Additive Combinatorics and Ergodic Methods in Fractals[1]

Tuomas Sahlsten

March 2, 2017

[1]Most of the material is heavily based on Hochman's papers [Ho1, Ho2] with additional material from the books [Fa, Ma, TV], Bourgain's papers [Bo1, Bo2] and Tao's paper [Ta]

Contents

1 Additive combinatorics

1.1 Size of sumsets and the inverse problem

The **sumset** of $A, B \subset \mathbb{R}^d$, $A, B \neq \varnothing$, is defined by

$$A + B := \{a + b : a \in A, b \in B\} \subset \mathbb{R}^d.$$

One of the main topics in additive combinatorics is the following

Inverse Problem. If $A + B$ is 'small' (in cardinality, volume, dimension) compared to A and B, then what kind of structure sets A and B must have?

We will see that if $A + B$ is 'small' compared to A and B, then A and B must have some form of algebraic/arithmetic features.

Let us first look at some easy bounds for the cardinality of finite subsets of $\mathbb{R}^d$. What we can immediately do is to obtain the following easy bounds for the **cardinality** $|\cdot|$:

Lemma 1.1. *If $A, B \subset \mathbb{R}^d$ are finite and non-empty, then*

$$\max\{|A|, |B|\} \leq |A + B| \leq |A||B| \tag{1.1}$$

Remark 1.2. These inequalities can be attained:

(1) The first inequality of (1.1) is an equality if and only if A or B is a singleton.

(2) The second inequality of (1.1) is an equality if and only if every element in $A+B$ is uniquely represented, i.e. if $a + b = a' + b'$ for $a, a' \in A$ and $b, b \in B$, then $a = a'$ and $b = b'$.

An example of $|A + B| = |A||B|$ are the sets

$$A = \{0, q, q^2, \ldots, q^n\} \quad \text{and} \quad B = \{0, 1, 2, \ldots, q - 1\}$$

for fixed numbers $q, n \in \mathbb{N}$

If we consider in some sense 'random' subsets of $\mathbb{R}^d$, then one would expect that $|A+B| \approx |A||B|$. For example, if $n \in \mathbb{N}$ is fixed and $A, B \subset \{1, \ldots, n\}$ are chosen randomly such that each index $j \in \{1, \ldots, n\}$ is chosen with some probability $0 < p < 1$, then the event

$$|A + B| \geq c|A||B|$$

for some constant $c > 0$ occurs with high probability.

When happens when $|A + B| \ll |A||B|$? Let us first study what is the minimal possible values for $|A + B|$ in terms of $|A|$ and $|B|$:

1.2 Minimal growth: APs and Cauchy-Davenport inequality

A classical result in convex geometry and geometric measure theory that concerns sumsets of convex bodies in $\mathbb{R}^d$ is the **Brunn-Minkowski inquality**:

Theorem 1.3 (Brunn-Minkowski)**.** *If $A, B \subset \mathbb{R}^d$ are convex, then*

$$\mathrm{vol}(A + B) \geq (\mathrm{vol}(A)^{1/d} + \mathrm{vol}(B)^{1/d})^d.$$

Moreover, this is an equality if and only if A and B are homothetic (i.e. equal up to a translation and a dilatation).

Here we can see that the minimal possible value for $\mathrm{vol}(A+B)$ is attained when A and B are similar to each other. In additive combinatorics (of discrete sets) there is an analogue of this phenomenon. The discrete analogue of a convex set is an arithmetic progression:

Definition 1.4. Any set $P \subset \mathbb{R}$ of the form

$$P = \{a, a+p, a+2p, \dots, a+(k-1)p\}$$

for some $p \in \mathbb{N}$ and $k \in \mathbb{N}$ is called a **arithmetic progression (AP)** of **gap** p and **length** k. Moreover, an **AP in** $\mathbb{R}^d$ is any product of d arithmetic progressions in $\mathbb{R}$.

An analogue of Brunn-Minkowski inequality (in $d=1$) is the **Cauchy-Davenport inequality**:

Theorem 1.5 (Cauchy-Davenport)**.** *If $A, B \subset \mathbb{R}$ are finite with $|A|, |B| \geq 2$, then*

$$|A+B| \geq |A| + |B| - 1.$$

Moreover, this is an equality if and only if A and B are APs for the same gap.

See for example the book by Tao and Vu [TV, Proposition 5.8] for a proof.

1.3 Linear growth: GAPs and Freiman's theorem

Let us now assume $A+B$ only satisfies some linear growth with respect to A and B. To formalise this, let us first assume that $A = B$.

Definition 1.6. We say that a finite $A \subset \mathbb{R}^d$ is **small doubling** with a constant $C > 0$ if

$$|A+A| \leq C|A|.$$

Note that by definition we can choose the 'constant' $C = |A|$ so every finite A is small doubling for $C = |A|$ by the easy bound (1.1). However, the interesting case is when $C > 0$ is some fixed universal constant and $|A|$ is large.

Example 1.7. If A is an AP, then by Theorem 1.5

$$|A+A| = 2|A| - 1 \leq 2|A|,$$

so A is small doubling with $C = 2$. Moreover, in higher dimensions if $A \subset \mathbb{R}^d$ is an AP, then

$$|A+A| \leq 2^d |A|. \tag{1.2}$$

E.g. if $A = \{1, \dots, n\}^d$, then $A + A = \{2, 3, \dots, 2n\}^d$ so this inequality follows.

Often we do not have a set A, which is an AP, but still satisfies a bound with $|A+A| \leq 2^k|A|$ for some $k > d$. This leads to the definition of GAPs:

Definition 1.8. A finite set $A \subset \mathbb{R}^d$ is a **generalised arithmetic progression (GAP) of rank** $k \in \mathbb{N}$ if

$$A = \Big\{ a + \sum_{i=1}^{k} k_i p_i : k_i = 0, 1, \dots, N_i \Big\}$$

for some gaps $p_i > 0$ and $N_i \in \mathbb{N}$, $i \in \mathbb{N}$.

If A is a GAP of rank k, it will satisfy the small doubling condition

$$|A + A| \leq 2^k |A|$$

with $C = 2^k$ and are obtained with the following procedure from APs as we can see in the following:

Example 1.9. Fix some numbers $n_1, \ldots, n_k \in \mathbb{N}$, $k \in \mathbb{N}$, and write $P_i = \{1, \ldots, n_i\}$ for $i = 1, \ldots, k$. Let $T : \mathbb{R}^k \to \mathbb{R}^d$ be an affine map (i.e. $T(x) = Ax+b$ for some integer $d\times k$ matrix A and translation $b \in \mathbb{R}^d$) such that T is injective on

$$P := P_1 \times P_2 \times \cdots \times P_k.$$

Let $A \subset \mathbb{Z}^d$ be the image $A := T(P)$. Then A is a GAP of rank k. Moreover, as since $P \subset \mathbb{R}^k$ is a k-dimensional AP, the bound (1.2) yields

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

The second last equality uses the injectivity of T on P.

Small doubling also passes to 'large' subsets:

Example 1.10. Suppose A is small doubling with $C > 0$ and $A' \subset A$ satisfies $|A| \geq \varrho|A|$ for proportion $0 < \varrho \leq 1$. Then A' is small doubling with the constant C/ϱ:

$$|A' + A'| \leq |A + A| \leq C|A| \leq \frac{C}{\varrho}|A'|.$$

The first inverse theorem in the linear growth regime is **Freiman's theorem**, which shows that every finite set $A \subset \mathbb{R}^d$ with small doubling can be obtained as one of the sets obtained in Examples 1.7, 1.9 or 1.10:

Theorem 1.11 (Freiman)**.** *If $A \subset \mathbb{R}^d$ is finite and small doubling for $C > 0$, then $A \subset P$ for some GAP of rank k and cardinality $|P| \leq C'|A|$ for some $k \in \mathbb{N}$ and $C' > 0$ that only depend on C.*

Here in fact $k = O(\log_2[C(1 + \log C)])$ and $C' = O(C^{O(1)})$ as $C \to \infty$ (or at least conjectured). For a proof, see for example the book by Tao and Vu [TV, Theorem 5.32 and Theorem 5.33]. Using Plünnecke-Rusza's inequalities for iterated sumsets one can also prove now an asymmetric version we were after:

Theorem 1.12 (Asymmetric Freiman's theorem)**.** *If $A, B \subset \mathbb{R}^d$ are finite with comparable size:*

$$C^{-1} \leq |A|/|B| \leq C$$

satisfying the small doubling condition

$$|A + B| \leq C|A|,$$

then $A, B \subset P$ for some GAP of rank k and cardinality $|P| \leq C'|A|$ for some $k \in \mathbb{N}$ and $C' > 0$ that only depend on C.

1.4 Power growth: multiscale analysis and Bourgain's inverse theorem

Let $A \subset \mathbb{R}^d$ be finite. What if we relax even more and only assume the following power growth:

$$|A + A| \leq C|A|^{1+\delta} \tag{1.3}$$

for some $C > 0$ and $\delta > 0$? What can be said about the structure of A?

If we just use Freiman's theorem the best we can obtain is that A is contained in a GAP of rank $\log_2 |A|^{O(\delta)}$ as $\delta \to 0$, which does not really give any new information on A. To gain more information we can use **multiscale analysis** to establish a satisfying inverse theorem for A. This inverse theorem shows that the set A will have a special tree structure where the scales are distributed to either 'uniform' or 'singular' distribution. For this purpose, let us introduce some terminology for multiscale analysis (in $\mathbb{R}$).

Definition 1.13 (Dyadic intervals)**.** The **dyadic interval partition** $\mathcal{D}_n$ (of **generation** $n \in \mathbb{N}$) of $\mathbb{R}$ is

$$\mathcal{D}_n := \{[k2^{-n}, (k+1)2^{-n}) : k \in \mathbb{Z}\}.$$

Then $\mathcal{D}_n$ are nested in the sense that every interval $I \in \mathcal{D}_n$ splits into two subintervals of length $2^{-(n+1)}$ in $\mathcal{D}_{n+1}$. We call these two subintervals the **children** or **descendants** of I. Subintervals $J \subset I$ from $J \in \mathcal{D}_{n+m}$ are called **grandchildren** (of generation m) of I.

Looking at those dyadic subintervals $I \in \mathcal{D}_n$ which contains points from a set A can give us an idea what A looks like locally (**'magnifications'** of A). The length 2^{-n} of each interval in $\mathcal{D}_n$ could be considered as a **scale** or **resolution** we look at A. Using the dyadic intervals I we can define covering numbers:

Definition 1.14 (Covering numbers)**.** The generation $n \in \mathbb{N}$ **covering number** of a set $A \subset \mathbb{R}$ is

$$N_n(A) = \sharp\{I \in D_n : A \cap I \neq \varnothing\}.$$

Example 1.15. If $A \subset [0,1]$ and $N_n(A) = 2^n$, then every generation n subinterval of $[0,1]$ contains a point from A, which means that A is in some sense 'uniformly distributed' in the scale 2^{-n}. If $N_n(A) = 1$, then A is concentrated on a single interval $I \in \mathcal{D}_n$ so A is in some sense 'singular' in the scale 2^{-n}. The condition $N_n(A) \approx 2^{ns}$ then means that A has a 'fractal distribution' in the scale 2^{-n}.

Let us now define some quantitative notions of 'uniformity' and 'singularity' of a set A in an interval I.

Definition 1.16. Suppose we are given a finite $A \subset \mathbb{R}$, a generation $n \in \mathbb{N}$ and another generation $m \in \mathbb{N}$, and a dyadic interval $I \in \mathcal{D}_n$ such that $A \cap I \neq \varnothing$. We say that A is

(1) m**-uniform** (in I) if A has points in all the descendants of I for the next m generations:

$$\text{for all } J \in \mathcal{D}_{n+m} \text{ we have } A \cap J \neq \varnothing.;$$

(2) m**-singular** (in I) if all the points of A in I are concentrated on a single m:th generation descendant of I:

$$\text{there exists } J \in \mathcal{D}_{n+m} \text{ such that } A \cap I \subset J.$$

Remark 1.17. Using covering numbers we see that m-uniformity of A in I happens if and only if

$$N_{n+m}(A \cap I) = 2^m$$

and m-singularity of A in I happens if and only if

$$N_{n+m}(A \cap I) = 1.$$

Let us now look at a Cantor-type example of a finite set $A \subset \mathbb{R}$, where (1.3) occurs. We first recommend the reader to get familiar into the construction of the middle 1/3 Cantor set and how it is done using either trinary expansions or construction intervals.

Example 1.18 (Erdős-Volkmann 1966, Schmeling-Shmerkin 2010)**.** First define an infinite Cantor set $C \subset [0,1]$ with these construction steps using the dyadic filtration $\mathcal{D}_1, \mathcal{D}_2, \ldots$, where we name at each stage which intervals we keep and which we remove, and in the end we take an intersection.

- **Level** 1: Take $[0,1]$ and keep both subintervals of $[0,1]$ from $\mathcal{D}_1$.
- **Level** 2: Take the keps subintervals from Level 1, and only keep the lef-hand subinterval from $\mathcal{D}_2$ of each of them.
- **Level** i^2: Assuming we have constructed the intervals of generation $1, 2, \ldots, i^2 - 1$, for then next i generations keep all the subintervals in $\mathcal{D}_{i^2+j}$ of generations $i^2+1, i^2+2, \ldots, i^2+i$. I.e. we repeat Level 1 construction for i generations.
- **Level** i^2+i: Reaching generation i^2+i, for the last $i+1$ generations, keep only the left-hand subinterval of each subinterval in $\mathcal{D}_{i^2+j}$ for all generations $i^2+1, i^2+2, \ldots, i^2+i$. I.e. we repeat Level 2 construction for $i+1$ generations.

This way we have constructed collections of nested intervals $\mathcal{A}_1 \subset \mathcal{D}_1, \mathcal{A}_2 \subset \mathcal{D}_2, \ldots$. Then just define

$$C = \bigcap_{n \in \mathbb{N}} \overline{\bigcup_{I \in \mathcal{A}_n} I}.$$

Now fix $n \in \mathbb{N}$ and use the set C (or the construction of the intervals $\mathcal{A}_n$) to define $A = A_n$ as the set of left-hand end points of the intervals in $\mathcal{A}_n$.

In fact, A can be written more simply using binary expansions as

$$A = \Big\{ \sum_{i=1}^{n} x_i 2^{-i^2} : x_i \in P_i, i = 1, \ldots, n \Big\} \subset \mathbb{R},$$

where $P_i := \{1, 2, 3, \ldots, 2^i\}$ for $i = 1, \ldots, n$. From this formulation it is easier to see that A satisfies the power growth (1.3) as we will see in the following.

First of all, every $a \in A$ is uniquely represented as a sum $\sum_{i=1}^{n} x_i 2^{-i^2}$ for some $x_i \in P_i$, $i = 1, \ldots, n$ (this can be checked for example using the dyadic intervals above or the geometric sum formula). Therefore, the affine map $T : \mathbb{R}^n \to \mathbb{R}$, defined by

$$T(x) := \sum_{i=1}^{n} x_i 2^{-i^2}, \quad x = (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

is an injection on the higher dimensional arithmetic progression $P := P_1 \times P_2 \times \cdots \times P_n$. Since A is the image $A = T(P)$, we have by the proof of Example 1.9 (for $k := n$, $d := 1$, $n_i := 2^i$) that A is a GAP of rank n and satisfies the small doubling with $C = 2^n$ as follows:

$$|A + A| \leq 2^n |A|.$$

On the other hand, the cardinality

$$|A| = \prod_{i=1}^{n} |P_i| = 2^{\sum_{i=1}^{n} i} = 2^{\frac{n(n+1)}{2}} = (2^n)^{\frac{n+1}{2}}$$

which yields

$$|A + A| \leq |A|^{\frac{2}{n+1}} |A| = |A|^{1+o(1)}$$

as $n \to \infty$. Thus A satisfies the power growth (1.3) for any fixed $\delta > 0$ as long as n is large enough.

We will find out that every set A satisfying the power growth (1.3) will have a similar decomposition as A in Example 1.18 into blocks of 'uniform' and 'singular' scales. This was first made precise in Bourgain's works [Bo1, Bo2] on sum-product theory, where multiscale analysis approach was introduced (see in particular [Bo2, Sections 2, 3 and 4] for the construction and analysis). Let us now state **Bourgain's inverse theorem**:

Theorem 1.19 (Bourgain [Bo1, Bo2])**.** *Let $0 < s < 1$, $\kappa > 0$, $\varepsilon > 0$, $C > 0$ and $m \in \mathbb{N}$. Then there exists $\alpha, \varrho, \delta > 0$ such that for all large enough $n \in \mathbb{N}$ and every finite subset $A \subset \mathbb{R}$ satisfying the following conditions.*

(a) ***Separation:*** *Points in A are 2^{-n} separated:*

$$|x - y| \geq 2^{-n} \text{ for every distinct } x, y \in A;$$

(b) ***Fractal dimensionality:*** *Cardinality of A is comparable to 2^{ns}:*

$$C^{-1} 2^{ns} \leq |A| \leq C 2^{ns};$$

(c) ***Non-concentration:*** *for all integers $\varrho n < \ell < n$ we have*

$$|A \cap I| \leq 2^{-\kappa \ell} |A|$$

for all $I \in \mathcal{D}_\ell$, $I \subset [0, 1]$;

(d) ***Power growth for sumsets:***

$$|A + A| \leq |A|^{1+\delta}.$$

Then there exists $A' \subset A$ with $|A'| \geq |A|^\alpha$ and a partition $\{1, 2, \ldots, n\} = U \cup S \cup E$ satisfying

(1) *if $k \in U$, we have*

$$\frac{\sharp\{I \in \mathcal{D}_k : A' \text{ is } m\text{-uniform in } I\}}{N_k(A')} > 1 - \varepsilon;$$

(2) *if $k \in S$, we have*

$$\frac{\sharp\{I \in \mathcal{D}_k : A' \text{ is } m\text{-singular in } I\}}{N_k(A')} > 1 - \varepsilon;$$

(3) *the cardinality*

$$|E| < \varepsilon n.$$

Remark 1.20. In this result the exponent α for which $|A'| \geq |A|^\alpha$ depend implicitly on the dimension s in the sense that $\alpha \to 0$ as $s \to 0$. Hence Theorem 1.19 gives nothing on sets which are '0-dimensional' (i.e. when $1/|A|$ is asymptotically much larger to the separation 2^{-n})

Remark 1.21. In higher dimensions Bourgain's inverse theorem is not true in this form but has an analogue where concentration to affine subsets of $\mathbb{R}^d$ play an important role. See for example formulations in Bourgain's and Gamburd's work [BG] for analogues in $\mathrm{SU}(d)$.

