}\ d\mu(x) \\
&= \int_{\mathbb{R}^d} \underline{D}(\mu, x)\ d\mu(x) \\
&\geq \int_{\mathrm{supp}(\mu)} \operatorname*{ess\,inf}_{x\sim\mu} \underline{D}(\mu, x)\ d\mu(x) \\
&\geq \underline{\dim}_H \mu.
\end{aligned}$$

For all integers $N \geq 1$, define the set

$$A_N = \left\{ x \in \mathbb{R}^d \mid \frac{-\log \mu(B(x, 2^{-n}))}{n} \leq d+1 \text{ for all } n > N \right\}.$$

Using Theorem 2.20 and the fact that μ is exact dimensional, we have

$$\limsup_{n\to\infty} \frac{-\log \mu(B_n(x))}{\log 2^n} = \overline{D}(\mu, x) = \underline{\dim}_H \mu \leq d$$

for μ-almost every x. Hence $\mu\left(\bigcup_{N=1}^{\infty} A_N\right) = 1$ and since $A_1 \subset A_2 \subset \ldots$, $\lim_{N\to\infty} \mu(A_N) = 1$. Let $\varepsilon > 0$ be arbitrary and choose N so that $\mu(A_N) > 1-\varepsilon$. On the set A_N, the functions $x \mapsto \frac{-\log \mu(B(x,2^{-n}))}{n}$ are uniformly bounded by $d+1$ for every $n > N$. Alter the definition of $B_n(x)$ so that it is the ball concentric with $\mathcal{D}_{n-1}(x)$ and of radius 2^{-n}. Now $B_n(x) \subset \mathcal{D}_{n-1}(x)$ and, using

Fatou's lemma and Lemma 1.5 (ii.), we have

$$\begin{aligned}
\overline{\dim}_e\mu &= \limsup_{n\to\infty} \frac{H(\mu,\mathcal{D}_n)}{n} \\
&= \limsup_{n\to\infty} \frac{H(\mu,\mathcal{D}_{n-1}) + O_d(1)}{n} \\
&\le \limsup_{n\to\infty} \int_{\mathbb{R}^d} \frac{-\log\mu(\mathcal{D}_{n-1}(x))}{\log 2^n}\, d\mu(x) \\
&\le \limsup_{n\to\infty} \int_{\mathbb{R}^d} \frac{-\log\mu(B_n(x))}{\log 2^n}\, d\mu(x) \\
&\le \overline{\dim}_P\mu + \limsup_{n\to\infty} \int_{\mathbb{R}^d\setminus A_N} \frac{-\log\mu(B_n(x))}{\log 2^n}\, d\mu(x).
\end{aligned}$$

It remains to show that the integral $I := \int_{\mathbb{R}^d\setminus A_N} \frac{-\log\mu(B_n(x))}{\log 2^n}\, d\mu(x)$ vanishes as n tends to infinity. For all N, m and $n > N$, define the sets

$$\begin{aligned}
U^m_{N,n} &= (\mathbb{R}^d \setminus A_N) \cap \left\{ x\in\mathbb{R}^d \mid \frac{-\log\mu(B(x,2^{-n}))}{\log 2^n} \le m+1 \right\}, \\
V^m_{N,n} &= (\mathbb{R}^d \setminus A_N) \cap \left\{ x\in\mathbb{R}^d \mid m < \frac{-\log\mu(B(x,2^{-n}))}{\log 2^n} \le m+1 \right\}
\end{aligned}$$

so that for any $n > N$, we have

$$\begin{aligned}
I &= \int_{U^d_{N,n}} \frac{-\log\mu(B_n(x))}{\log 2^n}\, d\mu + \sum_{m=d+1}^{\infty} \int_{V^m_{N,n}} \frac{-\log\mu(B_n(x))}{\log 2^n}\, d\mu \\
&\le (d+1)\varepsilon + \sum_{m=d+1}^{\infty} \int_{V^m_{N,n}} \frac{-\log\mu(B_n(x))}{\log 2^n}\, d\mu \\
&= (d+1)\varepsilon + \sum_{m=d+1}^{\infty} \sum_{D\in\mathcal{D}_n} \int_{D\cap V^m_{N,n}} \frac{-\log\mu(B_n(x))}{\log 2^n}\, d\mu.
\end{aligned}$$

Observe that if $\frac{-\log\mu(B_n(x))}{\log 2^n} > m$, then $\mu(B_n(x)) < 2^{-nm}$. Hence

$$\sum_{D\in\mathcal{D}_n} \int_{D\cap V^m_{N,n}} \frac{-\log\mu(B_n(x))}{\log 2^n}\, d\mu \le (m+1)2^{-mn}O(2^{nd}),$$

where $O(2^{nd}) = O_{\mathrm{diam(supp}(\mu))}(2^{nd})$ is the number of atoms of $\mathcal{D}_n$ inside the bounded support of μ. Moreover,

$$\sum_{m=d+1}^{\infty} (m+1)O(2^{-3mn+dn}) \le O_d(2^{-n}).$$

It follows that for any $\varepsilon > 0$, $\overline{\dim}_e\mu \le \overline{\dim}_P\mu + (d+1)\varepsilon$. Letting $\varepsilon \to 0$ finishes the proof.

□

By Lemma 1.7, we could write the entropy dimension of μ as the limit $\lim_{n\to\infty} \mathbb{E}_{0\le i<n}(H_m(\mu^{x,i}))$ for all integers m. However, counterexamples show that the convergence of the expected values does not imply that the entropy dimensions of the components $\mu^{x,i}$ would concentrate around that of μ. We introduce the following stronger notion.

Definition 2.25. A measure $\mu \in \mathcal{P}(\mathbb{R})$ has *uniform entropy dimension* α if for every $\varepsilon > 0$ and for large enough m,

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

3 Self-similar sets and measures

Many fractals possess a structure that contains infinitely many smaller-scale copies of itself. Examples of sets such as these are the $\frac{1}{3}$-Cantor set on the line which is the union of two smaller-scale copies of itself, and the Sierpinski gasket on the plane which is the union of three copies of itself. Sets such as these may be categorized as *self-similar*.

From now on, we are operating only in Euclidean spaces and begin by defining a mapping that transforms a set into a smaller-scale copy of itself.

Definition 3.1. Let D be a closed subset of $\mathbb{R}^n$. A mapping $\phi : D \to D$ is a *contraction*, if there exists a $0 < c < 1$ such that $|\phi(x) - \phi(y)| \le c|x - y|$ for all x and y in D. This c is called the *(contraction) ratio* of ϕ. If the equality holds, ϕ is called a *similarity*.

Remark 3.2. (i.) A contraction is continuous on D.

(ii.) In Section 4, we are primarily interested in self-similar sets on the line. Regarding this, it is useful to note that a similarity $\phi : \mathbb{R} \to \mathbb{R}$ may be written as $\phi(x) = \pm rx + a$, where r is the contraction ratio of ϕ and $a = \phi(0)$.

A similarity mapping gets its name from its property of transforming sets into geometrically similar ones; it scales them with a contraction ratio and translates them according to a constant (and possibly rotates the set, in higher dimensional spaces).

Definition 3.3. Let $\Lambda \subset \mathbb{N}$ be a finite set of indices, let $\{\phi_i\}_{i\in\Lambda}$ be a family of contractions on a closed set $D \subset \mathbb{R}^n$ and let $X \subset D$. We say that the set X is invariant under the family $\{\phi_i\}_{i\in\Lambda}$ if $X = \bigcup_{i\in\Lambda} \phi_i(X)$. If ϕ_i are similarities, we say that X is *self-similar*.

A finite family of contractions $\Phi = \{\phi_i\}_{i\in\Lambda}$ is called an *iterated function system*, or *IFS*. Theorem 3.4 is valid for a general IFS but after this we will only consider those consisting of similarity mappings.

We use the following notations: $\Lambda^n = \Lambda \times \ldots \times \Lambda$ is the set of length-n multi-indices, $I = (i_1, \ldots, i_n) \in \Lambda^n$ is a multi-index of length n, $\phi_I = \phi_{i_1} \circ \ldots \circ \phi_{i_n}$ and $r_I = r_{i_1} r_{i_2} \cdots r_{i_n}$ is the ratio by which ϕ_I contracts. It turns out that for any iterated function system, the set invariant under its contractions is unique. We call this unique set the *attractor* of the IFS.

Theorem 3.4. *Denote by $\mathcal{D}$ the family of all compact subsets of D and let $\Phi = \{\phi_i\}_{i\in\Lambda}$ be an IFS on D. Define $\phi : \mathcal{D} \to \mathcal{D}$ as a transformation on $\mathcal{D}$ by setting*

$$\phi(E) = \bigcup_{i\in\Lambda} \phi_i(E)$$

for every $E \in \mathcal{D}$ and write ϕ^k for the k-th iterate of ϕ:

$$\phi^0(E) = E,\ \phi^2(E) = \phi(\phi(E)),\ \ldots,\ \phi^k(E) = \phi(\phi^{k-1}(E))$$

Then there exists a unique $F \in \mathcal{D}$, called the attractor of Φ, such that

$$F = \bigcap_{k=1}^{\infty} \phi^i(E)$$

for any $E \in \mathcal{D}$ such that $\phi_i(E) \subset E$ for all i.

To prove this, we need to expand $\mathcal{D}$ into a metric space using the following notion of distance between sets.

Definition 3.5. Define the *δ-parallel body* of a set $A \in \mathcal{D}$ as the set of points within Euclidean distance δ of A; $A_\delta = \{x \in \mathbb{R}^n \mid d(x,a) \leq \delta \text{ for some } a \in A\}$. The *Hausdorff metric* on $\mathcal{D}$ is the mapping $d_H : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$,

$$d_H(A,B) = \inf\{\delta \mid A \subset B_\delta \text{ and } B \subset A_\delta\}.$$

The Hausdorff metric is indeed a metric; for any $A \in \mathcal{D}$, clearly $d_H(A,A) = 0$. Also, for all A, B, $d_H(A,B) = d_H(B,A)$ since the definition is symmetric.

The triangle inequality is also satisfied: if $A, B, C \in \mathcal{D}$, then

$$\begin{aligned}
& d_H(A, C) + d_H(C, B) \\
= \; & \inf\{\delta_1 \mid A \subset C_{\delta_1} \text{ and } C \subset A_{\delta_1}\} + \inf\{\delta_2 \mid B \subset C_{\delta_2} \text{ and } C \subset B_{\delta_2}\} \\
\geq \; & \inf\{\delta_1 + \delta_2 \mid A \subset C_{\delta_1} \subset B_{\delta_1+\delta_2} \text{ and } B \subset C_{\delta_2} \subset A_{\delta_2+\delta_1}\} \\
\geq \; & \inf\{\delta \mid A \subset B_\delta \text{ and } B \subset A_\delta\} \\
= \; & d_H(A, B).
\end{aligned}$$

Remark 3.6. Let ϕ be a contraction with ratio c and let $a \in A$ and $b \in B$. Note that since $d(a, b) \leq \frac{1}{c}\delta$ implies $d(\phi(a), \phi(b)) \leq \delta$, $A \subset B_{\frac{1}{c}\delta}$ implies $\phi(A) \subset \phi(B)_\delta$ and therefore

$$\begin{aligned}
d_H(\phi(A), \phi(B)) &= \inf\{\delta \mid \phi(A) \subset \phi(B)_\delta \text{ and } \phi(B) \subset \phi(A)_\delta\} \\
&\leq \inf\left\{c\frac{1}{c}\delta \mid A \subset B_{\frac{1}{c}\delta} \text{ and } B \subset A_{\frac{1}{c}\delta}\right\} \\
&= c \cdot d_H(A, B). \qquad (3.1)
\end{aligned}$$

We can now prove Theorem 3.4.

Proof of Theorem 3.4. Let Φ be an IFS and let ϕ be the mapping defined in the theorem. As continuous mappings, contractions map compact sets to compact sets. Let $E \in \mathcal{D}$ be such that $\phi_i(E) \subset E$ for all i. For example, we can choose $E = D \cap \overline{B}(0, r)$ with large enough r; for all i, $\phi_i(D) \subset D$ and since ϕ_i are contracting, $\phi_i(\overline{B}(0, r)) \subset \overline{B}(0, r)$ when r is large enough.

Since $\phi(E) \subset E$, $\phi^k(E) \subset \phi^{k-1}(E)$ for every k and therefore $\phi^k(E)$ is a decreasing sequence of non-empty compact sets. Define $F = \bigcap_{i=k}^{\infty} \phi^k(E)$ as the non-empty compact intersection of this sequence. Observe that since $\phi^k(E)$ is a decreasing sequence,

$$\bigcup_{i\in\Lambda} \phi_i(F) = \phi(F) = \phi\left(\bigcap_{i=1}^{\infty} \phi^k(E)\right) = \bigcap_{i=1}^{\infty} \phi^{k+1}(E) = F,$$

so F is invariant under Φ. We now show that this F is unique. Observe that if $A, B \in \mathcal{D}$,

$$d_H(\phi(A), \phi(B)) = d_H\left(\bigcup_{i\in\Lambda} \phi_i(A), \bigcup_{i\in\Lambda} \phi_i(B)\right) \leq \max_{i\in\Lambda} d_H(\phi_i(A), \phi_i(B)), \qquad (3.2)$$

where the last inequality follows from the fact that if δ is such that for every i the δ-parallel body $(\phi_i(A))_\delta$ contains $\phi_i(B)$, then $\bigcup_{i\in\Lambda}(\phi_i(A))_\delta = \left(\bigcup_{i\in\Lambda}\phi_i(A)\right)_\delta$ contains $\bigcup_{i\in\Lambda}\phi_i(B)$. Using the equality (3.1), we see that

$$\begin{aligned}\max_{i\in\Lambda} d_H(\phi_i(A),\phi_i(B)) &\leq \max_{i\in\Lambda}(c_i\cdot d_H(A,B))\\ &= \left(\max_{i\in\Lambda} c_i\right) d_H(A,B) \qquad (3.3)\end{aligned}$$

Therefore, if A and B are invariant under Φ, by combining (3.2) and (3.3) we get

$$d_H(A,B) = d_H(\phi(A),\phi(B)) \leq \left(\max_{i\in\Lambda} c_i\right) d_H(A,B),$$

implying that $d_H(A,B)=0$ and $A=B$. $\square$

From now on, we only consider IFSs consisting of similarity mappings.

Remark 3.7. By Theorem 3.4, any self-similar set is completely specified by the associated IFS.

We may extend the notion of self-similarity to probability measures. If we attach a probability to each of the similarities of an IFS, a self-similar measure distributes the mass of 1 to its attractor X with each cylinder $\phi_I(X)$ emphasized by a probability p_I.

Definition 3.8. Let Φ be an IFS and $(p_i)_{i\in\Lambda}$ a probability vector, i.e. p_i are positive and $\sum_{i\in\Lambda} p_i = 1$. A Borel probability measure μ on a metric space X is called self-similar with respect to Φ if it satisfies

$$\mu = \sum_{i\in\Lambda} p_i \cdot \phi_i\mu. \qquad (3.4)$$

Theorem 3.9. *Given an IFS Φ and a probability vector $(p_i)_{i\in\Lambda}$, the self-similar measure associated to them exists and is unique.*

Proof. Let X be the attractor of Φ. We define a new metric d on $\mathcal{P}(X)$ by setting

$$d(\mu,\nu) = \sup\left\{\left|\int f\ d\mu - \int f\ d\nu\right| \mid f \text{ is a Lipschitz mapping with } \mathrm{Lip}(f)\leq 1\right\}.$$

We show that $(\mathcal{P}(X),d)$ is a complete metric space. It is clear that d is a metric. Let $(\mu_n)_n$ be a Cauchy sequence so that

$$\lim_{n>m\to\infty}\left|\int f\ d\mu_n - \int f\ d\mu_m\right| = 0$$

uniformly in f. Since $\mathcal{P}(X)$ is compact in the weak-* topology, we may pick a weakly converging subsequence $(\mu_{n_k})_k$ with limit μ. Then $\mu_{n_k} \to \mu$ also in d; if this was not the case, there would be a $c > 0$ such that for any k there exists a Lipschitz function f with $\mathrm{Lip}(f) \leq 1$ such that $|\int f\ d\mu_{n_k} - \int f\ d\mu| > c$. But now, if k is large enough, we have $|\int f\ d\mu_{n_l} - \int f\ d\mu| > c/2$ for all $l > k$ which contradicts the weak convergence of $(\mu_{n_k})_k$. So for any $\varepsilon > 0$, we have

$$d(\mu_n, \mu) \leq d(\mu_n, \mu_{n_k}) + d(\mu_{n_k}, \mu) < \varepsilon$$

when n and k are large enough. Therefore $(\mathcal{P}(X), d)$ is complete.

