Fractal Geometry

Han Yu

WARWICK
Current address: Warwick
Email address: han.yu.2@warwick.ac.uk

2020 *Mathematics Subject Classification.* Primary TBD;
Secondary TBD

Key words and phrases. `amsbook`, AMS-LaTeX

TBD.

Contents

Preface

This draft manuscript is for the course MA3D4 Fractal Geometry at the University of Warwick. In this 30hrs course, we will learn some basic as well as some not basic results in the study of fractal geometry and related fields. The materials are divided naturally into 3 themes.

- Foundations and basic constructions: Hausdorff dimension, Hausdorff measure; Box dimension; self-similar sets; some graphs of functions;
- The geometry of fractal sets and measures: geometric theorems (Marstrand); self-similar measures; random fractals;
- **Non-examinable** Number theory, Combinatorics and fractal measures: algebraic linear self-similar systems; Shmerkin-Wu theorem; Bourgain's sum product theorem;

N.E. marks non-examinable parts. Such parts will not be tested in the exam. However, some N.E. parts are useful in gaining a solid understanding of other materials including the examinable ones. All of the N.E. parts can be treated as blackboxes, i.e. you are not required to understand their proofs.

Warning: This draft is not error free, at least for a long while!

Highly recommended books: [**1, 4, 7**].

Han Yu

Part 1

Basic Topics

Some common notations

Let f, g be real-valued functions on a variable t. We assume that g is never zero. We use $f = O_{t \to t_*}(g)$ or $f \ll_{t \to t_*} g$ if for some $C \geq 0$,

$$\lim_{t \to t_*} \frac{|f(t)|}{|g(t)|} < C.$$

We omit $t \to t_*$ if this is clear in the context. Often t_* is either 0 or ∞. In this case, we also write

$$g = \Theta(f), g \gg f.$$

We write $f \asymp g$ if $f \ll g$ and $f \gg g$. If $C = 0$, then we write

$$f = o(g).$$

CHAPTER 1

Some basic constructions of fractal sets

We study fractals in Fractal Geometry. However, it is not clear what a fractal is. In fact, there is no standard definition telling us what a fractal should be. Wikipedia nowadays suggests that *"In mathematics, a fractal is a geometric shape containing detailed structure at arbitrarily small scales, usually having a fractal dimension strictly exceeding the topological dimension."* This is more of an idea than a precise definition. In this chapter, we will see many examples of fractals that fit the general idea stated in Wikipedia.

1.1. Random constructions

1.1.1. Brownian Motion. There are a lot of naturally formed fractals in the real world. This is not at all strange! In fact, smooth and regular-looking objects should be considered as being outliers.

Here is a list of fractals in nature. I did not include them in the notes for the printability.

- Landscapes https://www.fractal-landscapes.co.uk/photos.html
- Snowflakes https://www.noaa.gov/stories/how-do-snowflakes-form-science-behind-snow
- Fungi https://southern-highlands.naturemapr.org/sightings/4429351
- River https://en.wikipedia.org/wiki/River_delta#/media/File:Lena_River_Delta_-_Landsat_2000.jpg
- Lightning https://commons.wikimedia.org/wiki/File:Pink_Lightning.jpg
- Browning Motion https://www.youtube.com/watch?v=ernnQJwaKTs&ab_channel=StephenCurry

One way to explain such fractal structures is that many macroscopic phenomena are actually caused by microscopic parts and interact with each other in a certain way. A perfect example of such a phenomenon is Brownian Motion. In order to have some intuitive ideas, let us consider that in a $1 \times 1 \times 1$ box, we sequentially throw in white balls of radius 10^{-5} one after the other. To simplify the imagination, let us consider that the balls are all starting to move from the centre of the box with a unit speed of a uniformly random direction. After we have put in enough (like 10^3 many) balls, we can start tracking a particular ball. To do this, we can choose a moving ball and dye it red. We can then get a red path. (Recommended: try to write codes to simulate this procedure.)

Sometimes, our problem is not to understand the accurate mechanism for generating a natural fractal. We can then free ourselves from considering the physics law and focus on modelling this fractal. For example, we just mentioned Brownian motion by introducing a 'realistic' environment. This is not necessary. In fact, we could have chosen the following path.