1.5 Power growth: Hochman's inverse theorem

A recent and more statistical approach to multiscale analysis, which is also useful in our study of self-similar sets, was taken by Hochman [Ho1], where an alternative inverse theorem for sets A with $|A+A| \leq |A|^{1+\delta}$ was established (amongst other things, which we will come back to later).

The difference to Bourgain's inverse theorem is that the assumptions (b) and (c) on 'fractal-dimensinality' and 'non-concentration' were both dropped from Theorem 1.19 and it was possible to prove results for the whole set A, not just a large subset A'. However, the price that is paid to do this is to relax the assumption on 'm-singularity', which leads to the concept of (m,ε)**-concentration**.

Definition 1.22 ((m,ε)-concentration)**.** Let $n \in \mathbb{N}$, $\varepsilon > 0$ and $m \in \mathbb{N}$. We say that $A \subset \mathbb{R}$ is (m,ε)**-concentrated in** $I \in \mathcal{D}_n$ if

$$1 \leq N_{m+n}(A \cap I) \leq 2^{\varepsilon m}.$$

In other words, this means that A is (m,ε)-concentrated in $I \in \mathcal{D}_n$ if we can remove a few generation $n+m$ dyadic subintervals of I intersecting A such that the resulting set is m-singular in I. This essentially means that A nearly concentrates all the points into a single subinterval of I and only a small portion of A ('dust') is left outside.

Theorem 1.23 (Hochman [Ho1, Ho2])**.** *Let $0 < s < 1$, $\varepsilon > 0$, $C > 0$ and $m \in \mathbb{N}$. Then there exists $\delta > 0$ such that for all large enough $n \in \mathbb{N}$ and every finite subset $A \subset \mathbb{R}$ satisfying the following conditions.*

(i) ***Separation:*** *Points in A are 2^{-n} separated:*

$$|x-y| \geq 2^{-n} \text{ for every distinct } x, y \in A;$$

(ii) ***Power growth for sumsets:***

$$|A+A| \leq |A|^{1+\delta}.$$

Then there exists a partition $\{1,2,\ldots,n\} = U \cup S \cup E$ satisfying

(1) *if $k \in U$, we have*

$$\frac{\sharp\{I \in \mathcal{D}_k : A \text{ is } m\text{-uniform in } I\}}{N_k(A)} > 1-\varepsilon;$$

(2) *if $k \in S$, we have*

$$\frac{\sharp\{I \in \mathcal{D}_k : A \text{ is } (m,\varepsilon)\text{-concentrated in } I\}}{N_k(A)} > 1-\varepsilon;$$

(3) *the cardinality*

$$|E| < \varepsilon n.$$

The proof of Theorem 1.23 is to reformulate the problem (1.3) using measures. Here one replaces sumsets $A+B$ by **convolutions** $\mu * \nu$ of measures μ, ν (image of $\mu \times \nu$ under $(x,y) \mapsto x+y$) and the cardinality by the **entropy** $H(\mu, \mathcal{P})$ with respect to some partition $\mathcal{P}$. Then a multiscale analysis statement for measures can be formulated using **Hochman's inverse theorem for entropy**, which we will come back later and state precisely.

2 Additive combinatorics for measures

2.1 Convolution

The natural analogue of sumsets for measures is the convolution:

Definition 2.24. Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$. Then the **convolution** of μ and ν is the measure $\mu * \nu \in \mathcal{P}(\mathbb{R})$ defined by the push-forward of the product measure $\mu \times \nu$ under the mapping $(x, y) \mapsto x + y$.

If $+ : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is the plus mapping, then $\mu * \nu = +(\mu \times \nu)$ (push-forward), that is, for $A \subset \mathbb{R}$ we have that the convolution of a set A gives the $\mu \times \nu$ measure of the set of pairs (x, y) with the sum in A:

$$\mu * \nu(A) = \mu \times \nu(+^{-1}A) = (\mu \times \nu)(\{(x, y) \in \mathbb{R} \times \mathbb{R} : x + y \in A\}).$$

This is a very abstract way to say a simple thing. If we fix a test function $f \in C_0(\mathbb{R})$ (i.e. compactly supported and continuous f on $\mathbb{R}$), then the definition fo the convolution is equivalent to:

$$\int f \, d(\mu * \nu) = \iint f(x + y) \, d\mu(x) \, d\nu(y) \tag{2.4}$$

as $C_0(\mathbb{R})$ functions can be used to define Borel probability measures.

There is a useful way to understand convolution using translations: If $\mu \in \mathcal{P}(\mathbb{R})$, then we define the **translation** of μ by $x \in \mathbb{R}$ is the measure $\mu + x$ defined by

$$(\mu + x)(A) = \mu(A + x), \quad A \subset \mathbb{R}.$$

Example 2.25.

(1) If $\mu \in \mathcal{P}(\mathbb{R})$ and $z \in \mathbb{R}$, then the convolution by the Dirac mass δ_z is the translation:

$$\mu * \delta_z = \mu - z.$$

Recall that the Dirac mass δ_z is defined by $\delta_z(A) = 1$ if $z \in A$ and 0 if $z \notin A$, for $A \subset \mathbb{R}$. Indeed, if $f \in C_0(\mathbb{R})$ we have by the definition of convolution

$$\int f \, d(\mu * \delta_z) = \int f(x + z) \, d\mu(x) = \int f \, d(\mu - z).$$

(2) This can be generalised for any discrete (atomic) measures. If μ and ν are discrete, that is, there are finite sets $X, Y \subset \mathbb{R}$ and functions $\mu : X \to [0, 1]$ and $\nu : Y \to [0, 1]$ such that

$$\mu = \sum_{x \in X} \mu(x) \delta_x \quad \text{and} \quad \nu = \sum_{y \in Y} \nu(x) \delta_y,$$

then as in (1) we have

$$\mu * \nu = \sum_{x \in X, y \in Y} \mu(x) \nu(y) \delta_{x+y}.$$

Thus $\mu * \nu$ is a discrete measure as well on the sumset $X + Y$.

Remark 2.26. In the examples above we see that the support of $\mu * \nu$ can be obtained from the sumset of the support:

$$\operatorname{supp}(\mu * \nu) = \operatorname{supp} \mu + \operatorname{supp} \nu.$$

This is also true in general for measures $\mu, \nu \in \mathcal{P}(\mathbb{R})$ by the definition of convolution. Thus this gives the link of sumsets and convolutions.

Now in Example 2.25(2), we have for the discrete measure μ and ν that

$$\begin{aligned}\mu * \nu &= \sum_{x\in X, y\in Y} \mu(x)\nu(y)\delta_{x+y} \\ &= \sum_{y\in Y} \nu(y) \sum_{x\in X} \mu(x)\delta_{x+y} \\ &= \sum_{y\in Y} \nu(y)\mu * \delta_y \\ &= \sum_{y\in Y} \nu(y)(\mu - y).\end{aligned}$$

Here in the second equality we used the fact that

$$\mu * \delta_y = \mu - y = \sum_{x\in X} \mu(x)\delta_{x+y}.$$

Now from this above representation for $\mu * \nu$ we can see that $\mu * \nu$ is a **weighted average** (w.r.t. ν) of the translates of μ by $-y$. This can be actually made precise using probability theoretic notation. Write

$$\mu_y := \mu * \delta_y = \mu - y.$$

Then

$$X(y) := \mu_y, \quad y \in \mathbb{R},$$

is a **measure valued random variable** on the probability space $(\mathbb{R}, \operatorname{Bor}\mathbb{R}, \nu)$. Then we have that the convolution

$$\mu * \nu = \mathbb{E}(X).$$

Indeed, we for any $f \in C_0(\mathbb{R})$ we have by the definition of the translate μ_y (recall Example 2.25(1)) that

$$\int f \, d(\mu * \nu) = \iint f(x+y) \, d\mu(x) \, d\nu(y) = \iint f \, d\mu_y \, d\nu(y).$$

As this holds for any $f \in C_0(\mathbb{R})$ we have

$$\mu * \nu = \int \mu_y \, d\nu(y) = \mathbb{E}(X).$$

We will sometimes also use the notation $y \sim \nu$ to denote that y is random and distributed according to a measure $\nu \in \mathcal{P}(\mathbb{R})$. This simply means that if we have a random variable (or random measure) like μ_y and we take expectation of this, then we use ν as a probability measure and integrate over y. Then we may also write

$$\mathbb{E}_{y\sim\nu}(\cdot) := \int \cdot \, d\nu(y).$$

For example, by the above arguments we have with this notation:

Lemma 2.27. *The convolution of $\mu, \nu \in \mathcal{P}(\mathbb{R})$ satisfies*

$$\mu * \nu = \mathbb{E}_{y\sim\nu}(\mu - y) = \mathbb{E}_{x\sim\mu}(\nu - x),$$

that is, $\mu * \nu$ is the ν-weighted avarage of translations of μ or equivalently the μ-weighted average of translations of ν.

2.2 Entropy

The natural analogue for the 'size' of measures is the concept of entropy:

Definition 2.28. Let $\mu \in \mathcal{P}(\mathbb{R})$ and $n \in \mathbb{N}$. Then the **Shannon entropy** of μ with respect to $\mathcal{D}_n$ is the number

$$H(\mu, \mathcal{A}) := -\sum_{I \in \mathcal{D}_n} \mu(I) \log \mu(I).$$

Here log is base 2 and we define $0 \log 0 := 0$.

Now we have the following basic properties of Shannon entropy. In the proofs, we need to deal with more general partitions and probabilistic conditioning, so we need to talk about Shannon entropy for different partitions than just $\mathcal{D}_n$.

Definition 2.29. Let $\mu \in \mathcal{P}(\mathbb{R})$ and $\mathcal{A}$ a countable partition of $\mathbb{R}$. Then the **Shannon entropy** of μ with respect to $\mathcal{A}$ is the number

$$H(\mu, \mathcal{A}) := -\sum_{A \in \mathcal{A}} \mu(A) \log \mu(A).$$

Here log is base 2 and we define $0 \log 0 := 0$.

The relation to the previous notation is that

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

We will now refer to several (information theoretic) properties of the entropy for which proofs can be found in the literature, see for example the book by Cover and Thomas [CT].

Entropy is related to the covering numbers $N_n(X)$ mentioned before. In some sense we could say that $H(\mu, \mathcal{A})$ tells us the exponential growth rate of the number of elements $A \in \mathcal{A}$ are needed to cover the support of μ in 'in average'. If we define the following covering number for the support $\operatorname{supp} \mu$:

$$N(\mu, \mathcal{A}) := \sharp\{A \in \mathcal{A} : \mu(A) > 0\}$$

then we obtain the following bounds:

Lemma 2.30 (Cover, Thomas [CT])**.** *Let $\mu \in \mathcal{P}(\mathbb{R})$ and $\mathcal{A}$ a countable partition of $\mathbb{R}$. We have*

$$0 \leq H(\mu, \mathcal{A}) \leq \log N(\mu, \mathcal{A}). \tag{2.5}$$

Moreover, for the extremes of this inequality we have the following characterisation:

(1) *We have $H(\mu, \mathcal{A}) = 0$ if and only if μ is **concentrated** on some atom of $\mathcal{A}$, that is, $\mu(A) = 1$ for some $A \in \mathcal{A}$.*

(2) *We have $H(\mu, \mathcal{A}) = \log N(\mu, \mathcal{A})$ if and only if μ is **uniformly distributed** on the atoms of $\mathcal{A}$ with positive mass, that is,*

$$\mu(A) = \frac{1}{N(\mu, \mathcal{A})}$$

for all $A \in \mathcal{A}$ with $\mu(A) > 0$.

We also have following further properties of the entropy:

Lemma 2.31 (Cover, Thomas [CT]). *Let $\mu, \nu, \mu_i \in \mathcal{P}(\mathbb{R})$, $i \in \mathbb{N}$, and let $\mathcal{A}, \mathcal{B}$ be countable partitions of $\mathbb{R}$. Then*

(1) *Entropy is increasing under refinements: If $\mathcal{A}$ **refines** $\mathcal{B}$, that is, for any $A \in \mathcal{A}$ there exists $B \in \mathcal{B}$ with $A \subset B$, then*
$$H(\mu, \mathcal{A}) \geq H(\mu, \mathcal{B}).$$

(2) *The map $\mu \mapsto H(\mu, \mathcal{A})$ is **concave**, that is, if $t \in [0,1]$, then*
$$H(t\mu + (1-t)\nu, \mathcal{A}) \geq tH(\mu, \mathcal{A}) + (1-t)H(\nu, \mathcal{A}).$$

(3) *The map $\mu \mapsto H(\mu, \mathcal{A})$ is satisfies also the following **'convexity bound'**: fix a probability vector $p = (p_1, p_2, \ldots, p_k) \in [0,1]^k$ (i.e. $\sum p_i = 1$), then*
$$H\Big(\sum_{i=1}^{k} p_i \mu_i, \mathcal{A}\Big) \leq \sum_{i=1}^{k} p_i H(\mu_i, \mathcal{A}) + H(p),$$
where
$$H(p) := -\sum_{i=1}^{k} p_i \log p_i.$$

2.3 Conditional entropy

Let $\mu \in \mathcal{P}(\mathbb{R})$. We would like to define what it means to condition to a subsets and partitions of $\mathbb{R}$. Let $\mu(B) > 0$. Then define the **conditional measure** μ_B as follows:
$$\mu_B(A) := \frac{\mu(B \cap A)}{\mu(B)}, \quad A \subset \mathbb{R}.$$

Definition 2.32. If $\mathcal{A}, \mathcal{B}$ are countable partitions of $\mathbb{R}$, then the **conditional entropy** of μ with respect to $\mathcal{A}$ **given** $\mathcal{B}$ is
$$H(\mu, \mathcal{A}|\mathcal{B}) := \sum_{B \in \mathcal{B}} \mu(B) H(\mu_B, \mathcal{A}).$$
Thus the conditional entropy is given by the average (expectation) of the entropies of μ_B along the information source $\mathcal{B}$.

Now a very useful indentity for the conditional entropy is given by the **join partition** of $\mathcal{A}$ and $\mathcal{B}$, defined by
$$\mathcal{A} \vee \mathcal{B} := \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}.$$
We have

Lemma 2.33 (Cover, Thomas [CT]). *If $\mu \in \mathcal{P}(\mathbb{R})$ and $\mathcal{A}$ and $\mathcal{B}$ are countable partitions of $\mathbb{R}$, then*
$$H(\mu, \mathcal{A} \vee \mathcal{B}) = H(\mu, \mathcal{A}) + H(\mu, \mathcal{A}|\mathcal{B}).$$

Moreover, we have

Lemma 2.34 (Cover, Thomas [CT]). *Let $\mu \in \mathcal{P}(\mathbb{R})$ and $\mathcal{A}, \mathbb{B}$ be countable partitions of $\mathbb{R}$. We have*
$$0 \leq H(\mu, \mathcal{A}|\mathcal{B}) \leq H(\mu, \mathcal{A}). \tag{2.6}$$
Moreover, for the extremes of this inequality we have the following characterisation:

(1) *We have* $H(\mu, \mathcal{A}|\mathcal{B}) = 0$ *if and only if* $\mathcal{B}$ ***refines*** $\mathcal{A}$*, that is, for any* $B \in \mathcal{B}$ *there exists* $A \in \mathcal{A}$ *with* $B \subset A$*.*

(2) *We have* $H(\mu, \mathcal{A}|\mathcal{B}) = H(\mu, \mathcal{A})$ *if and only if* $\mathcal{A}$ *and* $\mathcal{B}$ *are* ***independent****, that is,*

$$\mu_A(B) = \mu(B)$$

for every $A \in \mathcal{A}$ *of positive measure (thus* $\mathcal{B}$ *gives no new information).*

2.4 Entropy of convolutions and the inverse problem

From now on, we will assume our measures are supported on $[0, 1]$ as this is also a natural space when we do theory for self-similar measures. For much of the theory we will now present, we mostly refer to Tao's paper [Ta] and Hochman's paper [Ho1].

Coming back to the inverse problem for sumsets, we have the following inverse problem

Inverse Problem. If $\mu * \nu$ is 'small' (in entropy or dimension) compared to μ and ν, then what kind of structure measures μ and ν must have?

As for the sumsets (recall (1.1)), we have the following easy bounds for the entropy of convolutions:

Lemma 2.35 (Tao [Ta, Lemma 2.1])**.** *If* $\mu, \nu \in \mathcal{P}[0, 1]$ *and* $n \in \mathbb{N}$*, then*

$$\max\{H_n(\mu), H_n(\nu)\} - O(1) \leq H_n(\mu * \nu) \leq H_n(\mu) + H_n(\nu) + O(1) \tag{2.7}$$

Motivated by Lemma 2.35, we now could say that $H_n(\mu * \nu)$ is 'small" if

$$H_n(\mu * \nu) \ll H_n(\mu) + H_n(\nu).$$

Let us now assume we have a single measure $\mu \in \mathcal{P}[0, 1]$ and we look at self-convolutions. Now we would like to understand the structure of measure μ if

$$H_n(\mu * \mu) \ll 2H_n(\mu).$$

2.5 Tao's inverse theorem for entropy

Tao [Ta] studied the case with the following growth condition:

$$H_n(\mu * \mu) \leq H_n(\mu) + O(1)$$

For uniformly distributed and discrete measures with enough spacing compared to 2^{-n} (e.g. each cell of $\mathcal{D}_n$ contains at most one atom of μ), this condition is equivalent to

$$|\operatorname{supp}\mu + \operatorname{supp}\mu| \leq C|\operatorname{supp}\mu|,$$

for a constant that depends on the constant in $O(1/n)$ and $0 < r < 1$ (in particular, is independent of n). This is the small doubling condition required in Freiman's theorem (Theorem 1.11). Thus Freiman's theorem yields that $\operatorname{supp}\mu$ contains a GAP as a large subset with rank controlled by C.