Define $F : \mathcal{P}(X) \to \mathcal{P}(X)$ by setting $F(\mu) = \sum_{i \in \Lambda} p_i \cdot \phi_i \mu$. Then F is a contraction mapping in d: for any $\mu, \nu \in \mathcal{P}(X)$,

$$\begin{aligned} d(F(\mu), F(\nu)) &\leq \sum_{i \in \Lambda} p_i \sup \left\{ \left| \int f \circ \phi\ d\mu - \int f \circ \phi\ d\nu \right| \mid \mathrm{Lip}(f) \leq 1 \right\} \\ &= \sum_{i \in \Lambda} p_i r_i \sup \left\{ \left| \int r_i^{-1} f \circ \phi\ d\mu - \int r_i^{-1} f \circ \phi\ d\nu \right| \mid \mathrm{Lip}(f) \leq 1 \right\} \\ &\leq (\max_i r_i) d(\mu, \nu), \end{aligned}$$

since $r_i^{-1} f \circ \phi_i$ is a Lipschitz mapping with $\mathrm{Lip}(r_i^{-1} f \circ \phi_i) \leq 1$. Let $r = \max_i r_i$. We show that there exists a unique measure $\tau \in \mathcal{P}(X)$ such that $F(\tau) = \tau$; this result actually holds for any contraction in a complete metric space and is known as Banach fixed-point theorem.

Indeed, let $\theta \in \mathcal{P}(X)$ be arbitrary. Define the sequence $(\theta_n)_n \subset \mathcal{P}(X)$ by setting $\theta_n = F^n(\theta) = F \circ \cdots \circ F(\theta)$ for every $n \in \mathbb{N}$. This sequence is Cauchy;

$$d(\theta_n, \theta_{n+1}) \leq r^n d(\theta, F(\theta)) = M r^n,$$

where $M = d(\theta, F(\theta))$. Therefore, for all $m, n, n > m$,

$$\begin{aligned} d(\theta_n, \theta_m) &\leq \sum_{i=0}^{n-m-1} d(\theta_{m+i}, \theta_{m+i+1}) \leq M r^m \sum_{i=0}^{n-m-1} r^i \\ &= M r^m \frac{1 - r^{n-m}}{1 - r} \to 0 \end{aligned}$$

as $m \to \infty$. Since $(\mathcal{P}(X), d)$ is complete, this sequence converges to a probability measure $\tau \in \mathcal{P}(X)$. Since

$$d(F(\tau), \tau) \leq d(F(\tau), F^n(\tau)) + d(F^n(\tau), \tau) < \varepsilon$$

for any ε whenever n is large enough, we have $\tau = F(\tau) = \sum_{i\in\Lambda} p_i \cdot \phi_i\tau$. Finally, if η is also a probability measure on X, we have

$$\begin{aligned} d(F^n(\eta),\tau) &\leq d(F^n(\eta),F^n(\theta)) + d(F^n(\theta),\tau) \\ &\leq r^n d(\eta,\theta) + d(F^n(\theta),\tau) \\ &\to 0 \end{aligned}$$

as $n \to \infty$, so τ is unique. □

3.1 Dimension of a self-similar set

Given that the cylinders $\phi_i(X)$ of an IFS are sufficiently separated from each other, the structures of its attractor and any self-similar measure associated to it are understood quite well. The main result of this section, the open set condition, states this separation condition explicitly. When this condition holds, we can derive a consistent way of calculating the dimension of the attractor and any self-similar measure associated to the IFS.

If X is the attractor of an IFS on the real line with disjoint cylinders, its dimension is the solution s to the equation

$$\mathcal{H}^s(X) = \mathcal{H}^s\left(\bigcup_{i\in\Lambda}\phi_i(X)\right) = \sum_{i\in\Lambda}\mathcal{H}^s(r_iX + a_i) = \sum_{i\in\Lambda} r_i^s\mathcal{H}^s(X)$$

which simplifies to $\sum_{i\in\Lambda} r_i^s = 1$ if $\mathcal{H}^s(X)$ is positive and finite. We aim to loosen the requirement of disjointness by introducing the following definition.

Definition 3.10 (The open set condition (OSC))**.** Let Φ be an IFS. We say that Φ satisfies the *open set condition*, or OSC, if there exists a bounded, open $\emptyset \neq V \subset \mathbb{R}^n$ for which

$$\begin{aligned} &\bigcup_{i\in\Lambda}\phi_i(V) \subset V \\ &\phi_i(V) \cap \phi_j(V) = \emptyset \text{ for all } i \neq j. \end{aligned}$$

Additionally, if $V \cap X \neq \emptyset$, where X is the attractor of Φ, we say that Φ satisfies the strong OSC (SOSC).

Theorem 3.11. *Let Φ be an IFS that satisfies the open set condition and let X be its attractor. The dimension of X is the unique $s \geq 0$ that satisfies $\sum_{i\in\Lambda} r_i^s = 1$. Moreover, for this s, $0 < \mathcal{H}^s(X) < \infty$.*

In the proof of this theorem, we require the following geometrical result.

Lemma 3.12. *Let $\{V_i\}_{i\in\mathbb{N}}$ be a collection of disjoint open subsets of $\mathbb{R}^n$ such that each V_i contains a ball $B(x_i, a_1r)$ and is contained in a ball $B(y_i, a_2r)$. Then any ball $B(x,r)$ intersects at most $(1+2a_2)^n a_1^{-n}$ of the closures $\overline{V_i}$.*

Proof. If $\overline{V_i}$ intersects $B(x,r)$, then $\overline{V_i} \subset B(x,(1+2a_2)r)$, since the diameter of $\overline{V_i}$ is at most $2a_2r$. Suppose that q of the sets $\overline{V_i}$ intersect $B(x,r)$. Again, all of these $\overline{V_i}$ are contained in $B(x,(1+2a_2)r)$ and therefore the disjoint balls $B(x_i,a_1r) \subset \overline{V_i}$ are also contained in $B(x,(1+2a_2)r)$. So we have

$$\begin{aligned} q(a_1r)^n\mathcal{L}(B(0,1)) &= q\mathcal{L}(B(x_i,a_1r)) \\ &\leq \mathcal{L}(B(x,(1+2a_2)r)) \\ &= (1+2a_2)^n r^n \mathcal{L}(B(0,1)) \end{aligned}$$

which gives the stated bound for q. $\square$

We now prove Theorem 3.11, adapting the argument from Falconer's book [4], Theorem 9.3.

Proof of Theorem 3.11. We need only check that the Hausdorff measure of X is strictly positive and finite for an s that satisfies $\sum_{i\in\Lambda} r_i^s = 1$. As previously noted, we can write $X = \bigcup_{I\in\Lambda^n} \phi_I(X)$ for any n. We will see that these covers of X provide a suitable upper estimate to the Hausdorff measure.

Let $\delta > 0$. Because the mapping ϕ_I has a contraction ratio $r_I = r_{i_1}\cdots r_{i_n}$ with every $r_i < 1$ and the diameter of X is finite, we may choose such an n that $\operatorname{diam}(\phi_I(X)) = r_I\operatorname{diam}(X) \leq \delta$. Therefore, we have

$$\begin{aligned} \mathcal{H}^s_\delta(X) &\leq \sum_{I\in\Lambda^n} \operatorname{diam}(\phi_I(X))^s = \sum_{I\in\Lambda^n} r_I^s \operatorname{diam}(X)^s \\ &= \left(\sum_{i_1\in\Lambda} r_{i_1}^s\right)\cdots\left(\sum_{i_n\in\Lambda} r_{i_n}^s\right)\operatorname{diam}(X)^s = \operatorname{diam}(X)^s. \end{aligned}$$

Since $\delta > 0$ is arbitrary, it follows that $\mathcal{H}^s(X) \leq \operatorname{diam}(X)^s < \infty$.

It remains to find a lower bound for $\mathcal{H}^s(X)$. Let

$$\Lambda^{\mathbb{N}} = \{(i_1, i_2, \ldots) \mid i_k \in \Lambda\}$$

be the set of all infinite sequences of indices in Λ and let

$$\Lambda^{\mathbb{N}}_I = \{(i_1,\ldots,i_n,q_{n+1},\ldots) \mid q_k \in \Lambda\}$$

be the collection of sequences in $\Lambda^{\mathbb{N}}$ with a fixed initial segment $(i_1,\ldots,i_n) = I$. Set $\mathcal{E}_n = \{\Lambda^{\mathbb{N}}_I\}_{I\in\Lambda^n}$ and $\mathcal{E} = \bigcup_{n\in\mathbb{N}}\mathcal{E}_n$. Let ζ be a set mapping on $\mathcal{E}$ with $\zeta(\Lambda^{\mathbb{N}}_I) = r_I^s$ and $\zeta(\Lambda^{\mathbb{N}}) = 1$. We show that $\mathcal{E}$ and ζ satisfy the conditions of Theorem 2.14 and then use the outer measure given by the theorem in applying the mass distribution principle.

Clearly the sets $\Lambda_I^{\mathbb{N}}$ are disjoint for all $I \in \Lambda^n$ and $\bigcup_{I\in\Lambda^n} \Lambda_I^{\mathbb{N}} = \Lambda^{\mathbb{N}}$ for every n. Each $\Lambda_I^{\mathbb{N}} \in \mathcal{E}_n$ is a subset of the unique $\Lambda_J^{\mathbb{N}} \in \mathcal{E}_{n-1}$ with the same fixed initial segment of length $n-1$ and contains $|\Lambda|$ subsets of $\mathcal{E}_{n+1}$. By the definition of ζ,

$$\sum_{i\in\Lambda} \zeta(\Lambda_i^{\mathbb{N}}) = \sum_{i\in\Lambda} r_i^s = 1 = \zeta(\Lambda^{\mathbb{N}})$$

and for every $\Lambda_I^{\mathbb{N}}$ with $I \in \Lambda^n$, if $J_i \in \Lambda^{n+1}$ is the multi-index I extended with $i \in \Lambda$, $\Lambda_I^{\mathbb{N}} = \bigcup_{i\in\Lambda} \Lambda_{J_i}^{\mathbb{N}}$ and

$$\sum_{i\in\Lambda} \zeta(\Lambda_{J_i}^{\mathbb{N}}) = \sum_{i\in\Lambda} r_I^s r_i^s = r_I^s = \zeta(\Lambda_I^{\mathbb{N}}).$$

Finally, if $\Lambda_{(i_1)}^{\mathbb{N}} \supset \Lambda_{(i_1,i_2)}^{\mathbb{N}} \supset \ldots$,

$$\bigcap_{k\to\infty} \Lambda_{(i_1,\ldots,i_k)}^{\mathbb{N}} = \{(i_1, i_2, \ldots)\} \subset \Lambda^{\mathbb{N}}.$$

Hence the conditions hold.

Let μ^* be the outer measure on $\Lambda^{\mathbb{N}}$ given by Theorem 2.14. Transform μ^* into an outer measure on X by defining

$$\tilde{\mu}(A) = \mu^*\left(\left\{(i_1, i_2, \ldots) \mid \bigcap_{k=1}^{\infty} \phi_{(i_1,\ldots,i_k)}(X) \subset A\right\}\right)$$

for all subsets A of X.

Next we show that $\tilde{\mu}$ satisfies the conditions of the mass distribution principle. Let V be the open set given by the OSC. Denoting by $\overline{V}$ the closure of V, we have $\overline{V} \supset \phi(\overline{V}) = \bigcup_{i\in\Lambda} \phi_i(\overline{V})$. By the uniqueness of the attractor of an IFS (Theorem 3.4), the sequence of iterates $\phi^k(\overline{V})$ converges to X. Therefore, $X = \bigcap_{k=1}^{\infty} \phi^k(\overline{V}) \subset \overline{V}$ and for each $I \in \Lambda^n$, $\phi_I(X) \subset \phi_I(\overline{V})$.

Let B be any ball in $\mathbb{R}^d$ with radius $r < 1$. For each infinite sequence $(i_1, i_2, \ldots) \in \Lambda^{\mathbb{N}}$, curtail it after the first i_k for which

$$\left(\min_{i\in\Lambda} r_i\right) r \le r_{i_1} r_{i_2} \cdots r_{i_k} \le r \tag{3.5}$$

and denote the set of all multi-indices obtained this way by Q. Then for every $(i_1, i_2, \ldots) \in \Lambda^{\mathbb{N}}$ there is exactly one k such that $(i_1, \ldots, i_k) \in Q$. Since $\phi_i(V) \subset V$ are disjoint by assumption, for any pair $(i_1, \ldots, i_k), (j_1, \ldots, j_l) \in Q$, the sets $\phi_{i_1}(\phi_{(i_2,\ldots,i_k)}(V))$ and $\phi_{j_1}(\phi_{(j_2,\ldots,j_l)}(V))$ are disjoint and from this it follows that the family $\{\phi_I(V)\}_{I\in Q}$ is disjoint. Observe that we may write $X = \bigcup_{I\in Q}\phi_I(X)$; in the presentation $X = \bigcup_{i\in\Lambda}\phi_i(X)$, write $\phi_i(X) = \bigcup_{j\in\Lambda}\phi_{(i,j)}(X)$ for every index $(i) \notin Q$. Inductively, repeat this process for every $\phi_{(i_1,\dots,i_k)}(X)$ with $(i_1,\dots,i_k) \notin Q$. We will arrive at the conclusion eventually, since Q is finite and every infinite sequence $(i_1, i_2, \dots)$ has a prefix in Q. We conclude that

$$X = \bigcup_{I\in Q}\phi_I(X) \subset \bigcup_{I\in Q}\phi_I(\overline{V}). \tag{3.6}$$

Now, as in Lemma 3.12, choose a_1 and a_2 so that V contains a ball of radius a_1 and is contained in a ball of radius a_2. Then, for all $I \in Q$, the set $\phi_I(V)$ contains a ball of radius $r_I a_1$ and one of radius $(\min_{i\in\Lambda} r_i)a_1 r$, and is contained in a ball of radius $r_I a_2$ and hence in a ball of radius $a_2 r$, by inequality (3.5). Denote by Q_1 the set of those multi-indices $I \in Q$ such that B intersects $\phi_I(\overline{V})$. By Lemma 3.12, there are at most $q = (1+2a_2)^n a_1^{-n}(\min_{i\in\Lambda} r_i)^{-n}$ elements in Q_1.

Observe that if $\bigcap_{k=1}^{\infty}\phi_{(i_1,\dots,i_k)}(X) \subset (X\cap B)$, by (3.6), $\bigcap_{k=1}^{\infty}\phi_{(i_1,\dots,i_k)}(X) \subset \bigcup_{I\in Q}\phi_I(\overline{V})$ and $\bigcap_{k=1}^{\infty}\phi_{(i_1,\dots,i_k)}(X) \subset B$. More specifically, $\phi_I(\overline{V}) \cap B$ for the $I \in Q$ that is the unique prefix obtained from $(i_1, i_2, \dots)$. Now, since $\tilde{\mu}$ has its support contained in X, we have

$$\begin{aligned}\tilde{\mu}(B) = \tilde{\mu}(X\cap B) &= \mu\left(\left\{(i_1, i_2, \dots) \mid \bigcap_{k=1}^{\infty}\phi_{(i_1,\dots,i_k)}(X) \subset (X\cap B)\right\}\right)\\ &\le \mu\left(\bigcup_{I\in Q_1}\Lambda_I^{\mathbb{N}}\right) \le \sum_{I\in Q_1}\mu(\Lambda_I^{\mathbb{N}}) = \sum_{I\in Q_1} r_I^s \le \sum_{I\in Q_1} r^s \le q r^s,\end{aligned}$$

using the inequality (3.5).

Let $U \subset \mathbb{R}^n$ be such that $\operatorname{diam}(U) < 1$. Since U is contained in a ball of radius $\operatorname{diam}(U)$, we have $\tilde{\mu}(U) \le q \cdot \operatorname{diam}(U)^s$. Mass distribution $\tilde{\mu}$ hereby satisfies the mass distribution principle and it follows that $\mathcal{H}^s(X) \ge \frac{\tilde{\mu}(X)}{q} = q^{-1} > 0$.