- Consider the lattice $\mathbb{Z}^3$. Let us consider the point $P(0) = (0, 0, 0)$.
- Each time (modelled as an integer k), we can throw a dice and decide to move the point from $P(k)$ to $P(k+1)$ according to the outcome of the dice. (There are six possible directions to move and there are six faces of a dice.) We can then draw a red line from $P(k)$ to $P(k+1)$.
- After a long time, say, N steps, we can take a look at the picture we drew by rescaling the whole picture with a factor of $1/\sqrt{N}$.
- For large enough N, the picture will look like a path of Brownian Motion.

Paths formed by the bouncing hard-ball system look like the random walk picture. However, it is extremely difficult to rigorously establish the fact that those two pictures are alike in a certain mathematical sense. See [2]. Even without mathematical justification, it is still impressive to obtain randomly generated pictures that look just like those that come from natural and deterministic processes. Inspired by the artificial construction of Brownian Motion, we now look at some more ways of generating nice (and natural) fractals. Such techniques are quite useful in Computer Graphics for procedural content generation.

1.1.2. Landscape. We consider the function $T(x) = \sin(\pi x)$. Let $0 < a < 1 < b$ be numbers. Consider the function

$$W_{a,b}(x) = \sum_{j \geq 0} a^j T(b^j x).$$

This function is well-defined because the sum converges absolutely and uniformly on $\mathbb{R}$. Thus $W_{a,b}$ is a continuous function. If the product $ab < 1$, then $W_{a,b}$ is differentiable. If $ab \geq 1$, then $W_{a,b}$ is not differentiable. If $ab > 1$ then $W_{a,b}$ is nowhere differentiable and the graph looks like a fractal set. However, even for $ab < 1$, we can generate interesting pictures. Try the following Mathematica code that can run on Mathematica 13.3 (and perhaps on other versions as well).

```
W[a_, b_, x_] := Sum[a^k Sin[Pi  b^k  x], {k, 40}];(*Weierstrass
    function definition*)

a = 0.7; b = Pi;

Manipulate[Plot[W[a, b, x], {x, window, 0.005 + window}], {window, 0,
    1}]
```

LISTING 1.1. Mathematica 13.3 code

It is fun to set a pair of values for a, b and look at the generated graph. Here is a list of examples:

- Mountain line: $a = 0.4, b = \pi$.
- Stock price: $a = 0.65, b = \pi$.

We can also make the sum in $W_{a,b}$ to be random. Namely, we consider

$$R_{a,b}(x) = \sum_{j \geq 0} \pm a^j T(b^j x + \theta_j)$$

where each $\pm$ is chosen randomly for each j independently with equal weight and θ_j is a random variable taking values in $[0, 1]$ with uniform distribution. We can insert as much randomness as we wish. The only obstruction is our imagination. Try the following code.

```
a = 0.45; b = Pi;
c = 0.45; d = Pi;
RandomList = RandomInteger[1, {10, 10}];
Theta = RandomReal[1, {10, 10}];

T[x_, y_] := Sum[2*(RandomList[[k]][[j]] - 0.5) a^k Sin[Pi  (b^k  x +
    Theta[[k]][[j]])]*2*(RandomList[[k]][[j]] - 0.5) c^jSin[Pi  (d^j
    y + Theta[[k]][[j]])], {k, 10}, {j, 10}];

Plot3D[T[x, y], {x, 0, 0.5}, {y, 0, 0.5}, Mesh -> 10, MeshFunctions ->
     {#1 &, #2 &}, MeshShading -> {{Automatic, Automatic},{Automatic,
    Automatic}}, ColorFunction -> "DarkRainbow"]
```

Listing 1.2. Mathematica 13.3 code

1.1.3. Voronoi diagram. Let $[0,1]^k$ be the unit cube. Let $F \subset (0,1)^k$ be a finite set. Then for each $x \in [0,1]^k$, the minimal Euclidean distance $d(x,F)$ from x to the set F is well-defined. In general, it can happen that there are multiple points in F that achieve this minimal distance. However, for most of $x \in [0,1]^k$, the minimal distance $d(x,F)$ is achieved via a unique $f \in F$. We can then colour this point x by f. As a result, we decomposed $[0,1]^k$ into a disjoint union of polygons. This is the Voronoi decomposition of $[0,1]^k$ via F.