Therefore, one could consider the following **Tao's inverse theorem for entropy** as an analogue of Freiman's theorem for measures:

Theorem 2.36 (Tao [Ta]). *Suppose $\mu \in \mathcal{P}([0,1])$ satisfies*

$$H_n(\mu * \mu) \leq H_n(\mu) + C$$

for some $C > 1$ and $n \in \mathbb{N}$. Then μ is close (uniformly depending on C) to a discrete uniform measure on a GAP with rank at most C' that only depends on C.

Here the interesting case is when n is large and C is fixed, like in Freiman's theorem we consider large cardinality and fixed doubling constant.

Remark 2.37. Here the notion of 'close' here means with respect to a **transport distance** $\inf\{H_n(\tau) : \mu * \tau = \nu\}$ that is defined for measures μ and ν, but we will not state the definitions here precisely. We will not need Tao's results in our proof, we are just presenting some historical references without much details.

2.6 Hochman's inverse theorem for entropy

Hochman [Ho1] studied the case when $\delta > 0$ is fixed with the following growth condition:

$$H_n(\mu * \mu) \leq H_n(\mu) + \delta n. \tag{2.8}$$

For uniformly distributed and discrete measures with enough spacing compared to 2^{-n} (e.g. each cell of $\mathcal{D}_n$ contains at most one atom of μ), this condition is equivalent to

$$|\operatorname{supp}\mu + \operatorname{supp}\mu| \leq |\operatorname{supp}\mu|^{1+\delta}.$$

Thus $\operatorname{supp}\mu$ satisfies the power growth condition required in Bourgain's Inverse Theorem 1.19 so $\operatorname{supp}\mu$ (roughly) contains a large tree-like subset with a distribution of uniform and singular scales.

Analogously to Tao's setting, Hochman studied the measure theoretical analogue of Bourgain's result. We note here that Bourgain's and Hochman's results have some differences in the set theoretical setting, in some sense Hochman's theorem provides more statistical information on $\operatorname{supp}\mu$ but one needs to change (weaken) assumptions on what it means to be 'close to a singular measure'. We will possibly return to this later in the lectures.

To state Hochman's result precisely, let us introduce some notation on **scenery processes** and **multiscale analysis** for measures. Write $I_{x,n} \in \mathcal{D}_n$ as the unique 2^n-adic interval where $x \in \mathbb{R}$ belongs. If $\mu \in \mathcal{P}(\mathbb{R})$ define the **component measure (blow-up)**

$$\mu_{x,n} := \mu_{I_{x,n}}.$$

Recall that μ_B is the conditional measure $A \mapsto \mu(A \cap B)/\mu(B)$. Moreover, if $T_{x,n} : \mathbb{R} \to \mathbb{R}$ is the unique affine map mapping $I_{x,n}$ onto $[0,1)$, let

$$\mu^{x,n} := T_{x,n}\mu_{x,n}$$

be the **scenery** of μ at x in the scale 2^{-n}. I.e. the affine map $T_{x,n}$ is given by

$$T_{x,n}(y) = 2^n(y - y_{x,n}),$$

where $y_{x,n}$ is the left-hand end point of $I_{x,n}$ so in particular for $A \subset \mathbb{R}$ we have

$$\mu^{x,n}(A) = \frac{\mu(T_{x,n}^{-1}(A) \cap I_{x,n})}{\mu(I_{x,n})}.$$

Now remember that

$$\frac{1}{n}H_n(\mu) \approx 1$$

precisely when $\mu \in \mathcal{P}([0,1])$ is "**nearly uniformly distributed**" on the intervals $\mathcal{D}_n$ that meet $[0,1]$. If we fix $I \in \mathcal{D}_n$ such that $\nu \in \mathcal{P}(I)$, then

$$\frac{1}{m}H_{n+m}(\nu) \approx 1$$

precisely when ν is uniformly distributed on the $\mathcal{D}_{n+m}$ intervals meeting the 2^{-n}-interval I.

Similarly

$$\frac{1}{m}H_{n+m}(\nu) \approx 0$$

if ν is only supported on a few of the atoms in $\mathcal{D}_{n+m}$ (i.e. "**nearly singular**").

Hochman's inverse theorem for entropy states that the condition (2.8) yields that μ decomposes into essentially two kinds of scales: where the sceneries $\mu^{x,i}$ are nearly uniformly distributed (in the sense of above) and nearly singular on the other scales. There will be a remaining, negligible, collection of scales as well.

Theorem 2.38 (Hochman [Ho1])**.** *For any $\varepsilon > 0$ and $m \in \mathbb{N}$ there exists $\delta = \delta(\varepsilon, m) > 0$ such that for all large enough n and every $\mu \in \mathcal{P}([0,1])$ satisfying*

$$H_n(\mu * \mu) \leq H_n(\mu) + \delta n,$$

there exists a partition $U \cup S \cup E = \{1, 2, \ldots, n\}$ such that

(1) *if $i \in U$:*
$$\mu(\{x \in [0,1] : \tfrac{1}{m}H_m(\mu^{x,i}) > 1 - \varepsilon\}) > 1 - \varepsilon;$$

(2) *if $i \in S$:*
$$\mu(\{x \in [0,1] : \tfrac{1}{m}H_m(\mu^{x,i}) < \varepsilon\}) > 1 - \varepsilon;$$

(3) $|E| \leq \varepsilon n$.

Hochman's inverse theorem also has an asymmetric version for measures $\mu, \nu \in \mathcal{P}([0,1])$, which splits μ into uniform scales and ν into singular scales:

Theorem 2.39 (Hochman [Ho1])**.** *For any $\varepsilon > 0$ and $m \in \mathbb{N}$ there exists $\delta = \delta(\varepsilon, m) > 0$ such that for all large enough n and every $\mu, \nu \in \mathcal{P}([0,1])$ satisfying*

$$H_n(\mu * \nu) \leq H_n(\mu) + \delta n,$$

there exists a partition $U \cup S \cup E = \{1, 2, \ldots, n\}$ such that

(1) *if $i \in U$:*
$$\mu(\{x \in [0,1] : \tfrac{1}{m}H_m(\mu^{x,i}) > 1 - \varepsilon\}) > 1 - \varepsilon;$$

(2) *if $i \in S$:*
$$\nu(\{x \in [0,1] : \tfrac{1}{m}H_m(\nu^{x,i}) < \varepsilon\}) > 1 - \varepsilon;$$

(3) $|E| \leq \varepsilon n$.

2.7 Probabilistic notations on multiscale analysis

Let us now prove Hochman's theorem on self-similar measures using Hochman's inverse theorem for entropy. To do this we will first introduce some probabilistic notation on multiscale analysis that will be useful later on and restate Theorem 2.39 using this notation.

We consider $\mathbb{N}$ as the set of **generations** and each $n \in \mathbb{N}$ correspond to **scale** r^n (recall that $0 < r < 1$ is fixed), which is small when n is large. Now a given measure $\mu \in \mathcal{P}([0,1])$ induces a natural probability distribution on the component measures $\mu_{x,n}$, where x is chosen randomly with respect to μ and n uniformly randomly from $\mathbb{N}$.

Notation 2.40. Given a collection of measures $\mathcal{M} \subset \mathcal{P}([0,1])$, write the probability of $\mu_{x,n}$ being in $\mathcal{M}$ by

$$\mathbb{P}(\mu_{x,n} \in \mathcal{M}) := \mu(x \in [0,1] : \mu_{x,n} \in \mathcal{M}) = \int \mathbf{1}(\mu_{x,n} \in \mathcal{M})\, d\mu(x).$$

Moreover, for $U \subset \mathbb{N}$, let us write the average probability $\mu_{x,n} \in \mathcal{M}$ for all $n \in U$ by

$$\mathbb{P}_{n\in U}(\mu_{x,n} \in \mathcal{M}) := \frac{1}{|U|}\sum_{n\in U} \mu(x \in [0,1] : \mu_{x,n} \in \mathcal{M}) = \frac{1}{|U|}\sum_{n\in U} \mathbb{P}(\mu_{x,n} \in \mathcal{M}).$$

The expectation notation is defined similarly: if $f : \mathcal{P}([0.1]) \to \mathbb{R}$ is a Borel function, then

$$\mathbb{E}(f(\mu_{x,n})) := \int f(\mu_{x,n})\, d\mu(x)$$

and for $U \subset \mathbb{N}$ write

$$\mathbb{E}_{n\in U}(f(\mu_{x,n})) := \frac{1}{|U|}\sum_{n\in U} \mathbb{E}(f(\mu_{x,n})) = \frac{1}{|U|}\sum_{n\in U} \int f(\mu_{x,n})\, d\mu(x).$$

Now using the multiscale analysis notation Theorem 2.39 can be written as

Theorem 2.41 (Hochman [Ho1])**.** *For any $\varepsilon > 0$ and $m \in \mathbb{N}$ there exists $\delta = \delta(\varepsilon, m) > 0$ such that for all large enough n and every $\mu, \nu \in \mathcal{P}([0,1])$ satisfying*

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

there exists disjoint subsets $I, J \subset \{1, \dots, n\}$ with $|I \cup J| \ge (1-\varepsilon)n$ such that

$$\mathbb{P}(\tfrac{1}{m\log(1/r)} H(\mu_{x,i}, r^{i+m}) > 1-\varepsilon) > 1-\varepsilon, \quad \text{if } i \in I;$$

and

$$\mathbb{P}(\tfrac{1}{m\log(1/r)} H(\nu_{y,j}, r^{j+m}) < \varepsilon) > 1-\varepsilon, \quad \text{if } j \in J.$$

2.8 Entropy from component measures

Now using this notation we have the following identities:

Lemma 2.42. *If $\mu \in \mathcal{P}([0,1])$ and $n \in \mathbb{N}$, then*

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

Moreover, if $\mathcal{A}$ is a countable partition of $\mathbb{R}$, then the conditional contropy

$$H(\mu, \mathcal{A}|\mathcal{D}_n) = \mathbb{E}(H(\mu_{x,n}, \mathcal{A})).$$

Proof. The first identity is precisely the identity

$$\mu = \sum_{I \in \mathcal{D}_n} \mu(I)\mu_I,$$

(recall that μ_I is the conditional measure) which follows from $\mu = \sum_{I \in \mathcal{D}_n} \mu|_I$. For the second identity is again just another way of writing the identity

$$H(\mu, \mathcal{A}|\mathcal{D}_n) = \sum_{I \in \mathcal{D}_n} \mu(I) H(\mu_I, \mathcal{A}).$$

□

This allows us to compute the entropy using the "local" entropies of the component measures:

Lemma 2.43. *For any $\mu \in \mathcal{P}([0,1])$ and $m \in \mathbb{N}$ we have*

$$H_n(\mu) = \mathbb{E}_{1 \le i \le n}\Big(\frac{1}{m \log(1/r)} H(\mu_{x,i}, r^{i+m})\Big) + O\Big(\frac{1}{m \log(1/r)} + \frac{m}{n \log(1/r)}\Big).$$

as $n \to \infty$.

We can use this to control entropy using "number of scales" where the local entropy is small. I.e. we have the following

Lemma 2.44. *Let $\nu \in \mathcal{P}([0,1])$, $\varepsilon > 0$ and $m, n \in \mathbb{N}$ and let $J \subset \{1, \dots, n\}$ be those indices such that*

$$\mathbb{P}\Big(\frac{1}{m \log(1/r)} H(\nu_{x,j}, r^{j+m}) < \varepsilon\Big) > 1 - \varepsilon.$$

Then

$$H_n(\nu) = O(\varepsilon) + \frac{|\{1, \dots, n\} \setminus J|}{n} + O\Big(\frac{1}{m} + \frac{m}{n}\Big)$$

Now if we apply Hochman's inverse theorem for measures $\mu, \nu \in \mathcal{P}([0,1])$ satisfying the conditions of the theorem, then for the set J constructed we have the following control bound:

Lemma 2.45. *For the set J in Hochman's inverse theorem (Theorem 2.41), we have*

$$\frac{|J|}{n} = 1 - O(H_n(\nu)).$$

Thus if $\dim_{\mathrm{e}} \nu$ is close to 1 we know that asymptotically the proportion of generations in J is close to 0.

3 Proof of Hochman's inverse theorem for entropy

3.1 Probabilistic intepretation

We will now dedicate the rest of the paper to prove Hochman's inverse theorem for entropy (Theorem 2.41). Let us assume here $r = 1/2$ and log is base two. Then the formulation is

Theorem 3.46 (Hochman [Ho1])**.** *For any $\varepsilon > 0$ and $m \in \mathbb{N}$ there exists $\delta = \delta(\varepsilon, m) > 0$ such that for all large enough n and every $\mu, \nu \in \mathcal{P}([0,1])$ satisfying*

$$H_n(\mu * \nu) \le H_n(\mu) + \delta n,$$

there exists a partition $U \cup S \cup E = \{1, 2, \dots, n\}$ such that

(1) *if* $i \in U$*:*
$$\mathbb{P}^{\mu}(\tfrac{1}{m}H_m(\mu^{x,i}) > 1-\varepsilon) > 1-\varepsilon;$$

(2) *if* $i \in S$*:*
$$\mathbb{P}^{\nu}(\tfrac{1}{m}H_m(\nu^{x,i}) < \varepsilon) > 1-\varepsilon;$$

(3) $|E| \leq \varepsilon n$.

Let us now introduce some helpful probabilistic terminology. If X and Y are finite valued random variables on a probability space $(\Omega, \mathcal{F}, \mu)$, then

$$\mathcal{A} = \{X^{-1}(a) : a \in \text{range of } X\} \quad \text{and} \quad \mathcal{B} = \{Y^{-1}(b) : b \in \text{range of } X\}$$

form finite partitions of Ω. Then using our previously defined notation the entropy of the random variable X is given by

$$H(X) := H(\mu, \mathcal{A})$$

and the conditional entropy of X given Y is

$$H(X|Y) := H(\mu, \mathcal{A}|\mathcal{B}).$$

Moreover, the join entropy

$$H(X, Y) = H(\mu, \mathcal{A} \vee \mathcal{B}).$$

If μ happens to be atomic (discrete, depends on only finitely many $\omega \in \Omega$), then we just write

$$H(\mu) = \sum_{\omega \in \Omega} -\mu(\omega) \log \mu(\omega).$$

Given $k \in \mathbb{N}$ and $\nu \in \mathcal{P}(\mathbb{R})$, let us write the self-convolution

$$\nu^{*k} := \underbrace{\nu * \nu * \cdots * \nu}_{k \text{ times}}.$$

Now the main idea of the proof of Theorem 2.41 is to reduce the properties of the convolution $\mu * \nu$ to the properties of the measures $\mu * (\nu^{*k})$.

Let X_0 be a random variable on $\mathbb{R}$ whose distribution is the measure μ. Moreover, let $Y_1, Y_2, \ldots, Y_k$ be k independent copies of random variables on $\mathbb{R}$ whose distribution is ν. Then as, by definition, $\mu * (\nu^{*k})$ is the push-forward of $\mu \times \nu^k$ under the map

$$(x_0, y_1, y_2, \ldots, y_k) \mapsto x_0 + \sum_{i=1}^{k} y_i$$

we see that the distribution of the random variable $X_0 + Y_1 + \cdots + Y_k$ is $\mu * (\nu^{*k})$.

3.2 Entropy growth: Kaimanovich-Vershik lemma

In additive combinatorics, if we have a discrete sets $A, B \subset \mathbb{Z}$, we have already seen that $|A+B| \geq |A|$ and often there is growth unless A and B form arithmetic progressions of the same gap. What happens if we sum B more than once to A, that is, consider the sumset $A+B+B+\cdots+B$? Now a result by Plünnecke and Rusza tells us that in the sequence of sets $A+B, A+B+B, A+B+B+B, \ldots$ "most" of the growth to the size of the set happens during the first step. Write

$$B^{+k} := \underbrace{B + B + \cdots + B}_{k \text{ times}}.$$

Lemma 3.47 (Plünnecke-Rusza)**.** *If $A, B \subset \mathbb{Z}$ and $|A+B| \leq C|A|$, then there exists $A_0 \subset A$ with $|A_0| \geq |A|/C'$ for some constant $C' > 0$ depending on C such that*

$$|A_0 + B^{+k}| \leq C^k |A|.$$

This has an analogue for entropy.[1] As we have seen before, convolution increases entropy, that is for example for atopic measures μ and ν we have: $H(\mu * \nu) \geq H(\mu)$. Now what will happen if we convolve μ more than once with ν? Kaimanovich and Vershik established that in the sequence $\mu * \nu, \mu * \nu * \nu, \mu * \nu * \nu * \nu, \ldots$ "most" of the growth to the entropy happens during the first step.