Combining the upper and lower bounds, we have $0 < \mathcal{H}^s(X) < \infty$ and therefore $\dim X = s$. □

3.1.1 One dimensional Sierpinski gasket

We take an example of a particular self-similar set on the plane that is the attractor of a set of three similarities. In Section 4, we return to this particular attractor to inspect the dimension of its projection to the line that turns out not to be as well-behaving as the set itself.

Define F as the subset of $\mathbb{R}^2$ consisting of points with an expansion in base 3 with negative powers of the base and digits $(0,0),(0,1),(1,0)$; that is,

$$F = \left\{ \sum a_n 3^{-n} \mid a_n \in \{(0,0),(1,0),(0,1)\} \right\}.$$

The set F is known as the *one dimensional Sierpinski gasket* (the Sierpinski gasket is defined by replacing all occurrences of 3 with 2 in the definition above) and it is the attractor of the IFS Φ defined by contractions

$$\phi_1(x,y) = \left(\frac{x}{3}, \frac{y}{3}\right), \ \phi_2(x,y) = \left(\frac{x+1}{3}, \frac{y}{3}\right), \ \phi_3(x,y) = \left(\frac{x}{3}, \frac{y+1}{3}\right).$$

To see this, note that if $(x,y) \in F$, $\phi_1(x,y) = 3^{-1}(x,y) = \sum a_n 3^{-n-1} = \sum \overline{a}_n 3^n$, where $\overline{a}_0 = (0,0)$ and $\overline{a}_k = a_{k-1}$ for $k \geq 1$. In a similar way we see that $\phi_i(F) \subset F$ for all i. Writing $\phi(A) = \bigcup_{i=1}^{3} \phi_i(A)$ for all $A \subset \mathbb{R}^2$ and ϕ^k for the k-th iterate of ϕ, this implies that $\bigcap_{k=1}^{\infty} \phi^k(F) \subset F$. On the other hand, for any $(x,y) \in F$,

$$(x,y) = \sum_{n=1}^{\infty} a_n 3^{-n} = \lim_{k\to\infty} \sum_{n=1}^{k} a_n 3^{-n} = \lim_{k\to\infty} \phi_{i_1} \circ \ldots \circ \phi_{i_k}(a_0) \in \bigcap_{k=1}^{\infty} \phi^k(\{a_0\}),$$

so $F \subset \bigcap_{k=1}^{\infty} \phi^k(F)$ and hence F is the unique attractor of Φ.

Consider the open rectangle $U = (0,1/2)^2 \subset \mathbb{R}^2$. For each contraction ϕ_i and $(x,y) \in U$, $\phi_i(x,y) \in (0, \frac{1/2+1}{3})^2 = (0,1/2)^2$ and

$$\begin{aligned}
\phi_1(U) &= \left(0, \tfrac{1}{6}\right)^2, \\
\phi_2(U) &= \left(\tfrac{1}{3}, \tfrac{1}{2}\right) \times \left(0, \tfrac{1}{6}\right), \\
\phi_3(U) &= \left(0, \tfrac{1}{6}\right) \times \left(\tfrac{1}{3}, \tfrac{1}{2}\right)
\end{aligned}$$

are disjoint. Hence Φ satisfies the open set condition and $\dim F$ is the unique s for which

$$3\left(\tfrac{1}{3}\right)^s = 1,$$

that is, $\dim F = 1$.

3.2 Dimension of a self-similar measure

To obtain a formula for the dimension of a self-similar measure, we require the associated IFS to satisfy the strong version of the open set condition. However, in Euclidean spaces, those two conditions turn out to be equivalent, as proven by A. Schief in [11].

Proposition 3.13. *In $\mathbb{R}^d$, if an IFS Φ satisfies the OSC, it also satisfies the SOSC.*

Proof. See [11], Theorem 2.1. $\square$

The formula for the dimension was introduced by A. Deliu, J. Geronimo, R. Shonkwiler and D. Hardin in [1]. It also holds for an IFS with a countable number of similarities, see e.g. [9].

Theorem 3.14. *Let μ be a self-similar measure associated to an IFS Φ that satisfies the SOSC. Then the dimension of μ is given by*

$$\dim \mu = \frac{\sum p_i \log(p_i)}{\sum p_i \log(r_i)}.$$

Proof. See [1], Theorem 1. $\square$

A measure being exact dimensional implies that its structure is similar almost everywhere. Due to their regular structure, it is quite natural for self-similar measures to have this property. This was proven by D. Feng and H. Hu in [6].

Theorem 3.15. *Self-similar measures are exact dimensional.*

Proof. See [6], Theorem 2.12. $\square$

Adding to their regularity in one-dimensional case, it turns out that self-similar measures possess a uniform entropy dimension.

Proposition 3.16. *Let $\mu \in \mathcal{P}(\mathbb{R})$ be a self-similar measure and $\alpha = \dim \mu$. Then μ has uniform entropy dimension α.*

By the translation and scaling invariance of Hausdorff dimension, we may assume without loss of generality that the attractor $X \subset [0, 1]$ and that $0 \in X$. Observe that if μ has point mass, its dimension

$$\dim \mu = \inf\{\dim E \mid E \text{ is Borel and } \mu(E) > 0\} = 0 = \lim_{n\to\infty} H_n(\mu),$$

since μ gives positive mass to a countable set; particularly, for any given $\varepsilon' > 0$, μ is (ε', n)-atomic for a large n. Combining this with Lemma 1.11 and taking ε' small enough with respect to ε, we get $\mathbb{P}_{0\le i<n}\left(H_m(\mu^{x,i}) < \varepsilon\right) > 1-\varepsilon$, which is the statement of the proposition. From now on, we assume that μ distributes no mass to countable sets.

Before we can prove the proposition in general case, we require two lemmas regarding the component measures of μ. Consider first the case of the contractions in the associated IFS contracting by a uniform ratio r. Then $\Phi = \{\phi_i\}_{i\in\Lambda}$ with $\phi_i(x) = r(x - a_i)$ for all i. Fix a point $\tilde{x}$ in the attractor X and define probability measures

$$\mu_{x,k}^{[n]} = c \cdot \sum_{\substack{I \in \Lambda^n \\ \phi_I \tilde{x} \in \mathcal{D}_k(x)}} p_I \cdot \phi_I \mu,$$

where $c = c(x, \tilde{x}, k, n)$ is a normalizing constant. Observe that $\mu_{x,k}^{[n]}$ differs from $\mu_{x,k}$ in that instead of restricting $\mu = \sum_{I\in\Lambda^n} p_I \cdot \phi_I \mu$ to the atom $\mathcal{D}_k(x)$, we include or exclude each term in its entirety depending on whether $\phi_I \tilde{x} \in \mathcal{D}_k(x)$.

For $0 < \rho < 1$, we define an integer

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

so that ρ and $r^{l(\rho)}$ differ only by a multiplicative constant.

Lemma 3.17. *For every $\varepsilon > 0$, there is a $0 < \rho < 1$ such that, for all k and $n = l(\rho 2^{-k})$,*

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

Furthermore, ρ can be chosen independently of $\tilde{x}$ and the result holds for any translate of μ.

Proof ([7], p. 31). Since μ has no point mass by assumption, given $\varepsilon > 0$, there is a $\delta > 0$ such that every interval of length δ has μ-mass less than $\varepsilon^2/2$. Choose such an integer q that $r^q < \delta/2$ and let $\rho = r^q$.

Let $k \in \mathbb{N}$ and let $l = l(2^{-k})$ so that

$$2^{-k} r \leq r^{\left\lceil \frac{-k}{\log_2 r} \right\rceil} = r^l \leq 2^{-k}.$$

Let $I \in \Lambda^l$, and consider those $J \in \Lambda^q$ such that the support of $\phi_{IJ}\mu$ is not contained in an atom of $\mathcal{D}_k$. Since all the contractions ϕ_{IJ} contract with a ratio r^{l+q}, the support is contained in an interval of length $2^{-k} \cdot \delta/2$; hence there is an interval E of length $2^{-k} \cdot \delta$ centered at one of the endpoints of an atom of $\mathcal{D}_k$ that contains the support of ϕ_{IJ}. Since ϕ_I contracts with ratio at most 2^{-k}, the support of $\phi_I \mu$ is contained in an interval of length 2^{-k} and hence gives positive mass to at most two such intervals E. Combining this with the fact that $\phi_I \mu(E) < \varepsilon^2/2$ for each such E, we conclude that in the representation

$$\mu_I = \frac{1}{p_I} \sum_{J \in \Lambda^q} p_{IJ} \cdot \phi_{IJ} \mu,$$

at least $1-\varepsilon^2$ of the mass comes from terms supported entirely on just one atom of $\mathcal{D}_k$. Since this is true for all $I \in \Lambda^l$, it is also true for the representation

$$\mu = \sum_{U\in\Lambda^{l+q}} p_U \cdot \phi_U \mu = \sum_{I\in\Lambda^l} \left(\sum_{J\in\Lambda^q} p_{IJ} \cdot \phi_{IJ}\mu \right) = \sum_{I\in\Lambda^l} p_I \mu_I.$$

Denote by $\tilde{\mu}$ the sum of terms that are supported on a single atom. Then

$$\begin{aligned}
\mathbb{E}_{i=k}\left(\left\|\mu_{x,i} - \mu_{x,i}^{[n]}\right\|\right) &\leq \mathbb{E}_{i=k}\left(\|\mu_{x,i} - \tilde{\mu}_{x,i}\|\right) + \mathbb{E}_{i=k}\left(\left\|\tilde{\mu}_{x,i} - \mu_{x,i}^{[n]}\right\|\right) \\
&= \sum_{D\in\mathcal{D}_k} \|\mu_D - \tilde{\mu}_D\|\, \mu(D) + \sum_{D\in\mathcal{D}_k} \left\|\tilde{\mu}_D - \mu_D^{[n]}\right\| \mu(D) \\
&< \varepsilon^2,
\end{aligned}$$

since μ and $\tilde{\mu}$ agree on the mass of $D \in \mathcal{D}_k$ except for the terms only partly supported on D, but their mass is at most $\varepsilon^2/2$. Also, if $\phi_I(\tilde{x}) \notin \mathcal{D}_k(x)$ and the support of $\phi_I\mu$ is partly inside $\mathcal{D}_k(x)$, the mass of such sets is at most $\varepsilon^2/2$ and otherwise agrees with that given by $\tilde{\mu}$. Returning to the statement of the lemma, Markov's inequality gives us

$$\begin{aligned}
\mathbb{P}_{i=k}\left(\left\|\mu_{x,i} - \mu_{x,i}^{[n]}\right\| < \varepsilon\right) &= 1 - \mathbb{P}_{i=k}\left(\left\|\mu_{x,i} - \mu_{x,i}^{[n]}\right\| \geq \varepsilon\right) \\
&\geq 1 - \frac{\mathbb{E}_{i=k}\left(\left\|\mu_{x,i} - \mu_{x,i}^{[n]}\right\|\right)}{\varepsilon} \\
&> 1 - \varepsilon. \qquad (3.7)
\end{aligned}$$

The second statement follows from the fact that our choice of parameters did not depend on $\tilde{x}$ and the proof is invariant under translation of μ. $\square$

Lemma 3.18. *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 ([7], p. 32). Fix an integer k, let $\varepsilon > 0$ and choose using Lemma 1.6 an $0 < \varepsilon' < \varepsilon$ such that $\|\nu - \nu'\| < \varepsilon'$ implies $|\frac{1}{m}H(\nu, \mathcal{D}_{m+k}) - \frac{1}{m}H(\nu', \mathcal{D}_{m+k}))| < \varepsilon/2$ for every m and every $\nu, \nu' \in \mathcal{P}([0,1])$. This is possible, since for any k, $\|\nu - \nu'\| < \varepsilon'$ implies $\|\nu(2^{-k}\cdot) - \nu'(2^{-k}\cdot)\| < \varepsilon'$ and the scale-$(m+k)$ entropy of ν is the scale-m entropy of $\nu(2^{-k}\cdot)$.

Choose such an integer q that $r^q < \varepsilon'/2$ and denote $r^q = \rho$. Define μ' as a translation of μ scaled down by a factor of $\sigma \leq \rho$ and translated by an x_0,

$\mu'(A) = \mu(\sigma^{-1}A + x_0)$. With large m, since $H(\mu', \mathcal{D}_m) = H(\mu, \mathcal{D}_{m-\log_2 \sigma}) + O(1)$ by Lemma 1.5 (iv.), the (ii.) of the lemma gives us

$$|H(\mu, \mathcal{D}_{m-\lfloor \log_2 \sigma \rfloor}) - H(\mu', \mathcal{D}_m)| \leq O(1)$$

since any atom of the partition $\mathcal{D}_{m-\log_2 \sigma}$ intersects at most two atoms of the partition $\mathcal{D}_{m-\lfloor \log_2 \sigma \rfloor}$. By the (v.) of the same lemma,

$$|H(\mu', \mathcal{D}_m) - H(\mu, \mathcal{D}_m)| \leq O_\sigma(1).$$

Let $l = l(\rho 2^{-k})$. Combining the inequality above with the definition of $\alpha = \lim_{n\to\infty} H_n(\mu)$, we may choose m large enough so that

$$|H_m(\mu') - \alpha| < \varepsilon/2. \tag{3.8}$$

Since $\mu^{[l]}_{x,k} = \sum_{J\in\Lambda^l,\ldots} p_J \cdot \phi_J \mu$, we have

$$\begin{aligned} \frac{1}{m} H(\mu^{[l]}_{x,k}, \mathcal{D}_{k+m}) &\geq \sum_{J\in\Lambda^l,\ldots} \frac{p_J}{m} H(\phi_J\mu, \mathcal{D}_m) \\ &> \alpha - \varepsilon/2 \end{aligned}$$

by concavity of entropy and by the fact that for every J, since $\phi_J\mu$ is just a translation of μ scaled down by a factor of at most ρ, we may choose m so that (3.8) holds for every J.

By Lemma 3.17, $\left\|\mu_{x,k} - \mu^{[l]}_{x,k}\right\| < \varepsilon'$ with μ-probability greater than $1-\varepsilon'$, and hence by the choice of ε', we have $|\frac{1}{m}H(\mu_{x,k}, \mathcal{D}_{m+k}) - \frac{1}{m}H(\mu^{[l]}_{x,k}, \mathcal{D}_{m+k})| < \varepsilon/2$. Combining this with the identity $H_m(\mu^{x,k}) = \frac{1}{m}H(\mu_{x,k}, \mathcal{D}_{m+k})$ from Section 1 (equation (1.4)), we have the statement of the lemma. The second statement follows again from the translation invariance of the proof. $\square$

We are now ready to prove the uniformity of the entropy dimension for a self-similar measure.

Proof of Proposition 3.16 ([7], p. 32). Let $0 < \varepsilon < 1$ be given and fix an $\varepsilon' < \varepsilon^2/16$. Let m be large enough so that

$$\mathbb{P}_{0\leq i<n}\left(H_m(\mu^{x,i}) > \alpha - \varepsilon'\right) = \frac{1}{n}\sum_{k=0}^{n-1} \mathbb{P}_{i=k}\left(H_m(\mu^{x,i}) > \alpha - \varepsilon'\right) > 1 - \varepsilon' \tag{3.9}$$

for any n, by Lemma 3.18. Assuming n large enough, we have $|H_n(\mu) - \alpha| < \varepsilon'$ and therefore by Lemma 1.7,

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

Assuming the event of (3.9), $H_m(\mu^{x,i}) - \alpha + \varepsilon' > 0$, the inequality above combined with Markov's inequality implies that

$$\begin{aligned}
\mathbb{P}_{0\le i<n}\left(H_m(\mu^{x,i}) + \varepsilon' < \alpha + \varepsilon''\right) &= 1 - \mathbb{P}_{0\le i<n}\left(H_m(\mu^{x,i}) - \alpha + \varepsilon' \ge \varepsilon''\right) \\
&\ge 1 - \frac{\mathbb{E}_{0\le i<n}\left(H_m(\mu^{x,i})\right) - \alpha + \varepsilon'}{\varepsilon''} \\
&> 1 - \frac{2\varepsilon'}{\varepsilon''} \\
&\ge 1 - \varepsilon'', \qquad (3.10)
\end{aligned}$$

when $\varepsilon'' = 2\sqrt{\varepsilon'} < \varepsilon/2$. Hence, (3.9) combined with (3.10) gives

$$\mathbb{P}_{0\le i<n}\left(|H_m(\mu^{x,i}) - \alpha| < \varepsilon\right) \ge (1-\varepsilon')(1-\varepsilon'') = 1 - 2\varepsilon'' > 1 - \varepsilon,$$

which is the statement of the proposition; this finishes the proof of the case with uniform contraction ratios.