We can choose F randomly and keep refining the decomposition by introducing new points. We omit further details. See http://files.righto.com/fractals/vor.html

1.2. Dynamical Systems

Generally speaking, a dynamical system is a pair (X,T) where X is a set and $T : X \to X$ is a map. We want to understand the iterated actions of T. For example, we can take $X = \mathbb{C}$ and $T : z \to f(z)$ by a polynomial function f. We want to consider the following set

$$J_f = \{z : \sup_{n\geq 1} |f^n(z)| < \infty\}.$$

This is the set of points on the complex plane such that under the iterated action of f, the orbit $\{f^n(z)\}_{n\geq 1}$ is bounded.

- If $f(z) = z$, then $J_f = \mathbb{C}$.
- If $f(z) = z^2$, then $J_f = \mathbb{D}$, the unit disk $\{z : |z| \leq 1\}$.
- If $f(z) = z + 1$, then J_f is empty.
- What is J_f for $f(z) = z^2 + 1$?

For simplicity we write J_c for J_f with $f(z) = z^2 + c$ where $c \in \mathbb{C}$. Here are some pictures of J_c with different c. See https://en.wikipedia.org/wiki/Julia_set

Next, we see that for some of c, $0 \in J_c$. It is curious to see the collection of c with such a phenomenon. We define

$$M = \{c : 0 \in J_c\}.$$

We can now draw M. See https://en.wikipedia.org/wiki/Mandelbrot_set

The study of J_f for general f (not necessarily a polynomial) lies in the heart of Complex Dynamics, a difficult research area with a lot of nice pictures.

1.3. Iterated Function Systems (IFS)

Let $n \geq 1$ be an integer. An IFS on $\mathbb{R}^n$ is a collection of functions

$$\mathcal{F} = \{f_i : \mathbb{R}^n \to \mathbb{R}^n\}_{i \in I}$$

where I is an index set. Let U be a subset of $\mathbb{R}^n$. We say that U is (forward) invariant under $\mathcal{F}$ if

$$f_i(U) \subset U$$

for all $i \in I$. The idea is similar to that of Dynamical Systems. The difference here is that we could have more than one map to generate the 'dynamics'. Clearly, $\mathbb{R}^n$ is always invariant. However, this tells us nothing about the IFS. In fact, a large part of the study of IFSs is the study of possible invariant sets.

If I is a finite set and $\mathcal{F}$ is a collection of linear maps we call it a self-affine system. Any invariant set is called a self-affine set. If moreover $\mathcal{F}$ is a collection of Euclidean isometries, we call it a self-similar system and any invariant set is a self-similar set.

We will study linear IFSs in a later chapter. For now, let us see some examples.

Example 1.3.1:Middle Third Cantor set

Consider the IFS $\mathcal{F} = \{f_1 : x \to x/3, f_2 : x \to (x+2)/3\}$ on $\mathbb{R}$. There is a compact invariant set $K \subset [0,1]$. It can be constructed via the 'chopping the middle third method'. More precisely, consider the unit interval $I_0 = [0,1]$ We then remove the middle third part $(1/3, 2/3)$ and obtain a disjoint union of two intervals of length $1/3$. We can then remove from each of those intervals the middle third part. This procedure continues indefinitely. As a limit, we obtain K.

We can generalise the above construction in higher dimensional Euclidean spaces. The following list contains a few examples. You can run the code on Mathematica 13 and see how the construction goes in each example.

Example 1.3.2:Sierpinski Triangle

```
GraphicsRow[Table[MengerMesh[n], {n, 0, 4}]]
```

Listing 1.3. Mathematica 13.3 code

Example 1.3.3:Sierpinski Carpet

```
GraphicsRow[Table[SierpinskiMesh[n], {n, 0, 3}]]
```

Listing 1.4. Mathematica 13.3 code

Example 1.3.4:von Koch curve

```
GraphicsRow[Table[Graphics[KochCurve[n]], {n, 0, 6}]]
```

LISTING 1.5. Mathematica 13.3 code

Example 1.3.5:Menger Sponge

```
MengerMesh[4,3]
```

LISTING 1.6. Mathematica 13.3 code

1.4. Digit expansions in number fields (N.E.)

Many examples of IFSs can also be viewed from a number-theoretic perspective. For example, we can consider ternary expansion of numbers, i.e. digit expansions with base 3. Then the Middle Third Cantor set C is the closure of

$$\{x \in [0,1] : \text{a ternary expansion of } x \text{ has only } 0, 2 \text{ as digits}\}.$$

Notice that some rational numbers can have two different ternary representations. A lot of examples from the linear IFS can be constructed as missing-digit sets like the middle-third Cantor set. We can continue from this point of view and generalise the construction of missing-digit sets. Here, it is illustrative to consider a non-integral basis, i.e. expansions with respect to non-integers. To do this, we will make a small detour into algebraic number theory.