Lemma 3.48 (Kaimanovich-Vershik)**.** *Let Γ be a countable abelian group and $\mu, \nu \in \mathcal{P}(\Gamma)$ with finite entropies. Then for any $k \in \mathbb{N}$ we have*

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

Proof. Choose a random variable X_0 in Γ with the distribution μ and let $Y_1, \ldots, Y_n$ be independent random variables in Γ with distribution ν, that is, Y_i are i.i.d. with distribution ν. Define

$$X_n := X_0 + Y_1 + \cdots + Y_n.$$

Thus X_n is a random variable with a distribution $\mu * (\nu^{*n})$. Note that (X_n) is a Markov process. Now given a realisation $Y_1 = g \in \Gamma$, since Γ is Abelian, we have

$$X_n = X_0 + g + Y_2 + \cdots + Y_n = g + (X_0 + Y_2 + \cdots + Y_n).$$

Now as the vectors $(Y_1, \ldots, Y_{n-1})$ and $(Y_2, \ldots, Y_n)$ are identically distributed (distribution is $\nu^{*(n-1)}$), we have that the random variables X_n and $X_{n-1} + g$ are identically distributed. Hence the injectivity of the translation $h \mapsto h + g$ on Γ yields that

$$H(X_n | Y_1) = H(X_{n-1}).$$

Then by the definition of conditional expectation

$$\begin{aligned} H(Y_1 | X_n) &= H(Y_1, X_n) - H(X_n) \\ &= H(Y_1) + H(X_n | Y_1) - H(X_n) \\ &= H(Y_1) + H(X_{n-1}) - H(X_n) \\ &= H(\nu) + H(\mu * \nu^{*(n-1)}) - H(\mu * \nu^{*n}). \end{aligned}$$

Given X_n, then the variable $Y_1 = X_1 - X_0$ is independent of X_{n+1} (recall that (X_n) is Markov). Hence

$$H(Y_1 | X_n) = H(Y_1 | X_n, X_{n+1}) \leq H(Y_1 | X_{n+1}).$$

Using the equality for $H(Y_1|X_n)$ (and similarly for $H(Y_1|X_{n+1})$ we obtained above to this inequality, we obtain the following growth condition:

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

which gives the claim by telescoping. □

[1] The analogue was discovered by Kaimanovich and Vershik, but was popularised again by Tao in his blog post from 2009 (see https://terrytao.wordpress.com/2009/10/27/an-entropy-plunnecke-ruzsa-inequality/).

Let us now use Kaimanovich-Vershik Lemma 3.48 to establish the following δ-entropy version that we will use:

Lemma 3.49. *If $\mu, \nu \in \mathcal{P}([0,1])$, then for any $k \in \mathbb{N}$, as $n \to \infty$ we have*

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

We will prove Lemma 3.49 by discretising the claim and reducing to Kaimanovich-Vershik lemma. Define the 2^m-adic rational numbers as

$$\Gamma_m = \Big\{\frac{k}{2^m} : k \in \mathbb{Z}\Big\}.$$

Let $\mathcal{D}_m = \mathcal{I}_{2^{-m}}$ be the dyadic intervals of $\mathbb{R}$. Given $x \in \mathbb{R}$, let $D_m(x) \in \mathcal{D}_m$ be the unique dyadic interval containing x. Then each $D \in \mathcal{D}_m$ meets precisely one point in Γ_m. Define the discretisation map $\sigma_m : \mathbb{R} \to \Gamma_m$ by

$$\sigma_m(x) = y \text{ if } D_m(x) = D_m(y),$$

where $y \in \Gamma_m$ is the unique point belonging to $D_m(y)$. That is, $\sigma_m(x)$ chooses the dyadic rational from $D_m(x)$.

Definition 3.50. We say that $\mu \in \mathcal{P}(\mathbb{R})$ is **m-discrete** if it is supported on Γ_m.

If $\mu \in \mathcal{P}(\mathbb{R})$, then μ introduces a canonical m-discrete measure as the push-forward under σ_m:

$$\sigma_m\mu = \sum_{v \in \Gamma_m} \mu(D_m(v))\delta_v.$$

Then the normalised entropy

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

The following lemma gives us an error estimate in convolutions produced by discretisation:

Lemma 3.51. *If $m \in \mathbb{N}$ and $\mu_i \in \mathcal{P}(\mathbb{R})$ with $H_m(\mu_i) < \infty$, $i = 1, \ldots, k$, then*

$$|H_m(\mu_1 * \mu_2 * \cdots * \mu_k) - H_m(\sigma_m\mu_1 * \sigma_m\mu_2 * \cdots * \sigma_m\mu_k)| = O\Big(\frac{k}{m}\Big).$$

Proof. Define the projection $\pi : \mathbb{R}^k \to \mathbb{R}$ by

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

The distance

$$|\pi(x_1, \ldots, x_k) - \pi(\sigma_m x_1, \ldots, \sigma_m x_k)| \leq \sum_{i=1}^{k} |x_i - \sigma_m(x_i)| \leq \sum_{i=1}^{k} 2^{-m} \leq k.$$

Thus the maps π and $f := \pi \circ (\sigma_m \times \cdots \times \sigma_m)$ are k apart in the sup norm. On the other hand,

$$\mu_1 * \cdots * \mu_k = \pi(\mu_1 \times \cdots \times \mu_k)$$

and

$$\sigma_m\mu_1 * \cdots * \sigma_m\mu_k = f(\mu_1 \times \cdots \times \mu_k)$$

so these convolutions are obtained as π and f push-forwards of $\mu_1 \times \cdots \times \mu_k$. Then here one can apply a standard lemma for the properties of entropy (see for example [Ho1, Lemma 3.2(3)]), which is saying that then

$$|H(\pi(\mu_1 \times \cdots \times \mu_k), \mathcal{D}_m) - H(f(\mu_1 \times \cdots \times \mu_k), \mathcal{D}_m)| \leq Ck$$

for some universal constant $C > 0$ and then divide by m to get the claim. □

Let us now prove Lemma 3.49:

Proof of Lemma 3.49. Write $\tilde{\mu} = \sigma_n \mu$ and $\tilde{\nu} = \sigma_n \nu$, which are discrete measures in the countable Abelian group Γ_n of dyadic rationals. Applying Kaimanovich-Vershik lemma (Lemma 3.48) to the measures $\tilde{\mu}, \tilde{\nu} \in \mathcal{P}(\Gamma_n)$, we have

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

Now after dividing by n, Lemma 3.51 gives the claim. □

3.3 Convolutions approximate Gaussian: Berry-Esseen theorem

Next we will give a probabilistic terminology for Gaussian measures and introduce the Berry-Esseen theorem that gives a rate for the central limit theorem for repeated convolutions.

Definition 3.52. The **mean** or the **barycenter** of $\mu \in \mathcal{P}(\mathbb{R})$ is the point $\langle \mu \rangle \in \mathbb{R}$ defined by

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

The **variance** $\mathrm{Var}(\mu)$ of μ is then given by

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

Given measures $\mu_1, \ldots, \mu_k \in \mathcal{P}(\mathbb{R})$, the convolution $\mu = \mu_1 * \cdots * \mu_k$ satisfies:

$$\langle \mu \rangle = \sum_{i=1}^{k} \langle \mu_i \rangle \quad \text{and} \quad \mathrm{Var}(\mu) = \sum_{i=1}^{k} \mathrm{Var}(\mu_i).$$

Definition 3.53. A measure $\gamma \in \mathcal{P}(\mathbb{R})$ is **Gaussian** of mean m and variance σ^2 if μ has continuous density function

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Often one denotes $\gamma = N(m, \sigma^2)$.

Note that for a Gaussian measure γ with mean m and variance σ^2, we have do have consistent definition:

$$\langle \gamma \rangle = m \quad \text{and} \quad \mathrm{Var}(\mu) = \sigma^2.$$

When $m = 0$ and $\sigma = 1$, then we call γ a **standard** Gaussian $N(0, 1)$.

Now by the central limit theorem (using the interpretation of measures with random variables), we see that if $\mu_1, \mu_2, \cdots \in \mathcal{P}(\mathbb{R})$ have all positive variance, then (after rescaling) the convolutions $\mu_1 * \mu_2 * \cdots * \mu_k$ converge weakly to a Gaussian measure. Berry-Esseen theorem gives the rate for this convergence:

Theorem 3.54 (Berry-Esseen). *Let $\mu_1, \mu_2, \ldots, \mu_k \in \mathcal{P}(\mathbb{R})$ with finite third moments:*

$$\varrho_i := \int |x|^3 \, d\mu_i(x) < \infty.$$

Write $\mu = \mu_1 * \cdots * \mu_k$. Let γ be the Gaussian measure with the mean $\langle \mu \rangle$ and variance $\mathrm{Var}(\mu)$. If $I \subset \mathbb{R}$ is an interval, then for some universal constant $C > 0$ we have

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

Now note that if $\mu_1, \ldots, \mu_k$ are supported on δ-intervals (for example are components of a measure in δ-intervals), then the moments $\varrho_i = O(\delta^3)$ and $\mathrm{Var}(\mu_i) = O(\delta^2)$. If we have that $\sum_{i=1}^k \mathrm{Var}(\mu_i) \geq c\delta^2 k$ for some universal $c > 0$, then the inequality in Berry-Esseen Theorem 3.54 has the form

$$|\mu(I) - \gamma(I)| \leq \frac{C}{c^{3/2}} \cdot \frac{1}{\sqrt{k}}.$$

3.4 Components of convolutions and Gaussian measures

We will apply Berry-Esseen theorem for the components of a fixed measure ν. This is the way we will eventually use the assumption in Hochman's inverse theorem (Theorem 2.41) to gain information about components of μ and ν. Recall the multiscale analysis notation from Section 2.7. Then by that notation any given measure $\nu \in \mathcal{P}(\mathbb{R})$ satisfies for any $i \in \mathbb{N}$ the identity (recall Lemma 2.42)

$$\nu = \mathbb{E}(\nu_{y,i})$$

(which is just the way to write $\nu = \sum_{D \in \mathcal{D}_i} \nu(D)\nu_D$, where ν_D is the conditional measure on D). Now if $k \in \mathbb{N}$, also the product measure

$$\nu^{\times k} = \mathbb{E}(\nu_{y_1,i} \times \nu_{y_2,i} \times \cdots \times \nu_{y_k,i}),$$

where the components $\nu_{y_n,i}$ are chosen independently of each other (here y_n are random variables distributed according to ν). Hence the convolution ν^{*k} can be written by the linearity of the convolution

$$\nu^{*k} = \mathbb{E}(\nu_{y_1,i} * \nu_{y_2,i} * \cdots * \nu_{y_k,i})$$

using the projection map $\pi(y_1, \ldots, y_k) = y_1 + \cdots + y_k$.

Fix any generation $i \in \mathbb{N}$ (corresponding to the scale 2^{-i}) and write

$$\sigma_i^2 := \mathbb{E}(\mathrm{Var}(\nu_{y,i})).$$

If we choose $y_1, y_2, \cdots \in \mathbb{R}$ as random variables whose distribution is ν, then the law of large numbers yields that

$$\frac{1}{k} \sum_{j=1}^k \mathrm{Var}(\nu_{y_j,i}) \longrightarrow \sigma_i^2$$

with probability 1 when $k \to \infty$ (using $\nu^{\times\infty}$ as the probability measure). Hence for large enough $k \in \mathbb{N}$, there is a high probability that

$$\mathrm{Var}(\nu_{y_1,i} * \nu_{y_2,i} * \cdots * \nu_{y_k,i}) \approx \sigma_i^2 k.$$

Then by the Berry-Esseen theorem (and the discussion after the formulation on measures supported on δ-intervals) we obtain the following corollary:

Corollary 3.55. *If $\varepsilon, \sigma > 0$ and $m \in \mathbb{N}$ is large enough, and $k \in \mathbb{N}$ large enough (depending on m), the following holds. If $i \in \mathbb{N}$ and $\nu \in \mathcal{P}([0,1])$ satisfies*

$$\mathbb{E}(\mathrm{Var}(\nu_{y,i})) \geq \sigma^2 \times (2^{-i})^2.$$

Then with probability at least $1 - \varepsilon$ when the components $\nu_{y_1,i}, \dots, \nu_{y_k,i}$ the convolution

$$\eta := \nu_{y_1,i} * \nu_{y_2,i} * \cdots * \nu_{y_k,i}$$

satisfies for all intervals $I \subset \mathbb{R}$ that

$$|\eta(I) - \gamma(I)| \leq \frac{C}{\sigma^3} \cdot \frac{1}{\sqrt{k}},$$

where γ is a Gaussian measure with mean $\langle \eta \rangle$ and variance $\mathrm{Var}(\eta)$.

3.5 Entropy of convolution via components

Now to apply Berry-Esseen theorem in the component form and to link it to the assumption

$$H_n(\mu * \nu) \leq H_n(\mu) + \delta n,$$

let us give a way to compute entropy of convolutions via the components $\mu_{x,i}$ ('sceneries'). Before doing this, let us recall Lemma 2.43, which we used in the self-similar theorem but we did not prove from earlier. We will need it in the proof of inverse theorem so we will prove it now. Here it is written again (with $r = 1/2$ and $\log = \log_2$):

Lemma 3.56. *For any $\mu \in \mathcal{P}([0,1])$ and $m \in \mathbb{N}$ we have*

$$\frac{1}{n} H_n(\mu) = \mathbb{E}_{1 \leq i \leq n}\Big(\frac{1}{m} H(\mu_{x,i}, \mathcal{D}_{i+m})\Big) + O\Big(\frac{1}{m} + \frac{m}{n}\Big)$$

as $n \to \infty$.

Proof. By Lemma 2.42 (and its proof), we recall that the expectation

$$\mathbb{E}(H(\mu_{x,j}, \mathcal{D}_{j+m})) = H(\mu, \mathcal{D}_{j+m} | \mathcal{D}_j)$$

for any $j \in \mathbb{N}$. Hence recalling the definition of $\mathbb{E}_{1 \leq i \leq n}$, the statement is equivalent to

$$\frac{1}{n} H_n(\mu) = \frac{1}{n} \sum_{i=1}^{n} \Big(\frac{1}{m} H(\mu, \mathcal{D}_{j+m} | \mathcal{D}_j)\Big) + O\Big(\frac{1}{m} + \frac{m}{n}\Big).$$

Let k be the integer part of m/n. If $0 \leq u < m$, then the entropy

$$H(\mu, \mathcal{D}_{u+mk}) = H\Big(\mu, \bigvee_{i=0}^{k} \mathcal{D}_{u+im}\Big) = H(\mu, \mathcal{D}_u) + \sum_{i=1}^{k} H(\mu, \mathcal{D}_{u+(i+1)m} | \mathcal{D}_{u+im}). \tag{3.9}$$

Now, as μ is supported on $[0,1]$ and we have $u < m$, the entropy

$$H(\mu, \mathcal{D}_u) = O(m).$$

Moreover, by the choice of k, we have $|n - (u + mk)| < m$ so

$$H(\mu, \mathcal{D}_{u+mk}) = H(\mu, \mathcal{D}_n) + O(m).$$

Thus if we divide (3.9) by m, we have proved

$$\frac{1}{m}H(\mu,\mathcal{D}_n)=\sum_{i=1}^{k}\frac{1}{m}H(\mu,\mathcal{D}_{u+(i+1)m}|\mathcal{D}_{u+im})+O(1).$$

If we now sum over $0\le u<m$ and then divide by n gives

$$\frac{1}{n}H(\mu,\mathcal{D}_n)=\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{m}H(\mu,\mathcal{D}_{j+m}|\mathcal{D}_j)\Big)+O\Big(\frac{1}{m}+\frac{m}{n}\Big),$$

which is what we claimed. □

Now we can formulate the result for entropy of convolutions:

Lemma 3.57. *For any $\mu,\nu\in\mathcal{P}([0,1])$ and $m\in\mathbb{N}$ we have*

$$\frac{1}{n}H_n(\mu*\nu)\ge\mathbb{E}_{1\le i\le n}\Big(\frac{1}{m}H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_{i+m})\Big)+O\Big(\frac{1}{m}+\frac{m}{n}\Big)$$

as $n\to\infty$. Here (x,y) is distributed according to $\mu\times\nu$ (i.e. the choice of x and y are independent of each other).

Proof. By Lemma 3.56 (and its proof above) applied to the measure $\mu*\nu$, we have

$$\frac{1}{n}H_n(\mu*\nu)=\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{m}H(\mu*\nu,\mathcal{D}_{j+m}|\mathcal{D}_j)\Big)+O\Big(\frac{1}{m}+\frac{m}{n}\Big).$$

Thus we just need to check that for any $1\le i\le n$ the conditonal entropy

$$H(\mu*\nu,\mathcal{D}_{j+m}|\mathcal{D}_j)\ge\mathbb{E}(H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_{i+m}))+O(1).$$

Recall that Lemma 2.42 proved that

$$\mu=\mathbb{E}(\mu_{x,i})\quad\text{and}\quad\nu=\mathbb{E}(\nu_{y,i}).$$

Hence by the linearity of the convolution and that x and y were chosen independently, we have

$$\mu*\nu=\mathbb{E}(\mu_{x,i}*\nu_{y,i}).$$

Conditional entropy is concave[2], so

$$H(\mu*\nu,\mathcal{D}_{i+1}|\mathcal{D}_i)=H(\mathbb{E}(\mu_{x,i}*\nu_{y,i}),\mathcal{D}_{i+1}|\mathcal{D}_i)\ge\mathbb{E}(H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_{i+1}|\mathcal{D}_i)).$$

The convolution $\mu_{x,i}*\nu_{y,i}$ is supported on an interval of length 2^{i+1} as $\mu_{x,i}\in\mathcal{P}(D_i(x))$, $\nu_{y,i}\in\mathcal{P}(D_i(y))$ and the intervals in $\mathcal{D}_i$ have length 2^{-i}. Thus $\mu_{x,i}*\nu_{y,i}$ gives at most $O(1)$ measure for any interval $I\in\mathcal{D}_{i+1}$. Therefore,

$$H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_{i+1}|\mathcal{D}_i)=H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_{i+1}\vee\mathcal{D}_i)-H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_i)\ge H(\mu_{x,i}*\nu_{y,i},\mathcal{D}_{i+1})-O(1).$$

Thus we have proved the claim. □

[2]See Lemma 2.31. Here it is written for standard entropy but the estimate there is true for conditional entropy, see [CT].