Assume then that the IFS is non-uniformly contracting and write $\phi_i(x) = r_i x + a_i$ for the contractions. Let $0 < r < 1$, let n be given and denote by $\Lambda^{(n)}$ the set of arbitrarily long multi-indices $I \in \Lambda^* = \bigcup_{m=1}^{\infty} \Lambda^m$ such that $r_I < r^n \le r_J$, where J is obtained from I by removing the last element of the tuple. With this definition, the numbers $\{r_I \mid I \in \Lambda^{(n)}\}$ are all within a multiplicative constant of each other. Now define $\mu_{x,k}^{[n]}$ as before but summing over $\Lambda^{(n)}$ instead of Λ^n,

$$\mu_{x,k}^{[n]} = c \cdot \sum_{\substack{I \in \Lambda^{(n)}, \\ \phi_I(\tilde{x}) \in \mathcal{D}_k(x)}} p_I \cdot \phi_I \mu.$$

With this modification, all the previous arguments go through; the uniformity of contraction ratios was explicitly used only in the discussion leading to (3.7), but by the definition of $\Lambda^{(n)}$, the arguments regarding the length of the support of $\phi_I\mu$ hold and since each multi-index $I \in \Lambda^*$ has a unique prefix in $\Lambda^{(n)}$, we can compare the sum-presentations of $\tilde{\mu}$ and $\mu_{x,k}^{[n]}$ as in (3.7). □

4 Self-similar sets with overlaps

In this section, we focus only on iterated function systems on the line. Associated similarities are written $\phi_i(x) = r_i x + a_i$ or $\phi_i(x) = r_i(x + a_i)$ with $r_i < 1$ and $a_i \in \mathbb{R}$, whichever expression is more convenient.

From the previous section, we have a consistent way of calculating the dimension of a self-similar object when sufficient separation conditions hold.

However, generally we can expect no separation of this kind and there may be overlaps in the cylinders of the associated IFS. To compare the dimensions of an object with and without assuming the OSC and SOSC, we define the similarity dimension for self-similar sets and measures.

Definition 4.1. Let X be the attractor of an IFS Φ and let μ be the self-similar measure associated to a probability vector $(p_1, \ldots, p_{|\Lambda|})$. The *similarity dimension* of X, denoted by $\dim_{\mathcal{S}} X$, is the unique $s \geq 0$ satisfying $\sum_{i\in\Lambda} r_i^s = 1$. The similarity dimension of μ is defined by

$$\dim_{\mathcal{S}} \mu = \frac{\sum_{i\in\Lambda} p_i \log p_i}{\sum_{i\in\Lambda} p_i \log r_i}.$$

Remark 4.2. The similarity dimension is well-defined in the sense that we may use any iteration of the IFS in the definition: clearly

$$\sum_{I\in\Lambda^n} r_I^s = \left(\sum_{i_1\in\Lambda} r_{i_1}^s\right) \cdots \left(\sum_{i_n\in\Lambda} r_{i_n}^s\right) = 1$$

if and only if $\sum_{i\in\Lambda} r_i^s = 1$. Also, if we write $p_I = p_{i_1}p_{i_2}\cdots p_{i_n}$ for any $I \in \Lambda^n$, we have

$$\begin{aligned}\frac{\sum_{I\in\Lambda^n} p_I \log p_I}{\sum_{I\in\Lambda^n} p_I \log r_I} &= \frac{\sum_{i\in\Lambda}\sum_{I\in\Lambda^{n-1}} p_i p_I \log p_i + \sum_{i\in\Lambda}\sum_{I\in\Lambda^{n-1}} p_i p_I \log p_I}{\sum_{i\in\Lambda}\sum_{I\in\Lambda^{n-1}} p_i p_I \log r_i + \sum_{i\in\Lambda}\sum_{I\in\Lambda^{n-1}} p_i p_I \log r_I} \\ &= \frac{\sum_{i\in\Lambda} p_i \log p_i + \sum_{I\in\Lambda^{n-1}} p_I \log p_I}{\sum_{i\in\Lambda} p_i \log r_i + \sum_{I\in\Lambda^{n-1}} p_I \log r_I}\end{aligned}$$

and continuing the iteration, we arrive at

$$\frac{\sum_{I\in\Lambda^n} p_I \log p_I}{\sum_{I\in\Lambda^n} p_I \log r_I} = \frac{n\sum_{i\in\Lambda} p_i \log p_i}{n\sum_{i\in\Lambda} p_i \log r_i} = \dim_{\mathcal{S}} \mu.$$

Using these concepts, we may bound the dimensions in the case where the open set condition does not hold.

Lemma 4.3. *For any self-similar X and μ, the following bounds hold:*

(i.) $\dim X \leq \min\{1,\ \dim_{\mathcal{S}} X\}$

(ii.) $\dim \mu \leq \min\{1,\ \dim_{\mathcal{S}} \mu\}$

Proof. Because $X \subset \mathbb{R}$ and the support of μ is contained in X, clearly $\dim X \leq \dim \mathbb{R} = 1$ and $\dim \mu \leq \dim \mathbb{R} = 1$. The representation $X =$

$\bigcup_{I \in \Lambda^n} \phi_I(X)$ gives a suitable upper bound for the Hausdorff measure of X. As in the proof of Theorem 3.11, for any s and δ we can estimate

$$\begin{aligned}\mathcal{H}^s_\delta(X) &\leq \sum_{I \in \Lambda^n} \operatorname{diam}(\phi_I(X))^s = \sum_{I \in \Lambda^n} r_I^s \operatorname{diam}(X)^s \\ &= \left(\sum_{i_1 \in \Lambda} r_{i_1}^s\right) \cdots \left(\sum_{i_n \in \Lambda} r_{i_n}^s\right) \operatorname{diam}(X)^s,\end{aligned}$$

since $\operatorname{diam}(\phi_I(X)) \leq (\max_{i \in \Lambda} r_i)^n \operatorname{diam}(X) < \delta$ when n is large enough. If $s > \dim_{\mathcal{S}} X$, $\sum_{i_j \in \Lambda} r_{i_j}^s < 1$ for all $1 \leq j \leq n$. By letting $n \to \infty$, we see that $\mathcal{H}^s_\delta(X) = 0$ for every δ. Therefore $\dim X \leq \dim_{\mathcal{S}} X$, proving the first statement.

The bound for $\dim \mu$ is a bit more awkward. The reader is suggested to refer to the discussion below Theorem 4.9 where the following concepts are introduced in more detail.

Let $r = \prod_{i \in \Lambda} r_i^{p_i}$ and $n' = n \log_2(1/r)$. Define the n-th generation approximation of μ, $\nu^{(n)} = \sum_{I \in \Lambda^n} p_I \cdot \delta_{\phi_I(0)}$, as in the discussion after the statement of Theorem 4.9. In the proof of the theorem, it is noted that

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

Using this, we obtain

$$\begin{aligned}\dim \mu &= \lim_{n' \to \infty} \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{n'}) \\ &= \lim_{n' \to \infty} \frac{\sum_{D \in \mathcal{D}_{n'}} \left(\sum_{I \in \Lambda^n} p_I \delta_{\phi_I(0)}(D)\right) \log_2 \left(\sum_{I \in \Lambda^n} p_I \delta_{\phi_I(0)}(D)\right)}{-n'} \\ &\leq \lim_{n' \to \infty} \frac{\sum_{D \in \mathcal{D}_{n'}} \left(\sum_{I \in \Lambda^n} p_I \delta_{\phi_I(0)}(D) \log_2 p_I\right)}{n \sum_{i \in \Lambda} p_i \log_2 r_i} \\ &= \lim_{n' \to \infty} \frac{\sum_{I \in \Lambda^n} p_I \log_2 p_I}{n \sum_{i \in \Lambda} p_i \log_2 r_i} \\ &= \lim_{n' \to \infty} \frac{\sum_{i \in \Lambda} \sum_{I \in \Lambda^{n-1}} p_i p_I \log_2 p_i + \sum_{i \in \Lambda} \sum_{I \in \Lambda^{n-1}} p_i p_I \log_2 p_I}{n \sum_{i \in \Lambda} p_i \log_2 r_i} \\ &= \lim_{n' \to \infty} \frac{\sum_{i \in \Lambda} p_i \log_2 p_i + \sum_{I \in \Lambda^{n-1}} p_I \log_2 p_I}{n \sum_{i \in \Lambda} p_i \log_2 r_i}.\end{aligned}$$

By continuing the iteration, we arrive at the limit

$$\lim_{n' \to \infty} \frac{n \sum_{i \in \Lambda} p_i \log_2 p_i}{n \sum_{i \in \Lambda} p_i \log_2 r_i} = \dim_{\mathcal{S}} \mu.$$

□

We say that two cylinders of an IFS overlap if they are the same function and can be written as two different compositions of similarities.

Definition 4.4. Let $\Phi = \{\phi_i\}_{i\in\Lambda}$ be an IFS. We say that an *exact overlap* occurs in Φ if there are distinct multi-indices $I, J \in \Lambda^n$ such that $\phi_I = \phi_J$.

Remark 4.5. In the definition we could equivalently require overlapping of cylinders in any two scales: if there are multi-indices of different length, $I \in \Lambda^n$ and $J \in \Lambda^m$, such that $\phi_I = \phi_J$, then $IJ, JI \in \Lambda^{n+m}$ and $\phi_I \circ \phi_J = \phi_J \circ \phi_I$.

If an exact overlap occurs, then X and any associated self-similar measure can be expressed using an IFS that is a proper subset of $\{\phi_I\}_{I\in\Lambda^n}$, where n is the level at which the overlap of cylinders occurs. This always leads to a decrease in the similarity dimension of the attractor: if $\phi_K = \phi_J$ and s is such that $\sum_{I\in\Lambda^n} r_I^s = 1$, then $\sum_{I\in\Lambda^n,\ I\neq J} r_I^s < 1$. Also, any self-similar measure can be written $\mu = (p_K + p_J)\cdot \phi_K\mu + \sum_{I\in\Lambda^n,\ I\neq J,K} p_I \cdot \phi_I\mu$ and because $\log(p_K + p_J) > \log p_K + \log p_J$, the similarity dimension also decreases.

Definition 4.6. Let $\Phi = \{\phi_i\}_{i\in\Lambda}$. Define the distance between cylinders ϕ_I and ϕ_J by

$$d(I,J) = \begin{cases} \infty, & \text{if } r_I \neq r_J \\ |\phi_I(0) - \phi_J(0)|, & \text{if } r_I = r_J \end{cases}$$

and as a tool for identifying exact overlaps, define

$$\Delta_n = \min\{d(I,J) \mid I, J \in \Lambda^n,\ I \neq J\}.$$

Remark 4.7. We make a few observations regarding d and Δ_n.

(i.) Assume $d(I,J) = 0$ for some multi-indices I and J. Write $\phi_I(x) = r_I x + a_I$, where $a_I \in \mathbb{R}$ is the total translation of x by similarities $\{\phi_i\}_{i\in I}$. Then, $\phi_I(0) = a_I = a_J = \phi_J(0)$ by definition and $\phi_I = \phi_J$. Note that since $r_I = r_J$, we could define the distance using any $x \in X$ in the place of 0.

(ii.) An exact overlap occurs if and only if $\Delta_n = 0$ for some n.

(iii.) The number Δ_n converges to 0 exponentially. To see this, observe that the exponentially many multi-indices Λ^n give rise to only polynomially many contraction ratios r_I; by a basic "stars-and-bars" combinatoric argument, the amount of different length-n contraction ratios is $\binom{n+|\Lambda|-1}{|\Lambda|-1}$ which is a polynomial of degree $|\Lambda|$. Hence the number of different contraction ratios is bounded by $O(n^{|\Lambda|})$ and there are two distinct I, J with $r_I = r_J$ and $|\phi_I(0) - \phi_J(0)| \leq O_{\mathrm{diam}(X)}(|\Lambda|^{-n} n^{-|\Lambda|})$.

4.1 The dimension drop

We will now point our focus onto the dimension of a general self-similar measure. Assuming the separation conditions do not hold, it is largely unknown whether the bounds for its dimension presented in Lemma 4.3 hold with equalities or strict inequalities. The main motivation behind the results in this section is the following conjecture.

Conjecture 4.8. *If the bounds in Lemma 4.3 hold with strict inequalities, there is an exact overlap in the cylinders of the associated IFS.*

Utilizing the notion of distance between cylinders, the following theorem by Hochman [7] is our main result in this discussion.

Theorem 4.9. *Let Φ be an IFS on $\mathbb{R}$. If there is a self-similar measure μ associated to Φ that satisfies $\dim \mu < \min\{1, \dim_{\mathcal{S}} \mu\}$, then $\Delta_n \to 0$ super-exponentially, i.e.* $\lim(-\frac{1}{n} \log \Delta_n) = \infty$.

Particularly, if the conclusion fails, we know that $\dim \mu = \dim_{\mathcal{S}} \mu$ for every self-similar measure associated to Φ. As with Proposition 3.16, we will first consider the case in which Φ is uniformly contracting with contraction ratio of r.

Let μ be a self-similar measure associated to Φ and write $p_I = p_{i_1} p_{i_2} \cdots p_{i_n}$ for its probability vector $(p_i)_{i \in \Lambda}$ and for all multi-indices $I \in \Lambda^n$. We define the *n-th generation approximation* of μ by

$$\nu^{(n)} = \sum_{I \in \Lambda^n} p_I \cdot \delta_{\phi_I(0)}.$$

Lemma 4.10. *The n-th generation approximation $\nu^{(n)}$ is a probability measure on X and converges weakly to μ.*

Proof. Clearly $\nu^{(n)}(\emptyset) = 0$. Assume $A_1, A_2, \ldots \subset \mathbb{R}$ are disjoint. Then

$$\begin{aligned} \nu^{(n)} \left(\bigcup_{k=1}^{\infty} A_k \right) &= \sum_{I \in \Lambda^n} p_I \cdot \delta_{\phi_I(0)} \left(\bigcup_{k=1}^{\infty} A_k \right) \\ &= \sum_{I \in \Lambda^n} p_I \sum_{k=1}^{\infty} \delta_{\phi_I(0)}(A_k) = \sum_{k=1}^{\infty} \nu^{(n)}(A_k) \end{aligned}$$

and

$$\nu^{(n)}(X) = \sum_{I \in \Lambda^n} p_I = \left(\sum_{i_1 \in \Lambda} p_{i_1} \right) \cdots \left(\sum_{i_n \in \Lambda} p_{i_n} \right) = 1.$$

Let (a, b) be an open interval in $\mathbb{R}$. Since μ is a Borel measure, we have

$$\mu(a, b) = \lim_{k \to \infty} \mu\left(\left[a + \frac{1}{k}, b - \frac{1}{k}\right]\right).$$

Let $\varepsilon > 0$ be arbitrary and choose such N that $\mu(a, b) < \mu\left[a + \frac{1}{N}, b - \frac{1}{N}\right] + \varepsilon$. Choose also an integer $\tilde{n}$ so that $r^{\tilde{n}}\mathrm{diam}(X) < \frac{1}{N}$ and let $n \geq \tilde{n}$, $I \in \Lambda^n$. Since $|\phi_I(0) - \phi_I(x)| \leq r^n \mathrm{diam}(X)$ for any $x \in X$, we note that if $\phi_I(x) \in \left[a + \frac{1}{N}, b - \frac{1}{N}\right]$, then $\phi_I(0) \in (a, b)$. Denoting by Q the set of all multi-indices I of length n such that $\phi_I^{-1}(\left[a + \frac{1}{N}, b - \frac{1}{N}\right]) \neq \emptyset$, we have

$$\begin{aligned}\nu^{(n)}(a, b) &= \sum_{I \in \Lambda^n} p_I \cdot \delta_{\phi_I(0)}(a, b) = \sum_{I\colon\ \phi_I(0) \in (a,b)} p_I \geq \sum_{I \in Q} p_I \\ &\geq \sum_{I \in \Lambda^n} p_I \cdot \mu(\phi_I^{-1}(\left[a + \tfrac{1}{N}, b - \tfrac{1}{N}\right])) = \mu\left(\left[a + \tfrac{1}{N}, b - \tfrac{1}{N}\right]\right) > \mu(a, b) - \varepsilon.\end{aligned}$$

Letting $\varepsilon \to 0$ gives us $\nu^{(n)}(a, b) \geq \mu(a, b)$. Since this can be done for every $n \geq \tilde{n}$ and every open set in $\mathbb{R}$ is a countable union of disjoint open intervals, we have $\liminf_{n \to \infty} \nu^{(n)}(U) \geq \mu(U)$ for all open $U \subset \mathbb{R}$. $\square$

Note that the entropy of $\nu^{(n)}$ with respect to the partition $\mathcal{D}_{n'}$ may not be equal to the entropy with respect to the partition into its atoms which is $\sum_{I \in \Lambda^n} p_I \log_2 \left(\sum_{\phi_I(0) = \phi_J(0)} p_J\right)$. This is the case when there are distinct I and J of length n such that $\phi_I(0) \neq \phi_J(0)$ but they belong to the same atom of $\mathcal{D}_{n'}$. We want to use the approximation measure $\nu^{(n)}$ to identify exact overlaps in the cylinders of Φ and hence it is desirable that the aforementioned case does not occur with large n.