Let us consider the ring $\mathbb{Z}$. For each integer $b \geq 2$, there is a filtration of ideals

$$\cdots \subset (b^2) \subset (b) \subset (1) = \mathbb{Z}.$$

One can expend this chain as follows

$$\cdots \subset (b^2) \subset (b) \subset (1) = \mathbb{Z} \subset (b^{-1}) \subset (b^{-2}) \subset \ldots$$

Notice that $(b^{-1}) = \mathbb{Z}/b$ is no longer a set of integers. It can be checked that for all $t \in \mathbb{Z}$, there is a canonical isomorphism of rings

$$(b^t)/(b^{t-1}) \cong \mathbb{Z}/b\mathbb{Z}.$$

We can now 'complete' the above chain by considering one-sided infinite sequences with symbols in the finite set $\mathbb{Z}/b\mathbb{Z}$. A typical element is represented as

$$A_b = b_k b_{k-1} \ldots b_0 b_{-1} b_{-2} \ldots$$

for some integer k. We declare A_b as an element of $\mathbb{R}$ by giving a sense of convergence, i.e. a topology. In this way, we obtained b-ary expansions of real numbers.

If we 'complete' the chain in the other direction, then we do not have $\mathbb{R}$. In fact, if b is a prime number, we have the local field $\mathbb{Q}_b$ of b-adic numbers. To put $\mathbb{R}$ into the context, it is often written as $\mathbb{Q}_\infty$.

Let us consider a number field $\mathbb{K}$ with the ring of integers $\mathcal{O}_\mathbb{K}$. As $\mathcal{O}_\mathbb{K}$ may not be a principle ideal domain, it would be too restrictive to only introduce digit expansions with bases in $\mathcal{O}_\mathbb{K}$. A more proper way is to introduce 'digit' expansions with respect to ideals. Of course, if $\mathcal{O}_\mathbb{K}$ is a PID, there is no difference. Let I be a non-trivial ideal. Then $\mathcal{O}_\mathbb{K}/I$ is a finite set. As in the case of $\mathbb{Z}$, we can forge an infinite chain of fractional ideals

$$\mathcal{O}_\mathbb{K} \subset I^{-1} \subset I^{-2} \cdots \subset \mathbb{K}_\infty.$$

In this way, we obtained I-ary expansions of elements in the torus $\mathbb{K}_\infty/\mathcal{O}_\mathbb{K}$. We can finitely extend the I-ary ideal chain towards the left side and obtain I-ary expansions of $\mathbb{K}_\infty$. By using Galois embedding, we can identify $\mathbb{K}_\infty$ with

$$\mathbb{R}^s \times \mathbb{C}^{2r}$$

where $s, 2r$ are the numbers of real and complex embeddings and $s+2r$ is the degree of $\mathbb{K}/\mathbb{Q}$. By using the above ideas, we can now construct more exotic examples of Cantor sets.

Example 1.4.1:real quadratic example

Planned

Example 1.4.2:complex quadratic example

Planned

CHAPTER 2

Measures and dimensions

2.1. The Lebesgue measure: a review

2.1.1. σ-algebra and measure. The notion of σ-algebra is central in measure theory and related fields, e.g. probability theory. We now briefly recall the definition.

Definition 2.1.1:

Let X be a set. Consider the power set $\mathcal{P}(X)$. A σ-algebra $\mathcal{A}$ is a collection of subsets of X, i.e. $\mathcal{A} \subset \mathcal{P}(X)$ such that

- if $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$.
- given a countable collection $A_1, \ldots, A_i, \ldots$ of elements in $\mathcal{A}$, their intersection as well as their union are also in $\mathcal{A}$.
- (*) The empty set $\emptyset$ and X itself are in $\mathcal{A}$.

Note that (*) is redundant. A fact that will be useful is that for any collection $\mathcal{F} \subset \mathcal{P}(X)$, there is a smallest σ-algebra containing $\mathcal{F}$. That is, there is a σ-algebra $\sigma(\mathcal{F})$ such that any other σ-algebra that contains $\mathcal{F}$ also contains $\sigma(\mathcal{F})$.