3.6 Components of absolutely continuous measures

Recall Corollary 3.55 for Berry-Esseen theorem which yields that if $\nu \in \mathcal{P}([0,1])$ is given, then the convolutions of the components $\eta = \nu_{y_1,i} * \nu_{y_2,i} * \cdots * \nu_{y_k,i}$ are close to a Gaussian measure γ of mean $\langle \eta \rangle$ and variance $\mathrm{Var}(\eta)$. Now to link this information to "uniformity" of components, which is the conclusion of Hochman's inverse theorem, we will need to give a probabilistic statement in this language of the components of Gaussian measures. Gaussian measures are absolutely continuous with continuous density, so we will give a general statement for such measures.

Lemma 3.58. *Let $\gamma \in \mathcal{P}(\mathbb{R})$ be absolutely continuous with continuous density $d\gamma(x) = f(x)\,dx$ and assume that $f > 0$. Then for any $m \in \mathbb{N}$ there exists a generation $i \in \mathbb{N}$ such that*

$$\mathbb{P}\Big(\frac{1}{m} H(\gamma_{x,i}, \mathcal{D}_{i+m}) > 1 - O(\tfrac{1}{m})\Big) > 1 - O(\tfrac{1}{m}).$$

Proof. Fix $x \in \mathbb{R}$. Let $I = D_i(x)$ giving $\gamma_{x,i} = \gamma_I$. Let $J_1, \ldots, J_\ell$ be the intervals in $\mathcal{D}_{i+m}$ in I (note that as the dyadic intervals are nested, we have $\ell = 2^m$). Hence for any $u = 1, \ldots, \ell$ we have

$$\gamma_I(J_u) = \frac{\int_{J_u} f(x)\,dx}{\int_I f(x)\,dx}.$$

Denote

$$\underline{f}_I = \inf_{x \in I} f(x) \quad \text{and} \quad \overline{f}_I = \sup_{x \in I} f(x).$$

Then

$$\frac{\underline{f}_I}{\overline{f}_I} \cdot 2^{-m} \le \frac{\underline{f}_I}{\overline{f}_I} \cdot \frac{|J_u|}{|I|} \le \gamma_I(J_u) \le \frac{\overline{f}_I}{\underline{f}_I} \cdot \frac{|J_u|}{|I|} \le \frac{\overline{f}_I}{\underline{f}_I} \cdot 2^{-m}.$$

By the continuity and as $f > 0$ we have that when $|I| \to 0$, that is, when $i \to \infty$, we have

$$\frac{\underline{f}_I}{\overline{f}_I} \to 1 \quad \text{and} \quad \frac{\overline{f}_I}{\underline{f}_I} \to 1.$$

This then shows that

$$\frac{1}{m} H(\gamma_{x,i}, \mathcal{D}_{i+m}) \to 1$$

as $i \to \infty$ for any $x \in \mathbb{R}$. Now Egorov's theorem yields that this convergence is uniform in a set of large γ measure. Thus we have proved that for any $m \in \mathbb{N}$ there exists $i \in \mathbb{N}$ such that

$$\mathbb{P}\Big(\frac{1}{m} H(\gamma_{x,i}, \mathcal{D}_{i+m}) > 1 - O(\tfrac{1}{m})\Big) > 1 - O(\tfrac{1}{m}).$$

□

Now as a corollary of this Lemma 3.58 (and its proof), we obtain

Corollary 3.59. *Let $0 < a < b$ and $\varepsilon > 0$. Then for any large enough $m \in \mathbb{N}$ there exists $i \in \mathbb{N}$ satisfying*

$$C^{-1}\sqrt{a} \le 2^{-i} \le C\sqrt{a}$$

for some uniform constant $C > 0$ such that for any Gaussian $\gamma \in \mathcal{P}(\mathbb{R})$ with $\mathrm{Var}(\gamma) \in [a,b]$ we have

$$\mathbb{P}\Big(\frac{1}{m} H(\gamma_{x,i}, \mathcal{D}_{i+m}) > 1 - \varepsilon\Big) > 1 - \varepsilon.$$

Proof. In the proof of Lemma 3.58, the i is chosen such that it only depends on the modulus of continuity of f and how fast

$$\gamma(\{x \in \mathbb{R} : f(x) > t\}) \to 1$$

when $t \to 0$. If $\mathcal{G}[a,b]$ is the family of all Gaussian measures γ with $\mathrm{Var}(\gamma) \in [a,b]$ are uniform for these quantities. Now the generation i can be chosen to be proportional to $\sqrt{a}$. This is because what we are trying to claim is invariant under re-scaling and scaling a random measure (component) by t (i.e. zooming in by 2^{-i}) yields that the variance is scaled by $\sqrt{t}$. □

3.7 Components of large self-convolutions

Let us now return to the structure of the components $\eta_i := \nu_{y_1,i} * \nu_{y_2,i} * \cdots * \nu_{y_k,i}$ of a given $\nu \in \mathcal{P}([0,1])$ (recall Corollary 3.55). Combining Corollary 3.59 to Corollary 3.55 gives us then the following claim:

Proposition 3.60. *If $\varepsilon, \sigma > 0$ and $m \in \mathbb{N}$ is large enough, and $k \in \mathbb{N}$ large enough (depending on m), the following holds. If $i \in \mathbb{N}$ and $\nu \in \mathcal{P}([0,1])$ satisfies*

$$\mathbb{E}(\mathrm{Var}(\nu_{y,i})) \geq \sigma^2 \times (2^{-i})^2.$$

Then with probability at least $1-\varepsilon$ when the components $\nu_{y_1,i}, \ldots, \nu_{y_k,i}$ the convolution

$$\eta := \nu_{y_1,i} * \nu_{y_2,i} * \cdots * \nu_{y_k,i}$$

satisfies for a given $j \in \mathbb{N}$ such that 2^{-j} is propotional to $\sqrt{\sigma^2 k} \cdot r^i$ so that

$$\mathbb{P}\Big(\frac{1}{m} H(\eta_{x,j}, \mathcal{D}_{j+m}) > 1-\varepsilon\Big) > 1-\varepsilon.$$

In particular we have

$$\mathbb{E}\Big(\frac{1}{m} H((\nu^{*k})_{y,j}, \mathcal{D}_{j+m})\Big) > 1-2\varepsilon.$$

Proof. The first statement follows from Corollary 3.55 combined with the fact that Corollary 3.59 holds for any weak limit of Gaussian measures in $\mathcal{G}[a,b]$. For the second, concavity of conditional entropy (recall Lemma 2.31) gives

$$\begin{aligned}
\mathbb{E}\Big(\frac{1}{m} H(\nu_{y,j}^k, \mathcal{D}_{j+m})\Big) &= \frac{1}{m} H(\nu^{*k}, \mathcal{D}_{j+m} | \mathcal{D}_j) \\
&= \frac{1}{m} H(\mathbb{E}(\eta), \mathcal{D}_{j+m} | \mathcal{D}_j) \\
&\geq \frac{1}{m} \mathbb{E}(H(\eta, \mathcal{D}_{j+m} | \mathcal{D}_j)) \\
&= \mathbb{E}\Big(\frac{1}{m} H(\eta_{y,j}, \mathcal{D}_{j+m})\Big) \\
&\geq 1-2\varepsilon.
\end{aligned}$$

The last inequality follows from the fact that if $X \geq 0$ is a random variable with $X > 1-\varepsilon$ with probability at least $1-\varepsilon$ then satisfies $\mathbb{E}(X) \geq 1-2\varepsilon$. □

3.8 Completion of the proof

Let us now complete the proof of Hochman's Inverse Theorem 3.46, recall once more the statement:

Theorem 3.61 (Hochman [Ho1])**.** *For any $\varepsilon > 0$ and $m \in \mathbb{N}$ there exists $\delta = \delta(\varepsilon, m) > 0$ such that for all large enough n and every $\mu, \nu \in \mathcal{P}([0,1])$ satisfying*

$$H_n(\mu * \nu) \leq H_n(\mu) + \delta n,$$

there exists disjoint subsets $I, J \subset \{1, \dots, n\}$ with $|I \cup J| \geq (1-\varepsilon)n$ such that

$$\mathbb{P}(\tfrac{1}{m} H(\mu_{x,k}, \mathcal{D}_{k+m}) > 1 - \varepsilon) > 1 - \varepsilon, \quad \text{if } k \in I;$$

and

$$\mathbb{P}(\tfrac{1}{m} H(\nu_{x,k}, \mathcal{D}_{k+m}) < \varepsilon) > 1 - \varepsilon, \quad \text{if } k \in J.$$

To make the presentation and logic as clear as possible, we will write this explicitly the main steps.

Step 1. Assumptions and parameters

We will end up applying later on the Kaimanovich-Vershik lemma (in particular, the discretised Lemma 3.49) and there will be a $k \in \mathbb{N}$ constructed at some point and to make things clear, we will now write down carefully the dependences on the quantifiers, which will lead to the right order for Hochman's Inverse Theorem 3.46. In the upcoming steps we will find the following relations on the quantifiers:

(1) $\varepsilon > 0$ arbitrary;

(2) $m \in \mathbb{N}$ large comparted to ε;

(3) $k \in \mathbb{N}$ depends on ε and m (for Kaimanovich-Vershik);

(4) $\delta > 0$ small enough depending on ε, m and k.

(5) $n \in \mathbb{N}$ large enough depending on ε, m, k and δ.

Using these relations, we fix two measures $\mu, \nu \in \mathcal{P}([0,1])$ with

$$H_n(\mu * \nu) \leq H_n(\mu) + \delta n.$$

Then we construct $\varepsilon' = \varepsilon'(\varepsilon) > 0$ which depends on the given parameters, in particular $\varepsilon > 0$ and we have

$$\lim_{\varepsilon \to 0} \varepsilon' = 0.$$

Using this $\varepsilon' > 0$ we find disjoint subsets $I, J \subset \{1, \dots, n\}$ with $|I \cup J| \geq (1-\varepsilon')n$ satisfying

$$\mathbb{P}(\tfrac{1}{m} H(\mu_{x,k}, \mathcal{D}_{k+m}) > 1 - \varepsilon') > 1 - \varepsilon', \quad \text{if } k \in I;$$

and

$$\mathbb{P}(\tfrac{1}{m} H(\nu_{x,k}, \mathcal{D}_{k+m}) < \varepsilon') > 1 - \varepsilon', \quad \text{if } k \in J.$$

Thus the conclusion of Hochman's Inverse Theorem 3.46 with ε replaced by ε'.

Remark 3.62. Note that in the relation (2) above, we have that m is large enough in relative to ε. This is not what Hochman's Inverse Theorem 3.46 gives as the relation, it just says m should be arbitrary independent of ε. This can be removed with an argument, which we omit here. See [Ho1] for details.

Step 2. Applying Kaimanovich-Vershik lemma

Suppose we have constructed $\delta > 0$ and n is large. Fix $\mu, \nu \in \mathcal{P}([0,1])$ satisfying

$$H_n(\mu * \nu) \leq H_n(\mu) + \delta n. \tag{3.10}$$

Then by the Kaimanovich-Vershik lemma (the discretised version Lemma 3.49)

$$H_n(\mu * (\nu^{*k})) \leq H_n(\mu) + k\delta + O\Big(k\Big) \tag{3.11}$$

for any $k \in \mathbb{N}$ as $n \to \infty$. On the other hand, note that entropy does not increase convolution (recall Tao's bounds Lemma 2.35) so applying Lemma 2.35 here we have the following lower bound as well:

$$H_n(\mu * (\nu^{*k})) \geq H_n(\mu) - O\Big(1\Big).$$

Note that we aim to choose $k \in \mathbb{N}$ large enough relative to ε so this means that we may assume that $k\delta$ is arbitrarily small by just choosing $\delta > 0$ small enough depending on k. Thus the inequality we obtained in (3.11) above is actually nearly an equality

$$\frac{1}{n} H_n(\mu * (\nu^{*k})) \approx \frac{1}{n} H_n(\mu).$$

The error $O(k/n)$ can be ignored as we assume that n will be chosen large in relative to k.

What have we gained in (3.11) compared to the original inequality (3.10)? The key difference here is that we convolve in (3.10) the measure μ by some arbitrary measure ν of which we do not, *a priori*, do not know much. However, in (3.11) we convolve the measure μ with the k-fold self-convolution ν^{*k}, which is no longer "arbitrary" but (as we seen using Berry-Esseen analysis) very close to a fixed given measure, namely a Gaussian measure γ. Therefore, this yields that ν^{*k} must have large entropy at many small scales. We will now make this precise and complete the proof in the following steps.

Step 3. Reformulation with component measures

Let us now see what (3.11) from Step 2 implies for the components $\mu_{x,i}$ and $(\nu^{*k})_{y,j}$. So we assume that we have

$$H_n(\mu * (\nu^{*k})) \leq H_n(\mu) + k\delta + O\Big(k\Big)$$

for a large $k \in \mathbb{N}$ but $k\delta$ small and n large enough in relation to k and δ. Let $m \in \mathbb{N}$ be large enough and later on we will assume that k is large compared to m. By Lemma 3.56 and Lemma 3.57 we thus have

$$\mathbb{E}_{1\leq i\leq n}\Big(\frac{1}{m} H(\mu_{x,i} * (\nu^{*k})_{y,i}, \mathcal{D}_{i+m})\Big) < \mathbb{E}_{1\leq i\leq n}\Big(\frac{1}{m} H(\mu_{x,i}, \mathcal{D}_{i+m})\Big) + k\delta + O\Big(\frac{k}{m} + \frac{m}{n}\Big).$$

Now note that the O error can be ignored if we just have that m is large enough (depending on ε) and n is large enough (depending on m). Now again by Tao's bounds Lemma 2.35 (for any $1 \leq i \leq n$) that:

$$\frac{1}{m} H(\mu_{x,i} * (\nu^{*k})_{y,i}, \mathcal{D}_{i+m}) \geq \frac{1}{m} H(\mu_{x,i}, \mathcal{D}_{i+m}) - O\Big(\frac{1}{m}\Big).$$

Now we are in the following situation: there are two random variables X, Y satisfying

$$\mathbb{E}(X) \leq \mathbb{E}(Y) + c \quad \text{and} \quad X \geq Y - c$$

for some small $c > 0$. Therefore, $|X - Y| \leq \sqrt{2c}$ with probability at least $1 - \sqrt{2c}$. Thus we have shown that with high probability over $i \in \{1, \dots, n\}$ and components $\mu_{x,i}$ and $(\nu^{*k})_{y,i}$ we have

$$\frac{1}{m} H(\mu_{x,i} * (\nu^{*k})_{y,i}, \mathcal{D}_{i+m}) < \frac{1}{m} H(\mu_{x,i}, \mathcal{D}_{i+m}) + \delta' \tag{3.12}$$

for some $\delta' > 0$ small as long and we choose k large, $\delta > 0$ small enough in relative to k, m large in relative to k and n large enough compared to these. Moreover, the 'high probability' also depends on the parameters in this manner: it is at least $1 - \delta'$.

What did we gain now in (3.12)? The key difference to the inequality (3.11) is that in (3.11) the k was fixed and we chose n to be large, and now we can have m to be much smaller than k. Berry-Esseen allowed us to say that ν^{*k} looks like Gaussian at scales relative to 2^{-k}, but not to scales 2^{-n} like we would like to have. Allowing k to be much bigger than m ensures that the component $(\nu^{*k})_{y,i}$ looks like Gaussian up to scales $2^{-(i+m)}$. This allows us to conclude the claim of Hochman's inverse theorem.

Step 4. Applying Berry-Esseen theorem and convolution power analysis

Let us now conclude the proof and use the Berry-Esseen theorem and convolution power analysis to construct the sets desired disjoint subsets $I, J \subset \{1, \dots, n\}$ for Hochman Inverse Theorem 3.46.

Proof of Theorem 3.46. We will construct soon a parameter $\sigma > 0$ depending on ε and m. Using this parameter we will write

$$J := \{1 \leq j \leq n : \mathbb{E}(\mathrm{Var}(\nu_{y,j})) < \sigma^2\}$$

and let

$$I := \{1, \dots, n\} \setminus J.$$

Let $j \in J$. Now, intuitively, if we have small variance for a measure τ, then most of the measure τ is supported on a very small interval. Hence choosing $\sigma > 0$ small enough (depending on ε and m), the following holds

$$\mathbb{E}\Big(\frac{1}{m} H(\nu_{y,j}, \mathcal{D}_{j+m})\Big) < \varepsilon^2$$

whenever $j \in J$. This can be done preicely using this heuristics that small variance implies concentration. Thus we have

$$\mathbb{P}\Big(\frac{1}{m} H(\nu_{y,j}, \mathcal{D}_{j+m}) > \varepsilon\Big) < \varepsilon$$

for $j \in J$. By the corollary to Berry-Esseen theorem (Proposition 3.60) using these ε and σ, we have for any $i \in I$ there is $p \in \mathbb{N}$ such that $2^{-(i-p)}$ is propotional to $\sqrt{\sigma^2 k} \cdot 2^{-i}$ and

$$\mathbb{E}\Big(\frac{1}{m} H((\nu^{*k})_{y,i-p}, \mathcal{D}_{i-p+m})\Big) > 1 - 2\varepsilon.$$

Now p grows like $\log k$ and with some effort one can show that nearly all $i \in I$ satisfy $i - p \in I$, which is done in [Ho1]. This requires us to remove a few points from I but in a way that we have at least

$$|I \cup J| > (1 - \varepsilon)n$$

and for these $i \in I$ we have

$$\mathbb{P}\Big(\frac{1}{m} H((\nu^{*k})_{y,i}, \mathcal{D}_{i+m}) > 1 - \varepsilon'\Big) > 1 - \varepsilon' \tag{3.13}$$

for some $\varepsilon' > 0$ with $\varepsilon' \to 0$ as $\varepsilon \to 0$. See [Ho1] for details on this step. By Tao's bounds Lemma 2.35 we have

$$\frac{1}{m} H(\mu_{x,i} * (\nu^{*k})_{y,i}, \mathcal{D}_{i+m}) \geq \frac{1}{m} H((\nu^{*k})_{y,i}, \mathcal{D}_{i+m}) - O\Big(\frac{1}{m}\Big).$$

If we now combine (3.12) from Step 3 and (3.13) above we have that with high probability for the components $\mu_{x,i}$ and $(\nu^{*k})_{y,i}$ we have

$$1-\varepsilon'-O\Big(\frac{1}{m}\Big) \leq \frac{1}{m} H((\nu^{*k})_{y,i}, \mathcal{D}_{i+m}) - O\Big(\frac{1}{m}\Big) \leq \frac{1}{m} H(\mu_{x,i} * (\nu^{*k})_{y,i}, \mathcal{D}_{i+m}) \leq \frac{1}{m} H(\mu_{x,i}, \mathcal{D}_{i+m}) + \delta'.$$

Thus if $i \in I$ we have with high probability for the components $\mu_{x,i}$ (i.e. $x \sim \mu$) that

$$\frac{1}{m} H(\mu_{x,i}, \mathcal{D}_{i+m}) \geq 1 - \Big(\varepsilon' + \delta' + O\Big(\frac{1}{m}\Big)\Big).$$

Thus if we begin the proof with smaller $\varepsilon > 0$ and choosing the parameters as we did in Step 1, this completes the proof. □

4 Self-similarity

Let us now discuss on some of the applications of Hochman's inverse theorem. In the following sections we will give the key appplication to the theory of self-similar sets, which has consequences and links also to other applications such as Bernoulli convolutions, projection theorems and iterated function systems contracting on average.