Indeed, if the entropy with respect to the partition into its atoms is substantially greater than $H(\nu^{(n)}, \mathcal{D}_{n'})$, we are interested in the rate of n' at which this excess entropy appears. Since $\lim_{k \to \infty} H(\nu^{(n)}, \mathcal{D}_k) = H(\nu^{(n)})$, it does appear eventually. The excess entropy at scale k relative to the entropy at scale n' is the conditional entropy $H(\nu^{(n)}, \mathcal{D}_k | \mathcal{D}_{n'}) = H(\nu^{(n)}, \mathcal{D}_k) - H(\nu^{(n)}, \mathcal{D}_{n'})$.

The following theorem states that for a measure with dimension strictly less than 1, the excess entropy that appears at very small scale is negligible and hence $\nu^{(n)}$ may be used in identifying exact overlaps.

Theorem 4.11. *Let μ be a self-similar measure on $\mathbb{R}$ defined by an IFS with uniform contraction ratios. Let $\nu^{(n)}$ be the approximation of μ defined above and $n' = n\lfloor \log_2(1/r) \rfloor$. If $\dim \mu < 1$, then*

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

Proof ([7], p. 33). Observe that in the expression $\mu = \sum_{I\in\Lambda^n} p_I \cdot \phi_I\mu$, all $\phi_I\mu$ are translates of one another due to every ϕ_I having contraction ratio r^n. If we define $\tau^{(n)}$ by scaling μ down by r^n, $\tau^{(n)}(A) = \mu(r^{-n}A)$, for every $A \subset X$ we may write

$$\begin{aligned}\mu(A) &= \sum_{I\in\Lambda^n} p_I \cdot \mu \circ \phi_I^{-1}(A) \\ &= \sum_{I\in\Lambda^n} p_I \cdot \mu(r^{-n}(A - \phi_I(0))) \\ &= \sum_{I\in\Lambda^n} p_I \cdot (\delta_{\phi_I(0)} * \tau^{(n)})(A) \\ &= (\nu^{(n)} * \tau^{(n)})(A),\end{aligned}$$

using the bilinearity of convolution.

Define the functions $f, g : X \times X \to \mathbb{R}$ by $f(x,y) = x + 2^{-n'}y$ and $g(x,y) = x$. Observe that $f(\nu^{(n)} \times \mu) = \nu^{(n)} * \tau^{(n)}$ and $g(\nu^{(n)} \times \mu) = \nu^{(n)}$, and that $|f(x,y) - g(x,y)| = |y| \leq 2^{-n'}\mathrm{diam}(X)$ for all $x, y \in X$. Hence by Lemma 1.5 (iii.),

$$\begin{aligned}|H(\mu, \mathcal{D}_{n'}) - H(\nu^{(n)}, \mathcal{D}_{n'}| &= |H(\nu^{(n)} * \tau^{(n)}, \mathcal{D}_{n'}) - H(\nu^{(n)}, \mathcal{D}_{n'})| \\ &= O_{\mathrm{diam}(X)}(1). \end{aligned} \tag{4.1}$$

Fix a $q > \lfloor \log_2(1/r) \rfloor$ and use the notation $a \approx b$ to indicate that the difference tends to 0 as n tends to infinity. By (4.1),

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

Suppose now that $\alpha = \dim\mu < 1$. Consider the identity immediate from the conditional entropy formula in Lemma 1.4,

$$\begin{aligned}&\frac{1}{qn}H(\mu, \mathcal{D}_{qn}) \\ &= \frac{n'}{qn} \cdot \left(\frac{1}{n'}H(\mu, \mathcal{D}_{n'})\right) + \frac{qn - n'}{qn} \cdot \left(\frac{1}{qn - n'}H(\mu, \mathcal{D}_{qn}|\mathcal{D}_{n'})\right) \\ &= \frac{\lfloor \log_2(1/r) \rfloor}{q} \left(\frac{1}{n'}H(\mu, \mathcal{D}_{n'})\right) + \frac{q - \lfloor \log_2(1/r) \rfloor}{q} \left(\frac{1}{qn - n'}H(\mu, \mathcal{D}_{qn}|\mathcal{D}_{n'})\right).\end{aligned}$$

As n tends to infinity, the terms $\frac{1}{qn}H(\mu, \mathcal{D}_{qn})$ and $\frac{1}{n'}H(\mu, \mathcal{D}_{n'})$ tend to α. Since r and q are independent of n, we obtain the limit

$$\lim_{n\to\infty} \frac{1}{qn - n'}H(\mu, \mathcal{D}_{qn}|\mathcal{D}_{n'}) = \alpha. \tag{4.2}$$

From the identity $\nu^{(n)} = \mathbb{E}_{i=n'}\left(\nu_{y,i}^{(n)}\right)$ which is the equality (0.24) in Section 0, and the bilinearity of convolution, we get

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

Since $\nu_{y,n'}^{(n)}$ has its support contained in an interval of length $2^{-n'}$ and $\tau^{(n)}$ in an interval of length $r^n \mathrm{diam}(X) = O(2^{-n'})$, the support of each measure $\nu_{y,n'}^{(n)} * \tau^{(n)}$ is of order $O(2^{-n'})$. Hence,

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

By concavity of conditional entropy from Lemma 1.4,

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

Combining this with (4.2), we get

$$\limsup_{n\to\infty} \frac{1}{qn - n'} \mathbb{E}_{i=n'}\left(H(\nu_{y,i}^{(n)} * \tau^{(n)}, D_{qn})\right) \leq \alpha. \tag{4.3}$$

We also note the equality

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{qn} H(\mu, \mathcal{D}_{qn}) &= \lim_{n\to\infty} \frac{1}{qn - n'} H(\mu, \mathcal{D}_{qn-n'}) \\
&= \lim_{n\to\infty} \frac{1}{qn - n'} H(\tau^{(n)}, \mathcal{D}_{qn}) = \alpha,
\end{aligned} \tag{4.4}$$

since $\tau^{(n)}$ is μ scaled down by $2^{-n'}$. By Lemma 1.25, for each component $\nu_{y,n'}^{(n)}$,

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

By taking limits when $n \to \infty$, we get

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

for any component $\nu_{y,n'}^{(n)}$ and $\delta > 0$, and therefore

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

On the other hand, by Markov's inequality and (4.3),

$$\begin{aligned}
&\mathbb{P}_{i=n'} \left(\frac{1}{qn-n'} H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) + \delta < \alpha + \delta' \right) \\
&= 1 - \mathbb{P}_{i=n'} \left(\frac{1}{qn-n'} H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) - \alpha + \delta \geq \delta' \right) \\
&\geq 1 - \frac{\frac{1}{qn-n'} \mathbb{E}_{i=n'} \left(H(\nu_{y,i}^{(n)} * \tau^{(n)}, D_{qn}) \right) - \alpha + \delta}{\delta'} \\
&\geq 1 - \delta',
\end{aligned}$$

when $\delta' = \sqrt{\delta}$ and n is large. This combined with the previous limit gives

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

Replacing α with the limit from (4.4), we have

$$\lim_{n\to\infty} \mathbb{P}_{i=n'} \left(\left| \frac{1}{qn-n'} H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) - \frac{1}{qn-n'} H(\tau^{(n)}, \mathcal{D}_{qn}) \right| < \delta' \right) > 1-\delta' \tag{4.5}$$

for any $\delta > 0$.

Let $\varepsilon > 0$. By the uniformity of the entropy dimension (Proposition 3.16) and the assumption that $\alpha < 1$, for small enough ε, large enough m and large n,

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

The bounds for i here follow from the fact that the components of $\tau^{(n)}$ are identically distributed to those of μ when the scale is decreased by $2^{-n'}$. Choose now $\delta > 0$ so that δ' is smaller than ε and smaller than the $\delta(\varepsilon, m)$ in the statement of Theorem 1.28. Assuming n is sufficiently large, we apply the theorem in question to the measure $\nu_{y,i}^{(n)}$ in the event of equation (4.5). By the theorem, since $\frac{1}{qn-n'} H(\nu_{y,i}^{(n)} * \tau^{(n)}, \mathcal{D}_{qn}) < \frac{1}{qn-n'} H(\tau^{(n)}, \mathcal{D}_{qn}) + \delta'$, each component must satisfy $\frac{1}{qn-n'} H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn}) < \varepsilon$. Hence by (4.5),

$$\lim_{n\to\infty} \mathbb{P}_{i=n'} \left(\frac{1}{qn-n'} H(\nu_{y,i}^{(n)}, D_{qn}) < \varepsilon \right) > 1 - \delta' > 1 - \varepsilon. \tag{4.6}$$

Thus, from the definition of conditional entropy and the inequality above, we conclude that

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{qn-n'} H(\nu^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{n'}) &= \lim_{n\to\infty} \frac{1}{qn-n'} \sum_{D\in\mathcal{D}_{n'}} \nu^{(n)}(D) H(\nu_D^{(n)}, \mathcal{D}_{qn}) \\
&= \lim_{n\to\infty} \frac{1}{qn-n'} \mathbb{E}_{i=n'} \left(H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn}) \right) \\
&= \lim_{n\to\infty} \mathbb{E}_{i=n'} \left(\frac{1}{qn-n'} H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn}) \right) \\
&< \varepsilon(1-\varepsilon) + \varepsilon \\
&< \varepsilon'
\end{aligned} \tag{4.7}
$$

for some $\varepsilon' > 0$ that goes to 0 with ε. The second to last inequality follows from the fact that when $i = n'$, $\nu_{y,i}^{(n)}$ is supported on $2^{qn-n'}$ intervals of the partition $\mathcal{D}_{qn}$ and hence outside the event of (4.6), the random variable has the bound $\frac{1}{qn-n'} H(\nu_{y,i}^{(n)}, \mathcal{D}_{qn}) < \log_2(2^{qn-n'})/(qn-n') = 1$.

To get the limit in the form it was in the theorem, we have yet to do some simple modifications; by Lemma 1.5 (v.), since $qn'/qn = \lfloor \log_2(1/r) \rfloor$ for each n, we have

$$
\left| \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{qn'}) - \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{qn}) \right| \leq O_r\left(\frac{1}{n'}\right).
$$

Note that as a consequence of (4.7), we have

$$
\lim_{n\to\infty} \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{n'}) < \left(\frac{q}{\lfloor \log_2(1/r) \rfloor} - 1 \right) \varepsilon'.
$$

Combining these, we get

$$
\begin{aligned}
\lim_{n\to\infty} \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{qn'}|\mathcal{D}_{n'}) &= \lim_{n\to\infty} \left(\frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{qn'}) - \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{n'}) \right) \\
&\leq \lim_{n\to\infty} \left(\frac{1}{n'} \left(H(\nu^{(n)}, \mathcal{D}_{qn}) - H(\nu^{(n)}, \mathcal{D}_{n'}) \right) + O_r\left(\frac{1}{n'}\right) \right) \\
&= \lim_{n\to\infty} \frac{1}{n'} H(\nu^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{n'}) \\
&< \varepsilon''
\end{aligned}
$$

for $\varepsilon'' = \left(\frac{q}{\lfloor \log_2(1/r) \rfloor} - 1 \right) \varepsilon'$ which goes to 0 with ε. We may also depose of the requirement $q > \lfloor \log_2(1/r) \rfloor$, since in this case the last limit above equals 0 trivially. This finishes the proof. $\square$

By the theorem, $\frac{1}{n'}H(\nu^{(n)}, \mathcal{D}_{qn'})$ is close to the normalized scale-n' entropy of $\nu^{(n)}$ which in turn approaches the dimension of μ as n' grows. In particular, $\frac{1}{n'}H(\nu^{(n)}, \mathcal{D}_{qn'}) \approx \dim \mu < \dim_{\mathcal{S}} \mu$ implies that for every q, there are distinct multi-indices I and J of length n and $\phi_I(0), \phi_J(0)$ belonging to the same atom of $\mathcal{D}_{qn'}$ and therefore $\Delta_{n'} < 2^{-qn'}$; we will discuss this with more detail in the proof of Theorem 4.9 after generalizing the result above to an IFS with non-uniform contraction ratios.

Let $\Phi = \{\phi_i \mid \phi_i(x) = r_i x + a_i \text{ for all } x \in \mathbb{R}\}_{i\in\Lambda}$, let $(p_i)_{i\in\Lambda}$ be a probability vector and define

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

so that $\log_2 r$ is the expected value of $\log_2 r_i$ when r_i are chosen randomly, each with probability p_i. Note that if $I = (i_1, \ldots, i_n) \in \Lambda^n$ is chosen with probability p_I, by the law of large numbers,

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{\log_2 r_I}{n} - \log_2 r \right| < \varepsilon \right) = 1$$

for all $\varepsilon > 0$. Define again $n' = n\lfloor \log_2(1/r) \rfloor$. By the above, since $r^n = 2^{-n'}$, we have

$$r_I = r^{n(1+o(1))} = 2^{-n'(1+o(1))} \tag{4.8}$$

with probability that tends to 1 as n tends to infinity.