Once we have a space $(X, \mathcal{A})$ with $\mathcal{A}$ being a σ-algebra of X, we will call $(X, \mathcal{A})$ a measurable space. Now we want to introduce the notion of measure.

Definition 2.1.2:

Let $(X, \mathcal{A})$ be a measurable space. A measure μ is a function

$$\mu : \mathcal{A} \to G$$

where G is a complete topological abelian group. In addition, μ satisfies

- $\mu(\emptyset) = 0$ the identity element of G.
- For any countable collection of pairwise disjoint $A_1, \ldots,$ in $\mathcal{A}$,

$$\mu(\cup_i A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

In most cases, $G = \mathbb{R}$ or $\mathbb{C}$ with respect to the addition. In case $G = \mathbb{R}$, we will also call μ to be positive if $\mu(A) \geq 0$ for all $A \in \mathcal{A}$. We will always assume this is the case.

There is no guarantee that such a measure μ exists. However, in most of the cases, one can construct measures via certain extension theorems (e.g. Carathéodory's extension theorem). The idea is that one can explicitly construct μ on a much smaller collection $\mathcal{A}' \subset \mathcal{A}$. Then it is possible to extend μ to $\sigma(\mathcal{A}')$. If $\sigma(A') = \mathcal{A}$, then we have a satisfactory μ at hand. This condition can also be checked via certain generation theorems (e.g. Dynkin's $\pi - \lambda$ theorem). We will not provide more details on measure theory where one can find 'infinitely' many excellent learning materials (e.g. Wikipedia+references therein).

2.1.2. Borel sets. We will not encounter very general X as our universe. In fact, we will only consider metric spaces. Let X be a metric space. In this case, its metric topology $\mathcal{T}$ generates $\mathcal{B}(X) = \sigma(\mathcal{T})$. This σ-algebra is so important that we give it a name: the Borel σ-algebra of X. Each $A \in \mathcal{B}(X)$ is called a Borel set. In particular, any open/closed set is Borel. Not all Borel sets are open/closed. Then half interval $[0, 1) \subset \mathbb{R}$ demonstrates this fact.

We will not encounter very general metric space either. In fact, we will only consider those metric spaces X that are separable (i.e. the existence of a countable dense subset) and complete (i.e. the existence of all Cauchy sequences). Such a metric space is called a *Polish space*. Nearly all spaces that we will meet in everyday life are Polish, including any separable Banach space (basically, Polish linear space).

2.1.3. Continuity and Regularity of measures. Consider a (positive) measure μ with value in $[0, \infty]$. From the definition, it can be checked that

$$\mu(A) = \lim_{i\to\infty} \mu(A_i)$$

if A_i is an increasing list of measurable sets with union A. It can be also checked that

$$\mu(A) = \lim_{i\to\infty} \mu(A_i)$$

if A_i is a decreasing list of measurable sets with intersection A and A_1 has finite measure. These results are called the *continuity* of measures.

A related notion is the *regularity* of measures on Borel sets. In this context, we have a topology at hand.

Definition 2.1.3:Regular measure

Let $(X, \mathcal{B}, \mu)$ be a Borel measurable space. μ is regular if for any measurable set A,

$$\mu(A) = \sup\{\mu(K) : K \subset A, K \text{ is compact}\}$$

and

$$\mu(A) = \inf\{\mu(O) : O \supset A, K \text{ is open}\}.$$

Unlike the continuity, the regularity does not come for free. However, most of the measures that we will study in this course will have this regularity property. In most cases, it is not too difficult. We will illustrate some simple examples later in this chapter.

2.1.4. The Lebesgue measure. We will first encounter the Lebesgue measure on $\mathbb{R}^k$ where $k \geq 1$ is an integer. Here recall that for any Euclidean space, we have the Borel σ-algebra $\mathcal{B}(\mathbb{R}^k)$ which comes from the standard Euclidean metric.