4.1 Self-similar sets

Let Λ be a finite index set.

Definition 4.63. A finite family $\Phi = \{f_i\}_{i\in\Lambda}$ of maps $f_i : \mathbb{R}^d \to \mathbb{R}^d$ is called an **iterated function system (IFS)** if each of the maps f_i, $i \in \Lambda$, is a contraction: there exists $0 < r_i < 1$ such that

$$|f_i(x) - f_i(y)| \leq r_i |x - y|, \quad x, y \in \mathbb{R}^d.$$

Iterated function systems can be associated with a unique compact set X, known as **attractor**:

Theorem 4.64 (Hutchinson [Hu])**.** *For any IFS $\Phi = \{f_i\}_{i\in\Lambda}$, there exists a unique compact $X \neq \varnothing$ such that*

$$X = \bigcup_{i\in\Lambda} f_i(X). \tag{4.14}$$

Proof. Let $\mathcal{X} := \{X \subset \mathbb{R}^d : X \neq \varnothing \text{ is compact}\}$. Then $\mathcal{X}$ becomes a complete metric space with the **Hausdorff metric**

$$d_H(X, Y) := \inf\{\varepsilon > 0 : X \subset Y(\varepsilon), Y \subset X(\varepsilon)\},$$

where $Y(\varepsilon) := \{x \in \mathbb{R}^d : \operatorname{dist}(x, Y) < \varepsilon\}$ is the ε-neighborhood of Y. Moreover, the transformation $\Phi : \mathcal{X} \to \mathcal{X}$,

$$\Phi(X) := \bigcup_{i\in\Lambda} f_i(X), \quad X \in \mathcal{X},$$

satisfies

$$d_H(\Phi(X), \Phi(Y)) \leq \Big(\max_{i\in\Lambda} r_i\Big) d_H(X, Y), \quad X, Y \in \mathcal{X},$$

for the parameters $0 < r_i < 1$ associated to f_i. Since $\max_{i\in\Lambda} r_i < 1$, the map Φ is a contraction on the complete metric space $(\mathcal{X}, d_H)$. Hence by the Banach fixed-point theorem there exists a unique $X \in \mathcal{X}$ satisfying $X = \Phi(X)$, which is (4.14). □

Definition 4.65. A contraction $f_i : \mathbb{R}^d \to \mathbb{R}^d$ is called a **similitude** if

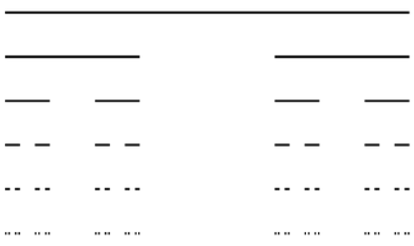
$$|f_i(x) - f_i(y)| = r_i|x - y|, \quad x, y \in \mathbb{R}^d,$$

which is equivalent to say that $f_i(x) = r_i A_i x + a_i$ for some $A_i \in O(d)$ and $a_i \in \mathbb{R}^d$. An IFS $\Phi = \{f_i\}_{i\in\Lambda}$ is **self-similar** if each $f_i \in \Phi$ is a similitude. The attractor X for a self-similar IFS is called a **self-similar set**.

Example 4.66. (1) The self-similar set X associated to the similitudes $f_i : \mathbb{R} \to \mathbb{R}$ defined by

$$f_1(x) = \tfrac{1}{3}x \quad \text{and} \quad f_2(x) = \tfrac{1}{3}x + \tfrac{2}{3}$$

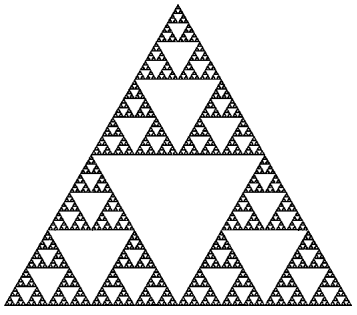
is the middle-third Cantor set.



(2) The self-similar set X associated to the similitudes $f_i : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f_1(x) = \tfrac{1}{2}x, \quad f_2(x) = \tfrac{1}{2}x + (\tfrac{1}{2}, 0) \quad \text{and} \quad f_3(x) = \tfrac{1}{2}x + (\tfrac{1}{4}, \tfrac{\sqrt{3}}{4})$$

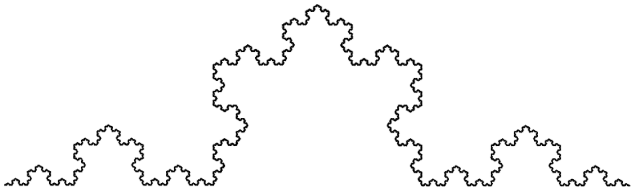
is the Sierpinski gasket.



(3) Let $A \in O(2)$ be the counter-clockwise rotation of $\mathbb{R}^2$ by the angle $60°$. Then the self-similar set X associated to the similitudes $f_i : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$f_1(x) = \tfrac{1}{3}x, \qquad f_2(x) = \tfrac{1}{3}Ax + (\tfrac{1}{3}, 0),$$
$$f_3(x) = \tfrac{1}{3}A^2x + (\tfrac{1}{2}, 0), \qquad f_4(x) = \tfrac{1}{3}x + (\tfrac{2}{3}, 0).$$

is the Koch snowflake curve.



4.2 Cylinder sets and iteration

From now on (and until the end of the lecture notes), we will study the case $d = 1$ and that the mappings in the self-similar IFS $\Phi = \{f_i\}_{i\in\Lambda}$ are all contracting with the same contraction rate $0 < r < 1$, that is, $r_i = r$ for every $i \in \Lambda$. This means that each f_i has the form

$$f_i(x) = rx + a_i$$

for some points $a_i \in \mathbb{R}$. Theory for those self-similar IFSs with varying contraction rates and in higher dimension has also been developed, but we will just concentrate on this simple case as most of the essential ideas are present there.

For a word (or n-tuple) $\mathbf{i} = i_1 i_2 \dots i_n \in \Lambda^n$ denote the composition

$$f_{\mathbf{i}} = f_{i_1} \circ f_{i_2} \circ \cdots \circ f_{i_n}.$$

In our case thus the composition

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

for $a_{\mathbf{i}} := \sum_{k=1}^{n} a_{i_k} r^{k-1} = f_{\mathbf{i}}(0)$. The set $f_{\mathbf{i}}(X)$ is called an n**th generation cylinder set** associated to the word $\mathbf{i} \in \Lambda^n$. Using (4.14) we obtain

$$X = \bigcup_{i\in\Lambda} f_i\Big(\bigcup_{j\in\Lambda} f_j(X)\Big) = \bigcup_{i,j\in\Lambda} f_i \circ f_j(X) = \bigcup_{\mathbf{i}\in\Lambda^n} f_{\mathbf{i}}(X).$$

Hence for any $n \in \mathbb{N}$ we have the following iterated version

$$X = \bigcup_{\mathbf{i}\in\Lambda^n} f_{\mathbf{i}}(X). \tag{4.15}$$

of (4.14).

4.3 Dimension

Different notions of dimensions are standard way to measure how 'big' or how much 'space' a set X occupies.

Definition 4.67. The **covering number** of a bounded set X at the scale $\varepsilon > 0$ is defined by

$$N(X, \varepsilon) := \min\{k \in \mathbb{N} : \text{we can cover } X \text{ by } k \text{ closed balls } B_i \text{ of diameter } \operatorname{diam}(B_i) = \varepsilon\}$$

The **box dimension** (or **Minkowski dimension**) of X is the exponential growth rate of the covering numbers $N(X, \varepsilon)$:

$$\dim_{\mathrm{B}} X = \lim_{\varepsilon\to 0} \frac{\log N(X, \varepsilon)}{\log(1/\varepsilon)}$$

provided that the limit exists (for self-similar X it does, to be proved later). This means that if $\alpha = \dim_{\mathrm{B}} X$, then $N(X, \varepsilon) = \varepsilon^{-\alpha+o(1)}$ as $\varepsilon \to 0$.

Example 4.68.

(1) For $X = [0, 1]$ it is a simple exercise to check $\dim_{\mathrm{B}} X = 1$.

(2) For the middle-third Cantor set one can also quite directly verify that

$$\dim_{\mathrm{B}} X = \frac{\log 2}{\log 3}$$

by using the cylinder sets for the optimal covering.

The value of box dimension is evaluated from coverings that have fixed upper bound for the size. The optimal coverings for a set may not be given by such covers. The notion of **Hausdorff dimension** (see [Fa, Ma]) $\dim X$ takes this into account:

$$\dim X = \inf\Big\{s \geq 0 : \forall \varepsilon > 0 \text{ we can cover } X \text{ by balls } B_i \text{ satisfying } \sum_i \operatorname{diam}(B_i)^s < \varepsilon\Big\}.$$

However, for self-similar sets this makes no difference:

Theorem 4.69 (Falconer [Fa])**.** *If X is self-similar, then* $\dim X = \dim_{\mathrm{B}} X$.

4.4 Bounding dimension

Let X be the self-similar set associated to $\Phi = \{f_i\}_{i\in\Lambda}$ (recall that $d = 1$ and $f_i(x) = rx + a_i$). If $\mathbf{i} \in \Lambda^n$, then as $f_{\mathbf{i}}(x) = r^n x + a_{\mathbf{i}}$ for some $a_{\mathbf{i}} \in \mathbb{R}$, we know that

$$\operatorname{diam} f_{\mathbf{i}}(X) \leq r^n \operatorname{diam} X.$$

Hence if we set $\varepsilon = r^n \operatorname{diam} X$, we can use (4.15) to establish

$$N(X, \varepsilon) \leq |\Lambda|^n.$$

(take an interval I containing X and use the interval cover $f_{\mathbf{i}}(I)$ of X). Hence

$$\dim X = \dim_{\mathrm{B}} X \leq \lim_{n\to\infty} \frac{\log |\Lambda|^n}{\log(\frac{1}{r^n \operatorname{diam} X})} = \frac{\log |\Lambda|}{\log(1/r)}.$$

Definition 4.70. The **similarity dimension** of the IFS $\Phi = \{f_i\}_{i\in\Lambda}$ is the number

$$\dim_{\mathrm{s}} \Phi := \frac{\log |\Lambda|}{\log(1/r)}.$$

Since we trivially have $\dim X \leq 1$, we obtain an upper bound:

$$\dim X \leq \min\{1, \dim_{\mathrm{s}} \Phi\}. \tag{4.16}$$

4.5 When is $\dim X = \min\{1, \dim_{\mathrm{s}} \Phi\}$?

The equality for (4.16) occurs if there is separation present in the choices of the maps f_i in the IFS Φ.

Definition 4.71. We say that Φ satisfies the

(1) **strong separation condition (SSC)**, if the union $\bigcup_{i\in\Lambda} f_i(X)$ is disjoint;

(2) **open set condition (OSC)**, if there exists a non-empty open set $U \subset \mathbb{R}$ such that $f_i(U) \subset U$ and the union $\bigcup_{i\in\Lambda} f_i(U)$ is disjoint.

Example 4.72. The middle-third Cantor set satisfies SSC and the Sierpinski gasket satisfies OSC but not the SSC.

If we have one of these separation conditions, we can obtain equality in (4.16):

Theorem 4.73 (Hutchinson [Hu])**.** *If Φ satisfies the SSC or OSC, then the self-similar set X satisfies*

$$\dim X = \dim_s \Phi.$$

Note that in particular SSC and OSC always fail if we have $\dim_s \Phi > 1$, which can be achieved if the maps in Φ 'overlap' as we will see in the next section.

4.6 When is $\dim X < \min\{1, \dim_s \Phi\}$?

Let us define the notion of 'exact' overlaps:

Definition 4.74. If there exists $n \in \mathbb{N}$ and $\mathbf{i}, \mathbf{j} \in \Lambda^n$, $\mathbf{i} \neq \mathbf{j}$, such that

$$f_{\mathbf{i}} = f_{\mathbf{j}},$$

then Φ is said to have **exact overlaps**.

Note that as $f_{\mathbf{i}}(x) = r^n x + a_{\mathbf{i}}$, having exact overlaps means that $a_{\mathbf{i}} = a_{\mathbf{j}}$. Suppose now that Φ has exact overlaps with some $\mathbf{i}, \mathbf{j} \in \Lambda^n$. Then in (4.15) we may omit one of the $f_{\mathbf{i}}(X)$ or $f_{\mathbf{j}}(X)$. In particular, if we put $\Lambda' = \Lambda^n \setminus \{\mathbf{j}\}$, then

$$X = \bigcup_{\mathbf{k}\in\Lambda'} f_{\mathbf{k}}(X).$$

Hence X is self-similar for the IFS $\Phi' = \{f_{\mathbf{k}}\}_{\mathbf{k}\in\Lambda'}$. Each $f_{\mathbf{k}}(x) = r^n x + a_{\mathbf{k}}$ and $|\Lambda'| = |\Lambda|^n - 1$, so (4.16) yields

$$\dim X \leq \dim_s \Phi' = \frac{\log(|\Lambda|^n - 1)}{\log(1/r^n)} < \frac{\log |\Lambda|^n}{\log(1/r^n)} = \dim_s \Phi.$$

Thus if we have $\dim_s \Phi < 1$, we obtain $\dim X < \min\{1, \dim_s \Phi\}$.

4.7 The dimension drop conjecture

Now it is expected that the drop in dimension in (4.16) should only occur when there are exact overlaps in the IFS:

Conjecture 4.75. *If $\dim X < \min\{1, \dim_s \Phi\}$, then Φ has exact overlaps.*

This conjecture is still open but recently substantial progress was made towards this conjecture by M. Hochman [Ho1].

Define

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

The choice of initial point 0 is irrelevant and we could put there any $x \in \mathbb{R}$ since $f_{\mathbf{i}}(x) = r^n x + a_{\mathbf{i}}$. In fact $\Delta_n = \min\{|a_{\mathbf{i}} - a_{\mathbf{j}}| : \mathbf{i}, \mathbf{j} \in \Lambda^n, \mathbf{i} \neq \mathbf{j}\}$.

Remark 4.76. (i) Φ has exact overlaps if and only if $\Delta_n = 0$. Thus Conjecture 4.75 is equivalent to asking that $\dim X < \min\{1, \dim_s \Phi\}$ yields $\Delta_n = 0$.

(ii) If $x \in X$, then each $f_{\mathbf{i}}(x) \in X$, $\mathbf{i} \in \Lambda^n$, so there has to exist $\mathbf{i}, \mathbf{j} \in \Lambda^n$, $\mathbf{i} \neq \mathbf{j}$, with

$$|f_{\mathbf{i}}(x) - f_{\mathbf{j}}(x)| \leq \frac{\operatorname{diam} X}{|\Lambda|^n - 1}.$$

Hence we always have that $\Delta_n \to 0$ (at least) exponentially.

(iii) If Φ satisfies the SSC or OSC, then $\Delta_n \to 0$ **at most** at exponential speed, that is, there exists $c > 0$ and $0 < \varrho < 1$ with

$$\Delta_n \geq c\varrho^n$$

for all $n \in \mathbb{N}$. (exercise)

Hochman found that a strict inequality in (4.16) yields faster than exponential decay for Δ_n:

Theorem 4.77 (Hochman [Ho1])**.** *If* $\dim X < \min\{1, \dim_s \Phi\}$*, then* $\Delta_n \to 0$ *super-exponentially, that is,*

$$-\frac{1}{n} \log \Delta_n \to \infty, \quad n \to \infty.$$

Theorem 4.77 has a wide-range of applications, which we will discuss later on, but let us mention concretely one on rational IFSs.

Definition 4.78. The IFS Φ is **rational** if the defining parameters $r \in \mathbb{Q}$ and $a_i \in \mathbb{Q}$, $i \in \Lambda$.

For rational IFSs we can use Theorem 4.77 to establish Conjecture 4.75:

Corollary 4.79. *If* $\dim X < \min\{1, \dim_s \Phi\}$ *and* Φ *is rational, then* Φ *has exact overlaps.*

Proof. Recall that

$$f_{\mathbf{i}}(x) = r^n x + \sum_{k=1}^{n} a_{i_k} r^{k-1} = r^n x + f_{\mathbf{i}}(0).$$

Since Φ is rational, we can find $p, q, p_i, q_i \in \mathbb{N}$ such that $r = p/q$ and $a_i = p_i/q_i$. Denote $Q = \prod_{i \in \Lambda} q_i$. Then we can write

$$f_{\mathbf{i}}(0) = \sum_{k=1}^{n} a_{i_k} r^{k-1} = \frac{P_{\mathbf{i}}}{Qq^n}$$

for some $P_{\mathbf{i}} \in \mathbb{N}$. Using this we see that $\Delta_n = \min\{|f_{\mathbf{i}}(0) - f_{\mathbf{j}}(0)| : \mathbf{i}, \mathbf{j} \in \Lambda^n, \mathbf{i} \neq \mathbf{j}\}$ is a rational number and with the denominator Qq^n. Suppose Φ has no exact overlaps, that is, $\Delta_n > 0$ for all $n \in \mathbb{N}$. Thus $\Delta_n \geq \frac{1}{Qq^n}$ for all $n \in \mathbb{N}$ and so $\Delta_n \to 0$ at most exponentially. Hence by Theorem 4.77 we must have $\dim X = \min\{1, \dim_s X\}$. □

We leave applications and further topics to the end of the lectures, let us now see how to prove Theorem 4.77.