In the non-uniform case, we cannot use the partition $\mathcal{D}_n$ to detect exact overlaps, since $\phi_I(0) = \phi_J(0)$ may happen for some $I, J \in \Lambda^n$ with different contraction ratios. To take this into account, we slightly modify the approximation measure of μ and define the probability measure $\tilde{\nu}^{(n)}$ on $\mathbb{R} \times \mathbb{R}$ by setting

$$\tilde{\nu}^{(n)} = \sum_{I\in\Lambda^n} p_I \cdot \delta_{(\phi_I(0), r_I)}.$$

Define also the partition of $\mathbb{R} \times \mathbb{R}$,

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

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

Theorem 4.12. *Let μ be a self-similar measure on $\mathbb{R}$ and $\tilde{\nu}^{(n)}$ as above. If* $\dim \mu < 1$, *then*

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

Proof ([7], p. 35). For any integer n, define the set of all length-n contraction ratios

$$R_n = \{r_I \mid I \in \Lambda^n\}.$$

As the number of length-n multi-indices $|\Lambda|^n$ gives rise to only polynomially many contraction ratios, we may write $|R_n| \leq O(n^{|\Lambda|})$. Since $\tilde{\nu}^{(n)}$ is supported on $|R_n|$ points of $\{\mathbb{R}\} \times \mathcal{F}$, we have $H(\tilde{\nu}^{(n)}, \{\mathbb{R}\} \times \mathcal{F}) = O(\log_2 n)$ and consequently for all k, since $\tilde{\mathcal{D}}_k = (\mathcal{D}_k \times \mathcal{F}) \vee (\{\mathbb{R}\} \times \mathcal{F})$,

$$\begin{aligned} H(\tilde{\nu}^{(n)}, \tilde{\mathcal{D}}_k) &= H(\tilde{\nu}^{(n)}, (\mathcal{D}_k \times \mathcal{F}) | (\{\mathbb{R}\} \times \mathcal{F})) + H(\tilde{\nu}^{(n)}, \{\mathbb{R}\} \times \mathcal{F}) \\ &= H(\nu^{(n)}, \mathcal{D}_k) + O(\log_2 n). \end{aligned}$$

Thus

$$H(\tilde{\nu}^{(n)}, \tilde{\mathcal{D}}_{qn} | \tilde{\mathcal{D}}_{n'}) = H(\nu^{(n)}, \mathcal{D}_{qn} | \mathcal{D}_{n'}) + O(\log_2 n)$$

and we only have to show that for every $q > 1$,

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

Since for every $\varepsilon > 0$, $|H(\nu^{(n)}, \mathcal{D}_{(1-\varepsilon)n'}) - H(\nu^{(n)}, \mathcal{D}_{n'})| \leq O(\log_2 2^{\varepsilon n'})$ by Lemma 1.5 (ii.), we may write

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

and hence it will suffice for us to show that

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

For a length-n contraction ratio $t \in R_n$, let

$$\begin{aligned} \Lambda^{n,t} &= \{I \in \Lambda^n \mid r_I = t\}, \\ p^{n,t} &= \sum_{I \in \Lambda^{n,t}} p_I, \end{aligned}$$

so $(p^{n,t})_{t \in R_n}$ is a probability vector. In the following we sometimes consider $I \in \Lambda^n$ and $t \in R_n$ as random elements drawn according to the probabilities p_I and $p^{n,t}$, respectively. Then, a random $I \in \Lambda^{n,t}$ is drawn according to the conditional probability that $I \in \Lambda^n$ with a given $t = r_I$; this is $p_I / p^{n,t}$. We introduce expressions $\mathbb{P}_{I \in \Lambda^n}(\cdot), \mathbb{P}_{I \in \Lambda^{n,t}}(\cdot)$ and $\mathbb{P}_{t \in R_n}(\cdot)$ in the same manner as in Section 0. With this notation, we may consider $I \mapsto \delta_{\phi_I(0)}$ as a discrete random variable and define

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

This is clearly a probability measure on $\mathbb{R}$ and represents the part of $\nu^{(n)}$ coming from contractions with ratio t; indeed,

$$\nu^{(n)} = \sum_{t \in R_n} p^{n,t} \frac{1}{p^{n,t}} \sum_{I \in \Lambda^{n,t}} p_I \cdot \delta_{\phi_I(0)} = \mathbb{E}_{t \in R_n}(\nu^{(n,t)}).$$

For $t > 0$, define $\tau^{(t)}$ as the measure μ scaled by t;

$$\tau^{(t)}(A) = \mu(t^{-1}A).$$

Much like in the proof of the previous theorem, by rearranging the sums and using bilinearity of convolution, we have

$$\begin{aligned}
\mu(A) &= \sum_{I \in \Lambda^n} p_I \cdot \mu(r_I^{-1}(A - \phi_I(0))) \\
&= \sum_{I \in \Lambda^n} p_I \cdot (\delta_{\phi_I(0)} * \tau^{(r_I)})(A) \\
&= \sum_{t \in R_n} p^{n,t} \left(\frac{1}{p^{n,t}} \sum_{I \in \Lambda^{n,t}} p_I \cdot (\delta_{\phi_I(0)} * \tau^{(t)}) \right)(A) \\
&= \sum_{t \in R_n} p^{n,t} \left(\nu^{(n,t)} * \tau^{(t)} \right)(A) \\
&= \mathbb{E}_{t \in R_n} \left(\nu^{(n,t)} * \tau^{(t)} \right)(A).
\end{aligned}$$

Now fix an arbitrary $\varepsilon > 0$. Replacing n' with $(1-\varepsilon)n'$ in the discussion leading to (4.2) in the previous proof, we obtain

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

and using concavity of conditional entropy,

$$\alpha \geq \limsup_{n \to \infty} \frac{1}{qn - (1-\varepsilon)n'} \mathbb{E}_{t \in R_n} \left(H(\nu^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn} | \mathcal{D}_{(1-\varepsilon)n'}) \right). \tag{4.9}$$

By (4.8),

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

or, replacing r_I with t,

$$\lim_{n \to \infty} \mathbb{P}_{t \in R_n} \left(2^{-n'(1+\varepsilon)} < t < 2^{-n'(1-\varepsilon)} \right) = 1. \tag{4.10}$$

Since

$$\begin{aligned}
&\lim_{n\to\infty} \frac{1}{qn-(1-\varepsilon)n'} H(\tau^{(2^{-n'(1\pm\varepsilon)})}, \mathcal{D}_{qn}) \\
&= \lim_{n\to\infty} \frac{1}{qn-(1-\varepsilon)n'} H(\mu, \mathcal{D}_{qn-(1\pm\varepsilon)n'}) \\
&= \alpha,
\end{aligned} \tag{4.11}$$

we conclude that

$$\lim_{n\to\infty} \mathbb{P}_{t\in R_n}\left(\frac{1}{qn-(1-\varepsilon)n'} H(\tau^{(t)}, \mathcal{D}_{qn}) \geq (1-\varepsilon)\alpha\right) = 1. \tag{4.12}$$

Assuming the event of (4.10) and using the fact that $\nu_{y,n'}^{(n,t)} * \tau^{(t)}$ has its support contained in intervals with combined length of at most $O(1)\cdot 2^{-n'(1-\varepsilon)}$, we have

$$\begin{aligned}
H(\nu_{y,n'}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn}|\mathcal{D}_{(1-\varepsilon)n'}) &= H(\nu_{y,n'}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn}) - H(\nu_{y,n'}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{(1-\varepsilon)n'}) \\
&= H(\nu_{y,n'}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn}) - O(1).
\end{aligned}$$

Using the above, equations (4.9), (4.10), and concavity of conditional entropy, we get

$$\begin{aligned}
\alpha &\geq \limsup_{n\to\infty} \frac{1}{qn-(1-\varepsilon)n'} \mathbb{E}_{t\in R_n}\left(\mathbb{E}_{i=n'}\left(H(\nu_{y,i}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn}|\mathcal{D}_{(1-\varepsilon)n'}\right)\right) \\
&= \limsup_{n\to\infty} \frac{1}{qn-(1-\varepsilon)n'} \mathbb{E}_{t\in R_n}\left(\mathbb{E}_{i=n'}\left(H(\nu_{y,i}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn}\right)\right).
\end{aligned} \tag{4.13}$$

By Lemma 1.25, for every component $\nu_{y,i}^{(n,t)}$,

$$\begin{aligned}
\frac{1}{qn-(1-\varepsilon)n'} H(\nu_{y,i}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn}) \geq{}& \frac{1}{qn-(1-\varepsilon)n'} H(\tau^{(t)}, \mathcal{D}_{qn}) \\
&+ O\left(\frac{1}{qn-(1-\varepsilon)n'}\right).
\end{aligned}$$

Combining this with (4.12), we get

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

for every $\delta > 0$. Let $\delta = \delta(\varepsilon)$ be such that it tends to 0 with ε but large enough so that $\alpha \leq (1-\varepsilon)\alpha + \delta/2$. Assuming n is large, we apply Markov's inequality and (4.13) to obtain

$$
\begin{aligned}
&\mathbb{P}_{t\in R_n, i=n'}\left(\frac{1}{qn-(1-\varepsilon)n'}H(\nu_{y,i}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn}) + \delta < (1-\varepsilon)\alpha + \delta'\right) \\
&= 1 - \mathbb{P}_{t\in R_n, i=n'}\left(\frac{1}{qn-(1-\varepsilon)n'}H(\nu_{y,i}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn}) + \delta - (1-\varepsilon)\alpha \geq \delta'\right) \\
&\geq 1 - \frac{\frac{1}{qn-(1-\varepsilon)n'}\mathbb{E}_{t\in R_n}\left(\mathbb{E}_{i=n'}\left(H(\nu_{y,i}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn})\right)\right) + \delta - (1-\varepsilon)\alpha}{\delta'} \\
&\geq 1 - \frac{\delta/2}{\delta'} \\
&\geq 1 - \delta'
\end{aligned}
$$

for $\delta' = \sqrt{\delta/2}$. Hence we have

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

and replacing α with the limit from (4.11),

$$
\begin{aligned}
&\lim_{n\to\infty}\mathbb{P}_{t\in R_n, i=n'}\left(\left|\frac{1}{qn-(1-\varepsilon)n'}\left(H(\nu_{y,i}^{(n,t)} * \tau^{(t)}, \mathcal{D}_{qn}) - H(\tau^{(t)}, \mathcal{D}_{qn})\right)\right| < \delta'\right) \\
&\geq 1 - \delta' \qquad (4.14)
\end{aligned}
$$

Let $\varepsilon' > 0$. As in the previous proof, we apply Proposition 3.16 to the measure $\tau^{(n)}$ and assuming $\alpha < 1$, ε' is small and m, n are large, we obtain

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

Continuing along the lines of the proof of the previous theorem, we choose δ smaller than the constant of the same name in Theorem 1.28 and apply the theorem in question to the components $\nu_{y,i}^{(n,t)}$. We conclude that given the event of (4.14), it follows that $\frac{1}{qn-(1-\varepsilon)n'}H(\nu_{y,i}^{(n,t)}, \mathcal{D}_{qn}) < \varepsilon'$. Hence, by (4.14),

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

We then conclude, using Lemma 1.4 (vi.) and the facts that $H((p^{n,t})_{t\in R_n}) \leq$

$O(\log_2 |R_n|) \leq O(\log_2 n)$ and $H(\nu^{(n,t)}, \mathcal{D}_{(1-\varepsilon)n'}) = H(\nu^{(n,t)}, \mathcal{D}_{n'}) + O(\varepsilon n')$,

$$\begin{aligned}
&\limsup_{n\to\infty} \frac{1}{qn-(1-\varepsilon)n'} H(\nu^{(n)}, \mathcal{D}_{qn}|\mathcal{D}_{(1-\varepsilon)n'}) \\
&= \limsup_{n\to\infty} \frac{1}{qn-(1-\varepsilon)n'} H\left(\mathbb{E}_{t\in R_n}(\nu^{(n,t)}), \mathcal{D}_{qn}|\mathcal{D}_{(1-\varepsilon)n'}\right) \\
&\leq \limsup_{n\to\infty} \mathbb{E}_{t\in R_n}\left(\frac{1}{qn-(1-\varepsilon)n'} H(\nu^{(n,t)}, \mathcal{D}_{qn}|\mathcal{D}_{n'})\right) + O(\varepsilon) \\
&= \limsup_{n\to\infty} \mathbb{E}_{t\in R_n}\left(\frac{1}{qn-(1-\varepsilon)n'} \sum_{D\in\mathcal{D}_{n'}} \nu^{(n,t)}(D) H(\nu_D^{(n,t)}, \mathcal{D}_{qn})\right) + O(\varepsilon) \\
&= \limsup_{n\to\infty} \mathbb{E}_{t\in R_n}\left(\frac{1}{qn-(1-\varepsilon)n'} \mathbb{E}_{i=n'}\left(H(\nu_{y,i}^{(n,t)}, \mathcal{D}_{qn})\right)\right) + O(\varepsilon) \\
&\leq \varepsilon'(1-\delta') + \delta' \frac{1}{qn-(1-\varepsilon)n'} \log_2\left(2^{qn-n'}\right) + O(\varepsilon) \\
&\leq \varepsilon''(\varepsilon, \delta'),
\end{aligned}$$

which is what we wanted to prove. □

We can now derive Theorem 4.9.

Proof of Theorem 4.9 ([7], p. 5). Let μ be a self-similar measure with $\dim \mu < \{1, \dim_{\mathcal{S}} \mu\}$. Define $\tilde{\nu}^{(n)}$ and n' as before. Note that the conclusion of Theorem 4.12 is equivalent to

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

for every $q > 1$. Hence, by our assumption, for any q and all sufficiently large n, we have

$$\frac{1}{n'} H(\tilde{\nu}^{(n)}, \tilde{\mathcal{D}}_{qn'}) < \dim_{\mathcal{S}} \mu.$$

Since $\tilde{\nu}^{(n)} = \sum_{I\in\Lambda^n} p_I \cdot \delta_{(\phi_I(0), r_I)}$, if each pair $(\phi_I(0), r_I)$ belonged to a different atom of $\tilde{\mathcal{D}}_{qn'} = \mathcal{D}_{qn'} \times \mathcal{F}$, we would have

$$\begin{aligned}
\frac{1}{n'} H(\tilde{\nu}^{(n)}, \tilde{\mathcal{D}}_{qn'}) &\geq -\frac{1}{n \log_2(1/r)} \sum_{I \in \Lambda^n} p_I \log_2 p_I \\
&= \frac{\sum_{i\in\Lambda} \sum_{I\in\Lambda^{n-1}} p_i p_I \log_2 p_i + \sum_{i\in\Lambda} \sum_{I\in\Lambda^{n-1}} p_i p_I \log_2 p_I}{n \sum_{i\in\Lambda} p_i \log_2 r_i} \\
&= \frac{\sum_{i\in\Lambda} p_i \log_2 p_i + \sum_{I\in\Lambda^{n-1}} p_I \log_2 p_I}{n \sum_{i\in\Lambda} p_i \log_2 r_i} \\
&\vdots \\
&= \frac{n \sum_{i\in\Lambda} p_i \log_2 p_i}{n \sum_{i\in\Lambda} p_i \log_2 r_i} \\
&= \dim_{\mathcal{S}} \mu
\end{aligned}$$

which is a contradiction. Thus there must be two distinct $I, J \in \Lambda^n$ for which $(\phi_I(0), r_I)$ and $(\phi_j(0), r_J)$ lie in the same atom of $\mathcal{D}_{qn'} \times \mathcal{F}$. This means that $\Delta_n \leq |\phi_I(0) - \phi_J(0)| \leq 2^{-qn'}$ and

$$\lim_{n\to\infty} \left(-\frac{1}{n} \log_2 \Delta_n \right) \geq \lim_{n\to\infty} \left(\frac{qn'}{n} \right) = q \lfloor \log_2(1/r) \rfloor$$

for every $q > 1$. $\square$

We can also derive from Theorem 4.9 a corresponding result for the dimension of the attractor.

Corollary 4.13. *If X is the attractor of an IFS on $\mathbb{R}$ and $\dim X < \min\{1, \dim_{\mathcal{S}} X\}$, then $\lim(-\frac{1}{n}\log_2 \Delta_n) = \infty$.*

Proof. Define the self-similar measure $\mu = \sum_{i\in\Lambda} r_i^{\dim_{\mathcal{S}} X} \cdot \phi_i \mu$ with similarity dimension

$$\dim_{\mathcal{S}} \mu = \frac{\sum_{i\in\Lambda} r_i^{\dim_{\mathcal{S}} X} \dim_{\mathcal{S}} X \log_2 r_i}{\sum_{i\in\Lambda} r_i^{\dim_{\mathcal{S}} X} \log_2 r_i} = \dim_{\mathcal{S}} X.$$

Since the support of μ is contained in X, we have $\dim \mu \leq \dim X$ so by assumption, $\dim \mu < \min\{1, \dim_{\mathcal{S}} \mu\}$. Hence by Theorem 4.9, Δ_n converges to 0 super-exponentially. $\square$

4.2 Applications

Theorem 4.9 and Corollary 4.13 settle a number of cases of the Conjecture 4.8. We take a look at three examples of this. First we discuss an IFS where the parameters of the contractions are algebraic numbers and show that in this case, the dimension drop can only happen in the presence of exact overlaps. Then we return to the one dimensional Sierpinski gasket and show that while by Mastrand's projection theorem the dimension remains unchanged when projected onto the line from (Lebesgue-) almost any direction, the number of such angles is in fact countable. Finally, we show that for a family of contractions with parameters being real analytic functions of t on a compact interval, the set of t in which the dimension drop can occur has packing (and therefore also Hausdorff) dimension 0.

4.2.1 IFSs with algebraic parameters

Let r_i and a_i be algebraic numbers and consider an IFS where the contractions are of the form $\phi_i(x) = r_i x + a_i$.

Theorem 4.14. *For IFSs on $\mathbb{R}$ defined by algebraic parameters, there is a dichotomy: either there are exact overlaps, or the attractor X satisfies* $\dim X = \min\{1, \dim_{\mathcal{S}} X\}$.

Before proving this, we bring up some definitions and results from algebraic number theory. Most of the proofs are omitted.

Definition 4.15. A complex number α is *algebraic*, if there is a polynomial f with rational coefficients such that $f(\alpha) = 0$. The *degree* of α is the degree of the unique minimal, monic polynomial f_α for which $f_\alpha(\alpha) = 0$. The *conjugates* of α are the other roots of f_α.

Definition 4.16. If E is a field containing F as a subfield and $x \in E$, define the *extension field* $F(x)$ as the smallest subfield of E containing both F and x. We call the extension finite, if it is finite dimensional as a vector space over F.