2.1.4.1. *Step 1:* The basic intuition of Lebesgue measure is the notions of 'length, area, volume'. Let us focus on the case when $k = 1$ while other cases are treated similarly. Consider the collection of open intervals

$$\mathcal{I} = \{(a, b) : -\infty \le a < b \le \infty\}.$$

This collection is not a topology nor a σ-algebra. However, it can be checked that the standard metric topology as well as $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$ are generated by $\mathcal{I}$. Now, we begin by defining

$$\lambda((a, b)) = b - a \in [0, \infty].$$

Here, $b-a$ is the diameter of the interval (a, b). For higher dimensions, we can start with the Cartesian product of intervals and give them the 'natural' volume which is the product of the lengths of their sides. It is possible to check that the measures we are given to the intervals (or higher dimensional boxes) satisfy the following scaling property

$$\lambda(tB) = t^k \lambda(B)$$

where $t > 0$ and B is a box in $\mathbb{R}^k$. Quite coincidentally, the scaling exponent k in t^k is the topological dimension of the space $\mathbb{R}^k$.

2.1.4.2. *Step 2:* After we agree with the length of intervals, we want to extend the notion of length to a larger class of sets. Consider

$$\mathcal{I}' = \text{finite unions of intervals}$$

whereby intervals, we mean all kinds of intervals, not only the open ones. We first declare that $\lambda(\{x\}) = 0$ for each $x \in \mathbb{R}$. We can then extend λ by (finite) additivity.

2.1.4.3. *Step 3:* Suppose that a countable collection of pairwise disjoint $I_1, \dots$ are in $\mathcal{I}'$ and their union I is also in $\mathcal{I}'$. Then we can check that

$$\sum_i \lambda(I_i) = \lambda(I).$$

To do this, we can assume that I is an interval and the general case would follow by the finite additivity that we forced to hold. Next, we can also assume that each I_i is an interval. Suppose that I is bounded then its closure is compact by Heine-Borel theorem. We can now enlarge each I_i to open intervals. Namely, for each $i \ge 1$ and $\varepsilon > 0$, we let

$$I_i' = (a_i - \varepsilon_i, b_i + \varepsilon_i)$$

where $\varepsilon_i = \varepsilon/2^i$ and a_i, b_i are defining I_i. We now shrink I to

$$I' = [a + \varepsilon, b - \varepsilon].$$

We agree that I_i' is a collection of open sets that covers I'. As I' is compact, we can find a finite sub-cover. Then we can deduce that

$$\lambda(I) - 2\varepsilon \le \sum_{i:\text{a finite index set}} \lambda(I_i) + 2\varepsilon \le \sum_i \lambda(I_i) + 2\varepsilon.$$

On the other hand, it is clearly true that

$$\sum_{i:\text{any finite index set}} \lambda(I_i) = \lambda(\cup_{i:\text{any finite index set}} I_i) \le \lambda(I).$$

From here we conclude that

$$\sum_i \lambda(I_i)$$

converges and the limit is equal to $\lambda(I)$. This is what is to be proved.

2.1.4.4. *Step 4:* We can now use Carathéodory's extension theorem (as a black-box). The outcome is that we can extend λ to all $\mathcal{B}(\mathbb{R})$. This λ is a measure. This measure is called the Lebesgue measure on $\mathbb{R}$. We can also construct the Lebesgue measure for $\mathbb{R}^k, k \geq 2$.

2.1.4.5. *The regularity:* It can be proved that the Lebesgue measure is regular. For example, when $k = 1$ we can first check the regularity at each interval. This is obvious from the construction. We want to prove the regularity at all Borel measurable sets. To do this, we call a set to be regular if $\lambda(A)$ satisfies the condition of regularity. Next, we check that being regular is kept by taking complements, countable unions and countable intersections. Thus we conclude that all Borel measurable sets are regular and this proves the regularity of the Lebesgue measure.

2.2. The Hausdorff measure and Hausdorff dimension

We now want to extend the notion of the Lebesgue measure. Before that, we want to introduce some motivations. We have constructed for each $k \geq 1$ a Lebesgue measure λ_k on $\mathbb{R}^k$. It is intuitive to associate a dimension for λ_k, which is k. Next, consider $\mathbb{R}^k$ for some $k \geq 2$. Let $L \subset \mathbb{R}^k$ be a line. It can be checked that $\lambda_k(L) = 0$. This captures the intuition that L should have dimension one which is smaller than k. Therefore its k-volume should be zero. More generally, for each manifold M of dimension $s \leq k-1$ we have $\lambda_k(M) = 0$.