5 Additive combinatorics and self-similar sets

5.1 Sumset approximations of self-similar sets

Let us now return to self-similar sets and see how additive combinatorics comes into play here. Recall that in Section 4.2 we agreed to fix a self-similar IFS $\Phi = \{f_i\}_{i\in\Lambda}$ consisting of similitudes $f_i(x) = rx + a_i$, $x \in \mathbb{R}$, with the same contraction ratio $0 < r < 1$. The attractor of Φ was denoted by X. We now further assume that $0 \in X$ and $X \subset [0, 1)$, which will not change any of the results as we can translate and rescale X (and thus all the maps) to obtain this.

Definition 5.80. For $n \in \mathbb{N}$ the n:th generation **approximation** of X is the set

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

Since $0 \in X$, we have $f_{\mathbf{i}}(0) \in X$ for all $\mathbf{i} \in \Lambda^n$. Hence $X_n \subset X$. Moreover, $|X_n| \leq |\Lambda|^n$ and we have that $|X_n| < |\Lambda|^n$ if and only if Φ has exact overlaps.

Recall that if $\mathbf{i} \in \Lambda^n$ and $x \in \mathbb{R}$, we have by the iteration formula

$$f_{\mathbf{i}}(x) = r^n x + \sum_{k=1}^{n} a_{i_k} r^{k-1} = r^n x + f_{\mathbf{i}}(0) = r^n x + a_{\mathbf{i}}. \tag{5.17}$$

Using this expression, we can show that the approximations X_n enjoy the following sumset properties:

Lemma 5.81. *For all $m, n \in \mathbb{N}$ we have*

$$X = X_m + r^m X \tag{5.18}$$

and

$$X_{m+n} = X_m + r^m X_n. \tag{5.19}$$

Proof. By (4.15) and (5.17) we can write

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

which is (5.18). For the second equality (5.19) let us first fix $\mathbf{i} \in \Lambda^m$ and $\mathbf{j} \in \Lambda^n$. Denote by $\mathbf{ij} \in \Lambda^{m+n}$ the **concatenation** of $\mathbf{i}$ and $\mathbf{j}$, that is, if $\mathbf{i} = i_1 i_2 \dots i_m$ and $\mathbf{j} = j_1 j_2 \dots j_n$, then $\mathbf{ij}$ is the word $\mathbf{ij} := i_1 i_2 \dots i_m j_1 j_2 \dots j_n$. Using the first equality in (5.17), we can write

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

Then the $m + n$ generation approximation has a form

$$X_{m+n} = \{f_{\mathbf{i}}(0) + r^m f_{\mathbf{j}}(0) : \mathbf{i} \in \Lambda^m, \mathbf{j} \in \Lambda^n\} = X_m + r^m X_n.$$

□

5.2 Box dimension of self-similar sets via sumsets

We mentioned earlier that for self-similar sets X the box dimension $\dim_{\mathrm{B}} X$ always exists:

Theorem 5.82. *If $X \subset \mathbb{R}$ is self-similar set, then $\dim_{\mathrm{B}} X$ exists.*

We will now use the above sum-set structure of self-similar sets to show this. We will later see that this will also help us to approach Hochman's theorem.

Recall the definition of the covering numbers of a set $A \subset \mathbb{R}$ in the scale $\varepsilon > 0$:

$$N(A, \varepsilon) = \min\{k \in \mathbb{N} : \text{we can cover } A \text{ by } k \text{ intervals } I_i \text{ of diameter } \varepsilon\}.$$

We need the following quantitative lemma on covering numbers of sumsets:

Lemma 5.83. *Let $\varepsilon > 0$ and $A, B \subset \mathbb{R}$ with $B \subset [0, \varepsilon)$. Then*

(1) *we have*
$$N(A + B, \varepsilon) \leq 2N(A, \varepsilon)$$

(2) *for any $0 < \delta \leq \varepsilon$ we have*
$$N(A + B, \delta) \geq \frac{1}{3} N(A, \varepsilon) N(B, \delta)$$

Proof. (1) Let $n := N(A, \varepsilon)$ and choose an optimal cover of A by intervals I_i, $i = 1, \ldots, n$, of diameter ε. Assume these intervals are half-open and wite $I_i = [a_i, a_i + \varepsilon)$ for some $a_i \in \mathbb{R}$. Denote

$$I_i' := I_i + [0, \varepsilon) = [a_i + \varepsilon, a_i + 2\varepsilon).$$

Then as $B \subset [0, \varepsilon)$ we can cover

$$A + B \subset \bigcup_{i=1}^{n} (I_i + B) \subset \bigcup_{i=1}^{n} (I_i + [0, \varepsilon)) = \bigcup_{i=1}^{n} (I_i \cup I_i').$$

Thus $A + B$ is covered by $2n$ intervals I_i, I_i' of length ε, which yields

$$N(A + B, \varepsilon) \leq 2n = 2N(A, \varepsilon).$$

(2) Let $n := N(A, \varepsilon)$ and choose an optimal cover of A by **disjoint**[3] intervals I_i, $i = 1, \ldots, n$, of diameter ε. Fix $0 < \delta \leq \varepsilon$, let

$$m := N(A + B, \delta).$$

Now choose an optimal cover of $A + B$ by intervals J_j, $j = 1, \ldots, m$, of diameter δ. Fix any $i = 1, \ldots, n$ and a point $a_i \in A \cap I_i$ (which is possible as $\{I_i\}$ cover A) and define the set of pairs of intervals

$$\mathcal{P} = \{(I_i, J_j) : (a_i + B) \cap J_j \neq \varnothing, j = 1, \ldots, m, i = 1, \ldots, n\}.$$

The collection $\mathcal{P}$ satisfies the following properties:

(i) *For each interval I_i* ***at least*** *$N(B, \delta)$ of the intervals $J_1, \ldots, J_m$ satisfy $(I_i, J_j) \in \mathcal{P}$.*

[3] Now we note that in the definition of box dimension we can choose the optimal interval cover of A (i.e. the one realising $N(A, \varepsilon)$) to be disjoint for example by using dyadic interval partitions $\mathcal{D}_j$, see for example Falconer's book [Fa, Equivalent definitions 3.1].

Indeed, if we fix i, then as $a_i \in A$ we have

$$a_i + B \subset A + B$$

so the intervals J_j cover $a_i + B$. Therefore, $a_i + B$ intersects at least $N(B, \delta)$ intervals J_j for $j = 1, \dots, m$ since $N(B, \delta)$ is the minimal number of δ intervals needed to cover B.

(ii) *For each interval* J_j ***at most*** 3 *intervals* $I_1, \dots, I_n$ *satisfy* $(I_i, J_j) \in \mathcal{P}$.

Indeed, if we fix j such that $J_j = [x_j, x_j + \delta)$ intersects $a + B$ for some $a \in A$, then

$$a \in [x_j - \varepsilon, x_j + \varepsilon)$$

since we assumed $B \subset [0, \varepsilon)$ and $\delta \leq \varepsilon$. Now recall that the diameter of each I_i is $\operatorname{diam}(I_i) = \varepsilon$. Since we chose the A cover $\{I_i\}$ to be disjoint, this means that a belongs to at most 3 intervals in I_i.

Now (i) and (ii) together show that the number of intervals J_j occuring for some $(I_k, J_j) \in \mathcal{P}$ is at least $\frac{1}{3}N(B, \delta)$ times the number of intervals I_i occuring for some $(I_i, J_l) \in \mathcal{P}$. On the other hand, the number of interval J_j occuring as a second coordinate in $\mathcal{P}$ is $m = N(A + B, \delta)$ and the number of intervals I_i occuring as the first coordinate in $\mathcal{P}$ is $n = N(A, \varepsilon)$. Therefore,

$$N(A + B, \delta) \geq \frac{1}{3}N(A, \varepsilon)N(B, \delta).$$

□

Let us now prove the existence of box dimension of self-similar X. Recall that we assumed $0 \in X \subset [0, 1)$.

Proof of Theorem 5.82. Let us first check the following:
Claim 1: *Box dimension* $\dim_{\mathrm{B}} X$ *exists if and only if*

$$\lim_{m \to \infty} \frac{1}{m} \log N(X_m, r^m) \quad \textit{exists.} \tag{5.20}$$

The approximation $X_m \subset X$ so

$$N(X_m, r^m) \leq N(X, r^m).$$

Moreover, by Lemma 5.81 (equation (5.18)) we have

$$X = X_m + r^m X$$

and the diameter $\operatorname{diam}(r^m X) \leq r^m$ (as $X \subset [0, 1)$) so by Lemma 5.83(1) with $\varepsilon = r^m$, $B = r^m X$ and $A = X_m$ we have

$$N(X_m, r^m) \geq \frac{1}{2}N(X, r^m).$$

This yields the claim.[4]

Now let us check the existence (5.20) for the approximations X_m:
Claim 2: (5.20) *is true*

[4]The existence of $\lim_{\varepsilon \to 0} \frac{\log N(X, \varepsilon)}{\log(1/\varepsilon)}$ is equivalent to the existence of $\lim_{m \to \infty} \frac{\log N(X, r^m)}{\log(1/r^m)}$ which is an easy exercise. Note that it is irrelevant that here we happen to have r.

If $n, m \in \mathbb{N}$, then as $X_n \subset X \subset [0,1)$ we have $r_m X_n \subset [0, r^m)$. Therefore by Lemma 5.81 (equation (5.19)) and Lemma 5.83(2) with $\varepsilon = r^m$, $\delta = r^{m+n}$, $A = X_m$ and $B = r^m X_n$ we have (as we have the invariance $N(tA, t\varepsilon) = N(A, \varepsilon)$ for all $t > 0$)

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

Write

$$s_m := \log N(X_m, r^m), \quad m \in \mathbb{N}.$$

Now we just need to check that the limit $\lim \frac{s_m}{m}$ exists. The inequality above shows that for any $m, n \in \mathbb{N}$ we have:

$$s_{m+n} \geq s_m + s_n - \log 3.$$

Let us modify s_m slightly to make a super-additive sequence. Write

$$\widetilde{s}_m := s_m - \log 3, \quad m \in \mathbb{N}.$$

Now

$$\widetilde{s}_{m+n} \geq s_m + s_n - \log 3 - \log 3 = \widetilde{s}_m + \widetilde{s}_n.$$

This means that $(\widetilde{s}_1, \widetilde{s}_2, \widetilde{s}_3, \dots)$ is super-additive so the limit $\lim \frac{\widetilde{s}_m}{m}$ exists. On the other hand

$$\frac{s_m - \widetilde{s}_m}{m} = \frac{\log 3}{m} \to 0, \quad m \to \infty,$$

so the limit $\lim \frac{s_m}{m}$ exists. □

5.3 Set theoretical "proof" of Hochman's theorem for self-similar sets

Let us recall our goal on self-similar sets, Hochman's Theorem 4.77:

Theorem 5.84. *If* $\dim X < \min\{1, \dim_{\mathrm{s}} \Phi\}$*, then*

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

super-exponentially, that is,

$$-\frac{1}{n} \log \Delta_n \to \infty, \quad n \to \infty.$$

We are not yet ready to prove this, but we will give a set theoretical "proof" which shows the key ideas and how (power growth) inverse theorems in additive combinatorics arise into the analysis. Don't worry, we will give a precise proof later using measure theoretical machinery, but as the set theoretical "proof" already contains most of the key ideas, this presentation may help the reader later to understand the technicalities in the upcoming measure theoretical precise proof.

The strategy will be based on to checking how the **failure** of Theorem 5.84 (i.e. only exponential decay of Δ_n) leads to size information on $X + Y$ for some relevant sets Y coming from approximations X_m, which yields to absurd combinatorial features for X that (roughly) should make no sense due to Bourgain's Inverse Theorem 1.19.

Write

$$\alpha := \dim X \quad \text{and} \quad \beta := \min\{1, \dim_{\mathrm{s}} \Phi\}.$$

Suppose $\alpha < \beta$ but $\Delta_n \to 0$ **at most exponentially**, that is, there exists $k \in \mathbb{N}$ with

$$\Delta_n \geq r^{kn}, \quad n \in \mathbb{N}. \tag{5.21}$$

(recall here that r here is the contraction of the IFS Φ, but it does not matter which r we use here!)

Note that this trivially yields that Φ has no exact overlaps. Let us use (5.21) to construct two intervals I_m and J_m of diameter r_m with certain cardinality properties for the approximations X_m.

Lemma 5.85 (Interval I_m)**.** *Write $\sigma := \frac{1}{2}(\beta - \alpha) > 0$. Then for every large enough $m \in \mathbb{N}$ there exists an interval I_m with diameter* $\operatorname{diam}(I_m) = r_m$ *such that*

$$|X_m \cap I_m| > r^{-\sigma m}.$$

Proof. By the proof of Theorem 5.82 we have

$$\alpha = \dim X = \lim_{m\to\infty} \frac{\log N(X_m, r^m)}{\log(1/r^m)}.$$

Thus for large enough $m \in \mathbb{N}$ we have

$$N(X_m, r^m) < r^{-(\alpha+\sigma)m} = r^{-(\beta-\sigma)m}.$$

The condition (5.21) in particular yields that Φ has no exact overlaps so

$$|X_m| = |\Lambda|^m = r^{-m \dim_{\mathrm{s}} \Phi} \geq r^{-m\beta}.$$

Thus if $\{J_i\}_{i=1}^{N(X_m, r^m)}$ is an optimal cover of X_m by intervals with $\operatorname{diam}(J_i) = r^m$, then at least one of the interval J_i must contain

$$\frac{|X_m|}{N(X_m, r^m)} > r^{-\sigma m}$$

points from X_m. Let I_m to be this interval J_i and thus $|X_m \cap I_m| > r^{-\sigma m}$. □

Lemma 5.86 (Interval J_m)**.** *For any $\delta > 0$ and large enough $m \in \mathbb{N}$ there exists an interval J_m with diameter* $\operatorname{diam}(J_m) = r^m$ *with $X_m \cap J_m \neq \varnothing$ and for $n := km$ we have*

$$N((X_m \cap J_m) + r^m X_n, r^{m+n}) < r^{-(1+\delta/2)\alpha n}.$$

Proof. Fix m and define the r-adic interval partition

$$\mathcal{I}_m^r := \{[kr^m, (k+1)r^m) : k \in \mathbb{Z}\}.$$

(recall that in Section 1.4 on multiscale analysis we used $r = 1/2$ and $\mathcal{I}_m^{1/2} = \mathcal{I}_m$.) Now $\mathcal{I}_m^r$ partitions $\mathbb{R}$ so as $X_{m+n} = X_m + r^m X_n$ by Lemma 5.81 (see (5.19)) so

$$X_{m+n} = \bigcup_{J \in \mathcal{I}_m^r} \Big((X_m \cap J) + r^m X_n \Big)$$

Setting $J = [kr^m, (k+1)r^m)$ for some $k \in \mathbb{Z}$ we have $X_m \cap J \subset [kr^m, (k+1)r^m)$ and $r^m X_n \subset [0, r^m)$ as $X_n \subset X \subset [0, 1)$. Therefore,

$$(X_m \cap J) + r^m X_n \subset [kr^m, (k+2)r^m).$$

The intervals $[kr^m, (k+2)r^m)$, $k \in \mathbb{Z}$, cover $\mathbb{R}$ and every $x \in \mathbb{R}$ is contained in at most two of such intervals. Therefore, an argument like in the proof of Lemma 5.83(2) gives us

$$N(X_{m+n}, r^{m+n}) \geq \frac{1}{2} \cdot N(X_m, r^m) \cdot \min_{J \in \mathcal{I}_r^m : X_m \cap J \neq \varnothing} N((X_{m+n} \cap J) + r^m X_n, r^{m+n}).$$

Now recall that the dimension

$$\alpha = \dim X = \lim_{m \to \infty} \frac{\log N(X_m, r^m)}{\log(1/r^m)}$$

by the proof of Theorem 5.82, so using the above inequality

$$\begin{aligned} \min_{J \in \mathcal{I}_r^m : X_m \cap J \neq \varnothing} N((X_{m+n} \cap J) + r^m X_n, r^{m+n}) &\leq 2 \cdot \frac{N(X_{m+n}, r^{m+n})}{N(X_m, r^m)} \\ &\leq 2 \cdot \frac{r^{-(\alpha+o(1))(m+n)}}{r^{-(\alpha+o(1))m}} \\ &= r^{-(\alpha+o(1))n} \end{aligned}$$

as $m \to \infty$ (recall that $n = km$). This is what we claimed if we fix $\delta > 0$ and choose m large enough that the above quantity is below $r^{-(1+\delta/2)\alpha n}$ and having minimum bounded by this can allow us to choose on $J_m \in \mathcal{I}_r^m$ with this property. In particular, $\operatorname{diam}(J_m) = r^m$. □

Theorem 5.87. *Suppose that* $\dim X < \min\{1, \dim_{\mathrm{s}} \Phi\}$ *and* $\Delta_n \geq r^{kn}$ *for all large enough* $n \in \mathbb{N}$*, that is, Hochman's Theorem 4.77 fails. Moreover, suppose the following assumption*

(A) *for all* $\delta > 0$ *and all large enough* $m \in \mathbb{N}$ *the intervals* I_m *and* J_m *from Lemmas 5.85 and 5.86 coincide.*

Then for all $\delta > 0$ *and all large enough* $n \in k\mathbb{N}$ *there exists a subset* $Y_n \subset [0,1]$ *satisfying*

$$N(Y_n, r^n) \geq N(X_n, r^n)^\gamma$$

and

$$N(X_n + Y_n, r^n) \leq N(X_n, r^n)^{1+\delta}$$

for some $\gamma > 0$*. Moreover, the set* Y_n *is obtained as a magnification (recall Section 1.4) of* $X_{n/k}$

Proof. Recall that $\sigma = \frac{1}{2}(\beta - \alpha)$ from Lemma 5.85, which is > 0 by $\dim X < \min\{1, \dim_{\mathrm{s}} \Phi\}$. Set

$$\tau := \frac{\sigma}{k}.$$

Let $n \in k\mathbb{N}$ be large enough such that that Lemmas 5.85 and 5.86 hold for $m \in \mathbb{N}$ satisfying $n = km$. Construct the intervals I_m and J_m from Lemmas 5.85 and 5.86. Thus by Assumption (A) above we have $I_m = J_m$. For our set Y_n let us choose

$$Y_n := r^{-m}(X_m \cap I_m) - c,$$

(the set Y_n is a **blow up** or **magnification** of X_m, note that $m = n/k$ depends on n), where we choose the translation $c \in \mathbb{R}$ so that $Y_n \subset [0,1]$.