Lemma 4.17. *Let α be an algebraic number of degree d. Then $\mathbb{Q}(\alpha)$ is a vector space over $\mathbb{Q}$ with basis $\{1, \alpha, \dots, \alpha^{d-1}\}$.*

Proof. See [3], Lemma 3.9. □

Lemma 4.18. *If $a_1, \dots, a_k$ are algebraic numbers, then $\mathbb{Q}(a_1, \dots, a_k) =: K$ is a finite extension of $\mathbb{Q}$. Moreover, there is a $\theta \in K$ such that $K = \mathbb{Q}(\theta)$.*

Proof. See [3], Lemmas 3.8., 3.9. and 3.10. □

Definition 4.19. The *norm* of an algebraic number α is defined as the determinant of the linear transformation $\pi_\alpha : \mathbb{Q}(\alpha) \to \mathbb{Q}(\alpha)$, $\pi_\alpha(x) = \alpha x$ and is denoted by

$$N(\alpha) = \det(\pi_\alpha).$$

Lemma 4.20. *Let α be an algebraic number and $\alpha = \alpha_1, \dots, \alpha_u$ its conjugates. Then then norm of α is an integer and is given by*

$$N(\alpha) = \prod_{i=1}^{u} \alpha_i.$$

Proof. See [10], Propositions 1.9. and 1.10. □

The final algebraic result we require is the following lemma, given by Hochman in [7].

Lemma 4.21. *Let $A \subset \mathbb{R}$ be a finite set of algebraic numbers over $\mathbb{Q}$. Then there is a constant $0 < s < 1$ such that for any degree-n polynomial P in the elements of A, either $P = 0$ or $|P| > s^n$.*

Proof ([7], p. 40). By Lemma 4.18, we may choose an algebraic number α such that $A = \{a_1, \dots, a_l\} \subset \mathbb{Q}(\alpha)$. Without changing the statement, we may multiply every element in A by an integer and by Lemma 4.17 assume that all the elements of A are integer polynomials of α of degree at most d and coefficients bounded by N for some d, N that depend only on α. Hence we may write the polynomial P as

$$P = \sum_{k=0}^{n} m_k a_k^k = \sum_{k=0}^{n} m_k \left(\sum_{i=0}^{d} n_{i,k} \alpha^i \right)^k = \sum_{k=0}^{nd} \tilde{n}_k \alpha^k,$$

where $\tilde{n}_k \in \mathbb{N}$ and $|\tilde{n}_k| \leq N$. Therefore it suffices to show that any such expression is either 0 or greater than s^n in absolute value for an $0 < s < 1$ independent of n. By replacing s with $s^{1/d}$ and changing nd to n', we may assume that $d = 1$.

Denote by $\sigma_1, \dots, \sigma_u$ the maximal set of automorphisms of $\mathbb{Q}(\alpha)$ (i.e. isomorphisms $\mathbb{Q}(\alpha) \to \mathbb{Q}(\alpha)$) with $\sigma_i(\alpha) \neq \sigma_j(\alpha)$ for distinct i, j. Note that $\sigma_i(\alpha) =: \alpha_i$ are the conjugates of α, since for the minimal monic polynomial f_α of α, $f_\alpha(\sigma_i(\alpha)) = \sigma_i(f_\alpha(\alpha)) = 0$. If $P \neq 0$, then by Lemma 4.20

$$\prod_{i=1}^{u} \sigma_i(P) = \prod_{i=1}^{u} \left(\sum_{k=0}^{n} \tilde{n}_k \sigma_i(\alpha)^k \right) \in \mathbb{Z}.$$

We conclude that

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

where $\alpha_{\max} = \max\{|\alpha_2|, \ldots, |\alpha_u|\}$. $\square$

We can now prove Theorem 4.14.

Proof of Theorem 4.14 ([7], p. 6). Let r_i, a_i be algebraic and write $\phi_i(x) = r_i x + a_i$. For all distinct multi-indices $I, J \in \Lambda^n$, the distance $|\phi_I(0) - \phi_J(0)|$ is a degree-n polynomial of r_i and a_i. Hence by Lemma 4.21 it either equals 0 or is greater than s^n for a constant $0 < s < 1$ depending only on the numbers r_i and a_i. Thus either $\Delta_n = 0$, implying an exact overlap, or $\Delta_n \geq s^n$ and

$$\lim_{n\to\infty}\left(-\frac{1}{n}\log_2 \Delta_n\right) \leq \lim_{n\to\infty}\left(\frac{s^n}{n}\right) = 0$$

and the statement follows from Corollary 4.13. $\square$

Remark 4.22. By the proof, in any IFS where the cylinders are either equal or exponentially separated, the dimension drop may only occur in the presence of exact overlaps.

4.2.2 Projecting the Sierpinski gasket

In Section 3 we considered the one dimensional Sierpinski gasket and showed that its dimension is indeed 1. We will now take a look at what happens to the dimension when we project the set onto the line.

Define the linear projection mappings $\pi_t(x, y) = tx + y$ and consider the set $F_t = \pi_t(F)$.

Lemma 4.23. *The set F_t is the attractor of an IFS Φ defined by the contractions*

$$\phi_1(x) = \frac{1}{3}x, \ \ \phi_2(x) = \frac{1}{3}(x+1), \ \ \phi_3(x) = \frac{1}{3}(x+t).$$

Proof. Recall from Section 3 that F is the attractor of

$$\left\{\theta_1 : (x, y) \mapsto \left(\tfrac{x}{3}, \tfrac{y}{3}\right), \theta_2 : (x, y) \mapsto \left(\tfrac{x+1}{3}, \tfrac{y}{3}\right), \theta_3 : (x, y) \mapsto \left(\tfrac{x}{3}, \tfrac{y+1}{3}\right)\right\}.$$

Hence

$$\pi_t(F) = \pi_t\left(\bigcup_{i=1}^{3}\theta_i(F)\right) = \bigcup_{i=1}^{3}\pi_t(\theta_i(F)),$$

where

$$\begin{aligned}
\pi_t(\theta_1(x, y)) &= \frac{tx + y}{3} = \phi_1(\pi_t(x, y)),\\
\pi_t(\theta_2(x, y)) &= \frac{tx + t + y}{3} = \phi_3(\pi_t(x, y)),\\
\pi_t(\theta_3(x, y)) &= \frac{tx + 1 + y}{3} = \phi_2(\pi_t(x, y)).
\end{aligned}$$

$\square$

Therefore the projection set F_t also has similarity dimension 1. It was conjectured by H. Furstenberg in 1970's that $\dim F_t = 1$ whenever t is irrational. Since R. Kenyon showed in [8] (Lemma 6.) that an exact overlap can only occur for certain rational values of t, this conjecture is a special case of Conjecture 4.8.

Theorem 4.24. *If $t \notin \mathbb{Q}$, then* $\dim F_t = 1$.

Proof ([7], p. 7). Fix a t and suppose that $\dim F_t < 1$. Let $\Lambda = \{0, 1, t\}$ so that $\phi_i(x) = \frac{1}{3}(x + i)$ for all $\phi_i \in \Phi$. For all multi-indices $I \in \Lambda^n$,

$$\phi_I(0) = \phi_{i_1} \circ \ldots \circ \phi_{i_n}(0) = \frac{1}{3}\left(\frac{1}{3}\left(\cdots\left(\frac{1}{3}i_n\right)\cdots + i_2\right) + i_1\right) = \sum_{k=1}^{n} i_k 3^{-k}$$

and hence

$$\begin{aligned} |\phi_I(0) - \phi_J(0)| &= \left|\sum_{k=1}^{n} i_k 3^{-k} - \sum_{k=1}^{n} j_k 3^{-k}\right| \\ &= \left|\sum_{i_k \neq t} i_k 3^{-k} - \sum_{j_k \neq t} j_k 3^{-k} - t q_{I,J}\right| \\ &= |p_{I,J} - t q_{I,J}|, \end{aligned}$$

where $q_{I,J} \geq 0$ (possibly multiplied by -1) is the sum of terms that are multiplied by t, and $p_{I,J}$ is what remains of the sum after removing $t q_{I,J}$. Both $p_{I,J}$ and $q_{I,J}$ are rational numbers belonging to the set

$$S_n = \left\{ \sum_{i=1}^{n} a_i 3^{-i} \mid a_i \in \{-1, 0, 1\} \right\}.$$

Therefore there are $p_n, q_n \in S_n$ such that $\Delta_n = |p_n - t q_n|$. By Corollary 4.13, Δ_n converges to 0 super-exponentially; particularly, for large enough n,

$$|p_n - t q_n| < 30^{-n}. \tag{4.15}$$

If $q_n = 0$ for this n, then $|p_n| < 30^{-n}$. Since p_n is a rational number with denominator 3^n, this can only happen if $p_n = 0$, implying that $\Delta_n = 0$ and $t \in \mathbb{Q}$.

Suppose then that $q_n \neq 0$ for all large n. Since q_n is rational with denominator 3^n, we have $q_n \geq 3^{-n}$. Dividing (4.15) by q^n we get $|t - p_n/q_n| < 10^{-n}$. By the triangle inequality,

$$\left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| \leq \left|t - \frac{p_{n+1}}{q_{n+1}}\right| + \left|t - \frac{p_n}{q_n}\right| < 2 \cdot 10^{-n}$$

for large enough n. Since $p_n, q_n, p_{n+1}, q_{n+1} \in S_{n+1}$, $\dfrac{p_{n+1}}{q_{n+1}} - \dfrac{p_n}{q_n}$ is rational with denominator at most $3^{n+1}3^{n+1} = 9^{n+1}$. Hence

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

But for large n, $2{\cdot}10^{-n} < 9^{-n+1}$, implying that $p_n/q_n = p_{n+1}/q_{n+1}$. Therefore, there is an N such that $|t - p_N/q_N| < 10^{-n}$ for all $n > N$, which means that $t = p_N/q_N \in \mathbb{Q}$. □

4.2.3 Parametric families of self-similar sets and measures

Let $S \subset \mathbb{R}$ be a compact interval, let $r_i : S \to (-1,1) \setminus \{0\}$ and $a_i : S \to \mathbb{R}$ for all $i \in \Lambda$. For each parameter $t \in S$, define contractions $\phi_{i,t} : \mathbb{R} \to \mathbb{R}$ by $\phi_{i,t}(x) = r_i(t)(x - a_i(t))$. For a multi-index $I \in \Lambda^n$, define $\phi_{I,t} = \phi_{i_1,t} \circ \cdots \circ \phi_{i_n,t}$ and set

$$\Delta_{I,J}(t) = \phi_{I,t}(0) - \phi_{J,t}(0).$$

If t is fixed, the distance $\Delta_n = \Delta_n(t)$ associated to an IFS $\{\phi_{i,t}\}_{i\in\Lambda}$ is always greater than the minimum of $|\Delta_{I,J}(t)|$ over distinct multi-indices $I, J \in \Lambda^n$ since it is the minimum over pairs with equal contraction ratios. Thus, if $\lim_{n\to\infty}(-\frac{1}{n}\log_2(\Delta_n)) = \infty$, the number $\min\{|\Delta_{I,J}(t)| \mid I, J \in \Lambda^n\}$ also converges to 0 super-exponentially and hence Theorem 4.9 may be used to conclude the following.

Theorem 4.25. *Let $\Phi_t = \{\phi_{i,t}\}_{i\in\Lambda}$ be a parametrized IFS as above, with r_i and a_i real analytic. For every $\varepsilon > 0$, let*

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

and

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

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

Proof. We show that if the conclusion of the theorem does not hold for a given $t \in S$, then t belongs to the set E. Indeed, assume $t \in S$ is such that $\dim \mu_t < \min\{1, \dim_{\mathcal{S}} \mu_t\}$. Then by Theorem 4.9 and the discussion above, $\min\{|\Delta_{I,J}(t)| \mid I, J \in \Lambda^n\} \to 0$ super-exponentially and for any $\varepsilon > 0$ there exists an integer $n(\varepsilon)$ such that whenever $n \geq n(\varepsilon)$, $|\Delta_{I,J}(t)| < \varepsilon^n$ for some pair $I, J \in \Lambda^n$. Hence $t \in \bigcup_{I,J\in\Lambda^n} \Delta_{I,J}^{-1}(-\varepsilon^n, \varepsilon^n)$ for all $n \geq n(\varepsilon)$; this means that $t \in E_\varepsilon$. Since ε was arbitrary, $t \in E$.

$\square$

We aim to show that the set E is small in the sense of packing dimension, implying the parameters in E are somehow "exceptional"; hence we call it the set of exceptional parameters. We extend the definition of $\Delta_{I,J}$ to infinite sequences $I, J \in \Lambda^{\mathbb{N}}$ by

$$\Delta_{I,J}(t) = \lim_{n\to\infty} \Delta_{I_n,J_n}(t),$$

where I_n and J_n are the length-n multi-indices consisting of the first n integers of I and J. The definition by limit is well-defined, since $(\phi_{I_n,t}(0))_{n\in\mathbb{N}}$ converges for all t; for any $\varepsilon > 0$,

$$|\phi_{I_n,t}(0) - \phi_{I_n,t}(\phi_{J,t}(0))| \leq |r_{I_n}(t)||\phi_{J,t}(0)| \leq \max_{i\in\Lambda, t\in S} |r_i(t)|^n < \varepsilon$$

for large enough n and any multi-index J. Since S is compact and the estimation above does not depend on the indices of I_n, the convergence is uniform over S and Λ^n. Hence, if a_i and r_i are real analytic in a neighbourhood of S, so are the functions $\Delta_{I,J}$.

Theorem 4.26. *Let $S \subset \mathbb{R}$ be a compact interval, let $r : S \to (-1,1) \setminus \{0\}$ and $a_i : S \to \mathbb{R}$ be real analytic, and let $\Phi_t = \{\phi_{i,t}\}_{i\in\Lambda}$ be the associated parametric family of iterated function systems, as above. Suppose that for all $I, J \in \Lambda^{\mathbb{N}}$,*

$$\Delta_{I,J} \equiv 0 \text{ on } S \text{ if and only if } I = J.$$

Then the set E of exceptional parameters in Theorem 4.25 has packing dimension 0.

The proof of this theorem is based on a transversality method. For multi-indices or sequences I, J, denote by $I \wedge J$ the longest common initial segment of I and J and let $|I \wedge J|$ denote its length, $|I \wedge J| = \min\{k \mid i_k \neq j_k\} - 1$. Let

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

so $0 < r_{\min} < 1$. For a k-times continuously differentiable function $F : S \to \mathbb{R}$, write $F^{(p)} = \frac{d^p}{dt^p}F$ for all $p \leq k$ and

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

Finally, we write

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

Definition 4.27. The family of iterated function systems $\{\Phi_t\}_{t\in S}$ is *transverse of order* k if r_i and a_i are k-times continuously differentiable and there is a constant $c > 0$ such that for every $n \in \mathbb{N}$ and distinct multi-indices $I, J \in \Lambda^n$,

$$|\Delta^{(p)}_{I,J}(t_0)| \geq c \cdot |I \wedge J|^{-p} \cdot r_{I\wedge J}(t_0)$$

for all $t_0 \in S$ and for $p = p(t_0) \in \{0, \dots, k\}$.

This notion of transversality is used by Hochman in [7].

Proposition 4.28. *Suppose r_i and a_i are real analytic on S. Suppose also that for $I, J \in \Lambda^{\mathbb{N}}$, $\Delta_{I,J} \equiv 0$ on S if and only if $I = J$. Then the associated family $\{\Phi_t\}_{t\in I}$ is transverse of order k for some k.*

Proof ([7], p. 38). For all i and t, write $\phi_{i,t}(x) = r_i(t)(x+a_i(t))$. Note that for all $t \in S$, we may extend r_i and a_i analytically to a complex neighbourhood U_t of t on which $|r_i|$ are still uniformly bounded away from 1; writing $r_i(z) = \sum_{k=1}^{\infty} a_k(z-t_0)^n$, r_i is complex analytic in the disc in which the power series converges, and by continuity there is an open neighbourhood in the complex plane in which $|r_i| < 1$. The function a_i extends similarly.

Define $\Delta_{I,J}(z)$ as before for multi-indices I, J and $z \in U_t$, and note that for $I, J \in \Lambda^{\mathbb{N}}$, the limit $\Delta_{I,J}(z) = \lim_{n\to\infty} \Delta_{I_n,J_n}(z)$ is uniform in U_t. Since the uniform limit of analytic functions on an open set is analytic, $\Delta_{I,J}(t)$ is real analytic on S.