Since $\Delta_m \geq r^{km} = r^n$, every point in $X_m \cap I_m$ is separated by at least r^n. Hence by Lemma 5.85 we have

$$N(r^m Y_n, r^{n+m}) = N(X_m \cap I_m, r^{n+m}) \geq |X_m \cap I_m| \geq r^{-\sigma m} = r^{-\tau n}$$

Recall that $N(tA, t\varepsilon) = N(A, \varepsilon)$ for all $t > 0$ so the above yields

$$N(Y_n, r^n) \geq r^{-\tau n} \geq N(X_n, r^n)^\gamma.$$

as $N(X_n, r^n) = r^{-n(\alpha+o(1))}$ if we set $\gamma = \tau/(2\alpha) > 0$. Moreover, since $I_m = J_m$ we have $X_n + Y_n = r^{-m}((X_m \cap J_m) + r^m X_n) + c$ for some $c \in \mathbb{R}$. Therefore by Lemma 5.86 we obtain.

$$N(X_n + Y_n, r^n) = N((X_m \cap J_m) + r^m X_n, r^{m+n}) \leq r^{-(1+\delta/2)\alpha n} \leq N(X_n, r^n)^{1+\delta}$$

as $N(X_n, r^n) = r^{-n(\alpha+o(1))}$. □

Theorem 5.87 would prove Hochman's Theorem 4.77 if we can show that the assumption (A) holds and the conclusion on sumsets would be impossible. Recalling Bourgain's inverse theorem it would make sense that the conclusion cannot hold since X is self-similar: one would not expect to have such a distribution of scales inside X (or the approximations X_m). However, as we said the inverse theorem is not yet precise and it is missing key properties on what do we mean by 'looks like' a tree. Moreover, for the assumption (A) it seems difficult to guarantee such an strong assumption, but if one allows some perturbation to the argument, the assumption (A) can be made to hold in a weak sense that is enough for us. We will now overcome these problems using measure theory.

6 Self-similar measures

6.1 Definitions

Let $\mathcal{P}(X)$ be the family of Borel probability measures on a metric space X. Given a self-similar IFS Φ on $\mathbb{R}$, let us define self-similar measures:

Definition 6.88. A measure $\mu \in \mathcal{P}(\mathbb{R})$ is a (uniform) **self-similar measure** if

$$\mu = \frac{1}{|\Lambda|} \sum_{i \in \Lambda} f_i \mu. \tag{6.22}$$

If $X = \operatorname{supp} \mu$, the support of μ, then as $f_i(\operatorname{supp} \mu) = \operatorname{supp} f_i\mu$ and Λ is finite, we have by (6.22)

$$X = \bigcup_{i \in \Lambda} f_i(X).$$

Hence X is the self-similar set associated to X. Moreover, self-similar measures are unique: it can be checked that the map

$$\mu \mapsto \frac{1}{|\Lambda|} \sum_{i \in \Lambda} f_i \mu$$

is a contraction in $\mathcal{P}(X)$ endowed with, for example, the 1st Wasserstein metric

$$d(\mu, \nu) := \sup \left\{ \left| \int f \, d\mu - \int f \, d\nu \right| : f \text{ is 1-Lipschitz } X \to \mathbb{R} \right\}$$

and we can use the Banach fixed point theorem again to make (6.22) unique.

Remark 6.89. The reason we call these "uniform" self-similar measures is because we weight each measure $f_i\mu$ by the equal weight $1/|\Lambda|$. General self-similar measures (see for example [Fa]) are often defined with any general weights $0 \leq p_i \leq 1$ (by putting them inside the sum in (6.22)) in place of $1/|\Lambda|$ which are required to satisfy $\sum_{i \in \Lambda} p_i = 1$. All the theory presented here will also follow for them, but for simplicity we stick to "uniform" self-similar measures.

6.2 Dimension of self-similar measures

Dimensions have analogue for measures as well.

Definition 6.90. The (Hausdorff) **dimension** of a measure $\mu \in \mathcal{P}(X)$ is

$$\dim \mu := \inf\{\dim A : A \subset \mathbb{R}, \mu(A) > 0\}.$$

If μ is the self-similar measure for Φ and X is the attractor, then $\mu(X) = 1$. Thus we can conclude:

Lemma 6.91. *If μ is the self-similar measure for Φ and X is the attractor, then*

$$\dim \mu \leq \dim X \leq \min\{1, \dim_s \Phi\}.$$

6.3 Hochman's theorem for self-similar measures

Let us now give a measure theoretical version of Hochman's Theorem 4.77:

Theorem 6.92 (Hochman [Ho1])**.** *If μ is the self-similar measure associated to Φ and we have*

$$\dim \mu < \min\{1, \dim_s \Phi\},$$

then $\Delta_n \to 0$ super-exponentially, that is,

$$-\frac{1}{n} \log \Delta_n \to \infty, \quad n \to \infty.$$

This is what we will prove in the following sections. However, let us first remind why this implies Theorem 4.77

6.4 Self-similar measures ⇒ self-similar sets

Using the measure theoretical Theorem 6.92 we can prove the set-theoretical Theorem 4.77:

Proof of Theorem 4.77 assuming Theorem 6.92. By the assumption in Theorem 4.77 we have $\dim X < \min\{1, \dim_s \varphi\}$. Hence by Lemma 6.91 we have that $\dim \mu < \min\{1, \dim_s \Phi\}$. Thus Theorem 6.92 applies so we have our claim. □

7 Proof of Hochman's theorem for self-similar measures

7.1 Convolution structure of self-similar measures

Let us now now turn to prove Hochman's theorem (Theorem 6.92) for self-similar measures, recall the statement:

Theorem 7.93. *If μ is the self-similar measure associated to Φ and we have*

$$\dim \mu < \min\{1, \dim_s \Phi\},$$

then $\Delta_n \to 0$ super-exponentially, that is,

$$-\frac{1}{n} \log \Delta_n \to \infty, \quad n \to \infty.$$

Recall that in Section 5.3, we reformulated the problem using sumsets. but do it measure theoretically using convolutions and entropy. Define the following uniform probability measure $\mu^{(n)}$ on the approximations $X_n = \{f_{\mathbf{i}}(0) : \mathbf{i} \in \Lambda^n\}$ as follows:

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

Note that if Φ has exact overlaps, we weight $\mu^{(n)}$ using the multiplicities of a given word $\mathbf{i} \in \Lambda^n$ instead of the uniform weight $1/|\Lambda|^n$. Then we have that $\mu^{(n)} \to \mu$ weakly.

Define a re-scaling map

$$S_t(x) := tx$$

and let $S_t\nu$ be the push-forward of a meeasure $\nu \in \mathcal{P}([0,1])$ under this map. In particular, $S_t\nu$ is a rescaling of ν under t. An analogue of Lemma 5.81 for sets is

Lemma 7.94.

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

and

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

Moreover, the analogue of Lemma 5.83 for the covering numbers

Lemma 7.95. *Let $\varepsilon > 0$ and $\mu, \nu \subset \mathcal{P}([0,1])$ with $\operatorname{supp}\nu \subset [0,\varepsilon]$. Then*

(1) *we have*

$$H(\mu * \nu, \varepsilon) \leq H(\mu, \varepsilon) + O(1)$$

(2) *for any $0 < \delta \leq \varepsilon$ we have*

$$H(\mu * \nu, \gamma) \geq H(\mu, \varepsilon) + H(\nu, \gamma) - O(1).$$

As for the box dimension of self-similar sets (Theorem 5.82, we can use this to obtain the following:

Theorem 7.96. *For the self-similar measure μ, the entropy dimension $\dim_{\mathrm{e}}\mu$ exists.*

Proof. Recalling the proof of Theorem 5.82 we use the above lemmas to conclude that

$$|H(\mu^{(m)}, r^m) - H(\mu, r^m)| = O(1),$$

which is the entropy analogue to (5.20)), and moreover, we can conclude with the lemmas the following superadditivity bound:

$$H(\mu^{(m+n)}, r^{m+n}) \geq H(\mu^{(m)}, r^m) + H(\mu^{(n)}, r^n) - O(1),$$

which is an entropy analogue to $N(X_{m+n}, r^{m+n}) \geq \frac{1}{3}N(X_m, r^m)N(X_n, r^n)$ proved in Theorem 5.82. These yield Theorem 7.96. □

As for the box dimension and Hausdorff dimension of self-similar sets, the Hausdorff dimension and the entropy dimension of μ agrees, which was proved by Feng and Hu:

Theorem 7.97 (Feng-Hu)**.** *For the self-similar measure μ, we have $\dim\mu = \dim_{\mathrm{e}}\mu$.*

We will use this result throughout the proof and refer to Feng and Hu's paper on obtaining this result.

7.2 Components of self-similar measures

Assume now on the contrary that

$$\dim \mu < \min\{1, \dim_{\mathrm{s}} \Phi\},$$

but $\Delta_n \to 0$ only exponentially, that is,

$$\Delta_n \geq r^{kn}$$

for some $k \in \mathbb{N}$ and for all $n \in \mathbb{N}$. Then in particular Φ does not have exact overlaps, so for the approximations $X_n = \{f_{\mathbf{i}}(0) : \mathbf{i} \in \Lambda^n\}$ we have

$$|X_n| = |\Lambda|^n.$$

Now as we argued for sets by constructing intervals I_m and J_m in Section 5.3, we can prove the following lemma, which is an analogue of Lemma 5.85:

Lemma 7.98. *There exists $\tau > 0$ such that if $n = km$ we have for all $m \in \mathbb{N}$ that*

$$\mathbb{P}\Big(\frac{1}{n\log(1/r)} H((\mu^{(m)})_{x,m}, r^{m+n}) > \tau\Big) > \tau.$$

and the following lemma, which is an analogue of Lemma 5.86:

Lemma 7.99. *For all $\delta > 0$ we have*

$$\lim_{m\to\infty} \mathbb{P}\Big(\frac{1}{n\log(1/r)} H((\mu^{(m)})_{x,m} * S_{r^m}\mu, r^{m+n}) < (1+\delta)\dim\mu\Big) = 1$$

where $n = km$.

Now if $\delta > 0$ is fixed, and m is large enough, the events in Lemmas 7.98 and 7.99 intersect. Hence using an argument as for sets, we have the following

Lemma 7.100. *Assume $\dim\mu < \min\{1, \dim_{\mathrm{s}}\Phi\}$ and $\Delta_n \geq r^{kn}$ for some $k \in \mathbb{N}$ and all $n \in \mathbb{N}$. Then there exists $\tau > 0$ such that for any $\delta > 0$ and for all large enough $n \in \mathbb{N}$ we can find a measure $\nu \in \mathcal{P}([0,1])$ such that*

$$H_{kn}(\nu) > \tau$$

and

$$H_{kn}(\mu * \nu) < H_{kn}(\mu) + \delta.$$

However, according to Hochman's inverse theorem for entropy this will be absurd as for self-similar measures (as we will soon), the component measures are in the sense of entropy uniform on "all" scales, recall in particular Lemma 2.45, which controls the number of singular scales by $H_{kn}(\nu)$.

7.3 Uniformity of the components

We have the following lemma that states that the components of the self-similar μ are close to being uniform in the sense of entropy on most scales:

Lemma 7.101. *Let μ be the self-similar measure for Φ. For any $\varepsilon > 0$ and m large enough and for any $n \in \mathbb{N}$ we have*

$$\mathbb{P}\Big(\frac{1}{m\log(1/r)} H(\mu_{x,n}, r^{n+m}) > \alpha - \varepsilon\Big) > 1 - \varepsilon$$

As a quantitative consequence we have the following lemma on scales:

Lemma 7.102. *Let μ be the self-similar measure for Φ. For any $\varepsilon > 0$ and m large enough and for any $n \in \mathbb{N}$ we have*

$$\mathbb{P}_{1\le i\le n}\Big(\frac{1}{m\log(1/r)}H(\mu_{x,i}, r^{n+m}) < \alpha+\varepsilon\Big) > 1-\varepsilon.$$

In particular

$$U := \Big\{i \in \{1,\dots,n\} : \mathbb{P}\Big(\frac{1}{m\log(1/r)}H(\mu_{x,i}, r^{n+m}) > \alpha+\varepsilon\Big) > \sqrt{\varepsilon}.\Big\}$$

has the cardinality bound $|U| < n\sqrt{\varepsilon}$.

7.4 Applying Hochman's inverse theorem

We will use Lemma 7.102 and Hochman's inverse theorem 2.41 to obtain the following statement on what happens to the convolution of $\mu * \nu$ for the self-similar μ when ν has positive entropy:

Lemma 7.103. *Let μ be the self-similar measure for Φ with $\dim\mu < 1$. Then for any small enough $\tau > 0$ and for all large enough $m \in \mathbb{N}$ we can find $\delta > 0$ such that for all large enough $n \in \mathbb{N}$ and for every $\nu \in \mathcal{P}([0,1])$ with*

$$H_n(\nu) > \tau$$

we have

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

Proof. Fix $\tau > 0$ small enough that $\dim\mu + \tau < 1 - \tau$. Suppose m is large enough such that Lemma 7.102 is true with $\varepsilon = \tau$. Let $\delta = \delta(\tau, m) > 0$ be the number for which the inverse Theorem 2.41. Fix Now choose n large enough such the inverse theorem and Lemma 7.102 hold for n and assume on the contrary that we have

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

Then by the inverse theorem we have disjoint sets $I, J \subset \{1,\dots,n\}$ such that $|I \cup J| > (1-\tau)n$ such that the conclusion of the inverse theorem holds. Now by the definition of the set U in Lemma 7.102 we have $I \subset U$, which yields $|I| \le |U| < n\sqrt{\tau}$. Hence $|J|/n \ge 1 - |I|/n - \tau \ge 1 - 2\sqrt{\tau}$. Thus by Lemma 2.44 we have

$$H_n(\nu) = O(\tau) + \frac{|\{1,\dots,n\}\setminus J|}{n} + O\Big(\frac{1}{n}\Big) \le O(\sqrt{(}\tau)) + O(1/n).$$

This yields the claim by letting n be large enough and τ small enough. □

7.5 Synthesis

Now we can prove Hochman's theorem (Theorem 6.92) on self-similar measures:

Proof of Theorem 6.92. Suppose $\dim\mu < \min\{1, \dim\Phi\}$ and $\Delta_n \ge r^{kn}$ for some $k \in \mathbb{N}$ and for all $n \in \mathbb{N}$. Now by Lemma 7.100 there exists $\tau > 0$ such that for any $\delta > 0$ and for all large enough $n \in \mathbb{N}$ we can find a measure $\nu \in \mathcal{P}([0,1])$ such that

$$H_{kn}(\nu) > \tau$$

and

$$H_{kn}(\mu * \nu) < H_{kn}(\mu) + \delta.$$

Hence by Lemma 7.103 for this τ there exists $\delta_0 > 0$ such that for all large enough $n \in \mathbb{N}$ we have

$$H_{kn}(\mu * \nu) > H_{kn}(\mu) + \delta_0$$

since $H_{kn}(\nu) > \tau$. This is a contradiction so the claim is correct. □

8 Further applications

References

[Bo1] Jean Bourgain: On the Erdős-Volkmann and Katz-Tao ring conjectures. *GAFA*, 13(2):334–365, 2003.

[Bo2] Jean Bourgain: The discretized sum-product and projection theorems. *J. Anal. Math.*, 112:193–236, 2010.

[BG] Jean Bourgain, Alex Gamburd: A spectral gap theorem in SU(d). *J. Eur. Math. Soc.*, 14(5), 1455–1511, 2012.

[CT] Thomas Cover, Joy Thomas: *Elements of information theory*. John Wiley (2nd ed), 2006.

[Fa] Kenneth Falconer: *Fractal geometry - Mathematical Foundations and Applications*, John Wiley (3rd ed), 2014.

[Ho1] Michael Hochman: On self-similar sets with overlaps and inverse theorems for entropy, *Ann. of Math.*, 180(2): 773-822, 2014.

[Ho2] Michael Hochman: Self-similar sets, entropy and additive combinatorics, *Geometry and Analysis of Fractals*, Springer Proceedings in Mathematics & Statistics, 88: 225–252, 2014.

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

[Ma] Pertti Mattila: *Geometry of sets and measures in Euclidean spaces*, Cambridge University Press, 1995.

[Ta] Terence Tao: Sumset and inverse sumset theory for Shannon entropy. *Combin. Probab. Comput.*, 19(4):603–639, 2010.

[TV] Terence Tao, Van Vu: *Additive combinatorics*, Cambridge University Press, 2006.