Denote by U, V the sequences obtained from $I, J \in \Lambda^{\mathbb{N}}$ by removing the common part $I \wedge J$ from both of them and write

$$\tilde{\Delta}_{I,J}(t) = \Delta_{U,V}(t).$$

Then

$$\Delta_{I,J}(t) = r_{I\wedge J}(t)\tilde{\Delta}_{U,V}(t).$$

We differentiate $\tilde{\Delta}_{I,J}(t)$ p times using the general Leibniz rule and the expression above to obtain

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

A calculation shows that we can approximate the derivative of $r_{I\wedge J}(t)^{-1}$ by

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

([7], p. 38). Therefore we have the bound

$$
\begin{aligned}
|\tilde{\Delta}^{(p)}_{I,J}(t)| &\leq \sum_{q=0}^{p} \left| \binom{p}{q} \cdot \frac{d^q}{dt^q}(r_{I\wedge J}(t)^{-1}) \cdot \Delta^{(p-q)}_{I,J}(t) \right| \\
&\leq \sum_{q=0}^{p} \left| \binom{p}{q} \cdot O_{q,r_{\min},R_q}\left(|I \wedge J|^q \cdot r_{I\wedge J}(t)^{-1}\right) \cdot \Delta^{(p-q)}_{I,J}(t) \right| \\
&\leq O_{p,r_{\min},R_p}\left(\max_{0\leq q\leq p} \left(|I\wedge J|^q \cdot r_{I\wedge J}(t)^{-1} \cdot |\Delta^{(q)}_{I,J}(t)| \right) \right). \qquad (4.16)
\end{aligned}
$$

Suppose that the family $\{\Phi_t\}$ is not transverse of any order. Then by definition, for any k and $c_k > 0$, there exist an $n(k)$ and distinct $I(k), J(k) \in \Lambda^{n(k)}$, and a parameter $t_k \in S$ such that

$$
|\Delta^{(p)}_{I(k),J(k)}(t_k)| < c_k \cdot |I(k) \wedge J(k)|^{-p} \cdot r_{I(k)\wedge J(k)}(t_k)
$$

for all $0 \leq p \leq k$. Combining this with (4.16), there exists $c'_k = O_{k,R_k}(c_k)$ such that

$$
\begin{aligned}
|\tilde{\Delta}^{(p)}_{I(k),J(k)}(t_k)| &\leq O_{p,r_{\min},R_p}\left(\max_{0\leq q\leq p} \left(|I(k)\wedge J(k)|^q \cdot r_{I(k)\wedge J(k)}(t_k)^{-1} \cdot |\Delta^{(q)}_{I(k),J(k)}(t_k)| \right) \right) \\
&\leq O_{p,r_{\min},R_p}\left(\max_{0\leq q\leq p} \left(|I\wedge J|^{q-p} \cdot c_k \right) \right) \\
&\leq c'_k.
\end{aligned}
$$

For all k, choose c_k so that $c'_k < 1/k$. Denote by $U(k), V(k)$ the multi-indices obtained from $I(k)$ and $J(k)$ by removing the common part $I(k) \wedge J(k)$, so that the first elements of $U(k)$ and $V(k)$ differ and $\Delta_{U(k),V(k)} = \tilde{\Delta}_{I(k),J(k)}$. We have

$$
|\Delta^{(p)}_{U(k),V(k)}(t_k)| \leq c'_k < 1/k \text{ for all } 0 \leq p \leq k. \qquad (4.17)
$$

Since S is a compact interval, there is a converging subsequence (t_{k_l}) of (t_k). Define $t_0 = \lim_{l\to\infty} t_{k_l}$ and $U = \lim_{h\to\infty} U(k_{l_h})$, $V = \lim_{h\to\infty} V(k_{l_h})$, the latter two convergences defined in the sense that all the coordinates of the multi-indices stabilize eventually to the corresponding coordinate in the limit sequence. The subsequence (k_{l_h}) is constructed in the following way.

For $h = 1$, choose $I(k_{l_1})$ so that there are infinitely many multi-indices with the same first element as $I(k_{l_1})$ in the sequence $(I(k_l))$. If we have chosen $I(k_{l_h}) \in \Lambda^{n(k_{l_h})}$ for some h, the next element in the subsequence is then chosen to be a multi-indice of length at least $n(k_{l_h}+1)$ (we may assume that $n(k)$ increases in k) with the first $n(k_{l_h})$ elements corresponding to $I(k_{l_h})$

and with infinitely many multi-indices of the same prefix still remaining in the sequence $I(k_l)$. The sequence $(J(k_{l_h}))$ can be obtained simultaneously.

Since $U(k_{l_h})$ and $V(k_{l_h})$ differ in their first symbols for all h, so do U and V and they are distinct. As before, the convergence $\Delta_{U(k_{l_h}),V(k_{l_h})} \to \Delta_{U,V}$ is uniform in S and, since $\Delta_{U(k_{l_h}),V(k_{l_h})}$ are analytic, the same holds for the p-th derivatives. Using (4.17), we obtain

$$|\Delta^{(p)}_{U,V}(t_0)| = \lim_{h\to\infty} |\Delta^{(p)}_{U(k_{l_h}),V(k_{l_h})}(t_{k_{l_h}})| = 0$$

for all $p \geq 0$. Since $\Delta_{U,V}$ is real analytic and vanishes along with its derivatives in t_0, Liouville's theorem asserts that $\Delta_{U,V} \equiv 0$ on S, which contradicts the hypothesis. $\square$

Before we move onto the implications of transversality, we need the following lemma.

Lemma 4.29. *Let $k \in \mathbb{N}$, and let F be a k-times continuously differentiable function on a compact interval $S \subset \mathbb{R}$. Let $M \geq \|F\|_{S,k}$, and let $0 < b < 1$ be such that for every $x \in S$ there is a $p \in \{0, \ldots, k\}$ with $|F^{(p)}(x)| > b$. Then for every $0 < \rho < (b/2)^{2^k}$, the set $F^{-1}(-\rho, \rho) \subset S$ can be covered by $O_{k,M,|S|}(1/b^k)$ intervals of length at most $2(\rho/b)^{1/2^k}$ each.*

Proof ([7], p. 38). We prove the statement by induction on k. If $k = 0$, by hypothesis $|F^{(0)}(x)| = |F(x)| > b$ for all $x \in S$. Hence $F^{-1}(-\rho, \rho) = \emptyset$ for $0 < \rho < b/2 = (b/2)^{2^0}$, so the assertion is trivial.

Assume then the statement holds for $k-1$. Denote by S' a maximal closed interval in $F^{-1}[-b, b]$ and let $G = F'|_{S'}$. Assuming F satisfies the hypothesis for k, $|G^{(p)}(x)| = |F^{(p+1)}(x)| > b$ for $0 \leq p \leq k-1$ and $M \geq \|F\|_{S,k} \geq \|G\|_{S',k-1}$, so G satisfies the hypothesis for $k-1$ with the same values of b and M. Additionally, $\sqrt{b\rho} < \sqrt{\rho} < (b/2)^{2^{k-1}}$, so by the induction hypothesis we know that $G^{-1}(-\sqrt{b\rho}, \sqrt{b\rho})$ can be covered by $O_{k,M,|S|}(1/b^{k-1})$ intervals of length at most $2(\sqrt{b\rho}/b)^{1/2^{k-1}} = 2(\rho/b)^{1/2^{k-1}}$ each.

Denote by U the union of this cover and by S'' the set of the closures of the maximal subintervals in $S' \setminus U$. Since removing $O_{k,M,|S|}(1/b^{k-1})$ intervals from S' splits it into at most $O_{k,M,|S|}(1/b^{k-1}) + 1$ intervals, we know that $|S''| \leq O_{k,M,|S|}(1/b^{k-1})$. Now, on each interval in S'' we have $|G| = |F'| \geq \sqrt{b\rho}$, so by continuity of F', either $F' \geq \sqrt{b\rho}$ or $F' \leq -\sqrt{b\rho}$ in every interval of S''. Therefore, for all $I \in S''$, F is strictly monotonic in I and $I \cap F^{-1}(-\rho, \rho)$ is an interval of length at most $2\rho/\sqrt{b\rho} = 2\sqrt{\rho/b} \leq 2(\rho/b)^{1/2^k}$. As a summary of this discussion, we have covered $S' \cap F^{-1}(-\rho, \rho)$ by $O_{k,M,|S|}(1/b^{k-1})$ intervals of length $2(\rho/b)^{1/2^k}$ each.

We have yet to show that there are $O_{k,M,|S|}(1/b)$ maximal intervals $S' \subset F^{-1}([-b,b])$ such as the one in the discussion above. Assume S' is an interval of this kind. If $S' = S$, the proof is complete, since then only one interval is included in the cover. Otherwise, by continuity of F and maximality of S', there is a point $a \in S'$ with $|F(a)| = b$. There is also a point $a' \in S'$ with $|F(a')| < \rho < (b/2)^{2k}$. Since $|F'| \leq M$, we conclude that $|S'| \geq |a' - a| \geq (b-\rho)/M \geq b/2M$. Moreover, since the intervals S' are disjoint, their number is at most $|S|/(b/2M) = O_{k,M,|S|}(1/b)$. This completes the induction step. □

Theorem 4.30. *If $\{\Phi_t\}_{t\in S}$ satisfies transversality of order $k \geq 1$ on the compact interval S and the mappings r_i, a_i are real analytic, then the set E of exceptional parameters in Theorem 4.25 has packing dimension* 0.

Proof ([7], p. 39). We aim to calculate the box dimension of E and then use it in bounding the packing dimension. Write

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

Since r_i, a_i are real analytic, $\sup_{I_n,J_n\in\Lambda^n} \|\Delta_{I_n,J_n}\|_{S,k} < \infty$ for all n. As the limit function $\Delta_{I,J}$ is also real analytic for all $I, J \in \Lambda^{\mathbb{N}}$, the complex extensions of the derivatives $\Delta^{(p)}_{I_n,J_n}$, $I_n, J_n \in \Lambda^n$ converge locally uniformly to $\Delta^{(p)}_{I,J}$ and since S is compact, $\Delta^{(p)}_{I_n,J_n}(t) \to \Delta^{(p)}_{I,J}(t)$ uniformly on S. Hence M is finite.

Since the family $\{\Phi_t\}$ is transverse, there is a constant $c > 0$ such that for every $t \in S$, every n and all distinct multi-indices $I, J \in \Lambda^n$,

$$|\Delta^{(p)}_{I,J}(t)| > c \cdot |I \wedge J|^{-p} \cdot r_{\min}^{|I\wedge J|} \text{ for some } p \in \{0, \dots, k\}.$$

By replacing c with $c' < 1$, we may assume that $c < 1$.

Let n be given and fix distinct $I, J \in \Lambda^n$. Let $b = b(n) = cn^{-k} r_{\min}^n$ so that the hypothesis of Lemma 4.29 is satisfied for the function $\Delta_{I,J}$ and this b. Then, for all $0 < \rho < (b/2)^{2^k}$, the set $\{t \in S \mid |\Delta_{I,J}(t)| < \rho\}$ can be covered by $O_{k,M,c,|S|}(1/b^k)$ intervals of length $2(\rho/b)^{1/2^k}$ each.

Let now $\varepsilon > 0$ be such that $\rho = \varepsilon^n$ satisfies $\rho < (b(n)/2)^{2^k} = (cn^{-k} r_{\min}^n)^{2^k}$ for all n. By the discussion above, we find covers of at most $O_{k,M,c,|S|}(1/b^k)$ intervals for all sets $\{t \in S \mid |\Delta_{I,J}(t)| < \varepsilon^n\}$, for every distinct $I, J \in \Lambda^n$. Ranging over all such pairs we find that the set

$$E_{\varepsilon,n} := \bigcup_{I,J\in\Lambda^n, I\neq J} (\Delta_{I,J})^{-1}(-\varepsilon^n, \varepsilon^n)$$

can be covered by $O_{k,M,c,|S|}(|\Lambda|^n/b(n)^k)$ intervals of length at most $2(\varepsilon^n/b(n))^{1/2^k}$ each. Now, defining

$$E'_\varepsilon = \bigcup_{N=1}^{\infty} \bigcap_{n>N} E_{\varepsilon,n},$$

we have for the set E_ε in Theorem 4.25 that

$$E_\varepsilon = \bigcup_{N=1}^{\infty} \bigcap_{n>N} \left(\bigcup_{I,J \in \Lambda^n} (\Delta_{I,J})^{-1}(-\varepsilon^n, \varepsilon^n) \right) \subset E'_\varepsilon.$$

Hence the set of exceptional parameters $E = \bigcap_{\varepsilon>0} E_\varepsilon \subset E'_\varepsilon$. For each ε and N, adding to the error term when convenient, we have

$$\begin{aligned}
\overline{\dim}_B \left(\bigcap_{n>N} E_{\varepsilon,n} \right) &\leq \limsup_{n\to\infty} \frac{\log\left(O_{k,M,c,|S|}(|\Lambda|^n/b(n)^k)\right)}{-\log\left((\varepsilon^n/b(n))^{1/2^k}\right)} \\
&= \limsup_{n\to\infty} \frac{\log\left(O_{k,M,c,|S|}(1)(n^k)^k |\Lambda|^{nk}/r_{\min}^{nk}\right)}{O_k(1)\log\left(n^k\varepsilon^n/r_{\min}^n\right)} \\
&= O_k(1) \limsup_{n\to\infty} \frac{\log\left(n^k|\Lambda|^n/r_{\min}^n\right) + \log O_{k,M,c,|S|}(1)}{\log(n^k\varepsilon^n/r_{\min}^n)} \\
&= O_k(1) \limsup_{n\to\infty} \frac{\log(n^{k/n}|\Lambda|/r_{\min})}{\log(n^{k/n}\varepsilon/r_{\min})} \\
&= O_k(1) \frac{\log(|\Lambda|/r_{\min})}{\log(\varepsilon/r_{\min})}
\end{aligned}$$

which tends to 0 as $\varepsilon \to 0$, uniformly in N. Since $\dim_P \leq \overline{\dim}_B$, the intersection has packing dimension 0 and hence

$$\dim_P E \leq \dim_P E'_\varepsilon = \sup_N \dim_P \left(\bigcap_{n>N} E_{\varepsilon,n} \right) = 0.$$

□

We now obtain Theorem 4.26 by combining Proposition 4.28 and Theorem 4.30.

References

[1] Anca Deliu, J. S. Geronimo, R. Shonkwiler, and D. Hardin, *Dimensions associated with recurrent self-similar sets*, Math. Proc. Cambridge Philos. Soc. **110** (1991), no. 2, 327–336. MR 1113431

[2] Carl-Gustav Esseen, *On the Liapounoff limit of error in the theory of probability*, Ark. Mat. Astr. Fys. **28A** (1942), no. 9, 19. MR 0011909

[3] J-H. Evertse, *Algebraic numbers and algebraic number fields*, Unpublished lecture notes.

[4] K. Falconer, *Fractal geometry: Mathematical foundations and applications*, John Wiley & Sons Ltd, West Sussex, 1990.

[5] Ai-Hua Fan, Ka-Sing Lau, and Hui Rao, *Relationships between different dimensions of a measure*, Monatsh. Math. **135** (2002), no. 3, 191–201. MR 1897575

[6] De-Jun Feng and Huyi Hu, *Dimension theory of iterated function systems*, Comm. Pure Appl. Math. **62** (2009), no. 11, 1435–1500. MR 2560042

[7] Michael Hochman, *On self-similar sets with overlaps and inverse theorems for entropy*, Ann. of Math. (2) **180** (2014), no. 2, 773–822. MR 3224722

[8] Richard Kenyon, *Projecting the one-dimensional Sierpinski gasket*, Israel J. Math. **97** (1997), 221–238. MR 1441250

[9] Manuel "Morán and José-Manuel Rey, *Geometry of self-similar measures*, Ann. Acad. Sci. Fenn. Math. **22** (1997), no. 2, 365–386. MR 1469797

[10] F. Oggier, *Algebraic numbers and algebraic integers*, Unpublished lecture notes.

[11] Andreas Schief, *Separation properties for self-similar sets*, Proc. Amer. Math. Soc. **122** (1994), no. 1, 111–115. MR 1191872

[12] M. Simonnet, *Measures and probabilities*, Springer-Verlag New York, Inc, 1996.

[13] V. Suomala, *Measure and integration*, Unpublished lecture notes.