From here, we encounter a natural question: Can we introduce naturally a measure on M that captures the same idea as the Lebesgue measure? To have some intuitions, for L being a line, we can intuitively think that it should carry a notion of length. That is, for each $A \subset \mathbb{R}^k$, we can associate $A \cap L$ a length, i.e. we can transfer λ_1 to L in a natural way. More generally, for each manifold M of dimension s, we can also transfer λ_s to M in a natural way. To do this, we need to invoke the definition of a manifold that it is locally an Euclidean ball. We can then transfer the Lebesgue measure locally to M. The only problem is achieving consistency. Namely, for two charts A_1, A_2, we need to make sure that the Lebesgue measure transferred to A_1, A_2 are consistent with that to $A_1 \cap A_2$. This can be accomplished with the help of a partition of unity. We omit further details.

Now we have a more general question: Can we introduce a measure on a Borel $F \subset \mathbb{R}^k$ which may not be a manifold? To answer this question, we will introduce the notion of the Hausdorff measure.

2.2.1. Hausdorff measure. Let $k \geq 1$ be an integer. We consider the Euclidean space $\mathbb{R}^k$. Let $s \in [0, k]$ be a number. We are going to construct a regular measure $\mathcal{H}^s$ on $\mathbb{R}^k$ that can be interpreted as an s-dimensional measure.

Definition 2.2.1:s-Hausdorff measure

For each $F \subset \mathbb{R}^k$, define
$$\mathcal{H}^s_\delta(F) = \inf \left\{ \sum \mathrm{diam}(B_i)^s : \cup_i B_i \supset F, \mathrm{diam}(B_i) \leq \delta \right\}$$
where $\delta > 0$ is a positive number. Moreover, we write
$$\mathcal{H}^s(F) = \lim_{\delta \to 0} \mathcal{H}^s_\delta(F).$$

The limit $\lim_{\delta\to 0}\mathcal{H}^s_\delta(F)$ exists and it can be infinite. If $s = k$, then we can check that $\mathcal{H}^k$ is a scaled version of the Lebesgue measure λ_k. More generally, for each integer $s \leq k-1$ and an s-manifold M, $\mathcal{H}^s_{|M}$ is a scaled version of the Lebesgue measure on M that we constructed above. It can be checked that $\mathcal{H}^s$ is a measure on $\mathbb{R}^k$ with respect to the Borel σ-algebra. An intuitive way of understanding the construction of the Hausdorff measure is that we really want to associate the ball B_r of radius r with a measure $\approx r^s$ instead of $\approx r^d$.

If we want to check the regularity of $\mathcal{H}^s$, we face a problem that unless $s = k$, any open set of $\mathbb{R}^k$ has infinite measure. Thus we conclude that $\mathcal{H}^s$ is not regular on $\mathbb{R}^k$ if $s < k$. However, it is possible to find subsets $F \subset \mathbb{R}^k$ such that the restriction $\mathcal{H}^s_{|F}$ is regular. For example, this is the case when s is an integer and M is a smooth s-dimensional manifold.

2.2.2. Hausdorff dimension. We now prove the following result.

Theorem 2.2.2:

Let $F \subset \mathbb{R}^k$ be a Borel set. Then there is a number $s \in [0,k]$ such that for each $s' < s$
$$\mathcal{H}^{s'}(F) = \infty$$
and for each $s' > s$,
$$\mathcal{H}^{s'}(F) = 0.$$

Remark 2.2.3:

On the other hand, there is no control on $\mathcal{H}^s(F)$. It can be $0, <\infty, \infty$.

Definition 2.2.4:Hausdorff dimension

Let $F \subset \mathbb{R}^k$ be a Borel set. Then its Hausdorff dimension is
$$\dim_{\mathrm{H}} F = \inf\{s \in [0,k] : \mathcal{H}^s(F) = 0\} = \sup\{s \in [0,k] : \mathcal{H}^s(F) = \infty\}.$$

Proof of Theorem 2.2.2. Let $s > 0$ be such that $\mathcal{H}^s(F) = 0$. We are going to show that $\mathcal{H}^{s'}(F) = 0$ for all $s' > s$. To do this, let $\varepsilon > 0$. Then for all small enough $1 > \delta > 0$,
$$\mathcal{H}^s_\delta(F) < \varepsilon.$$
Thus, we can find a collection of sets $\{B_i\}_i$ that covers F and $\mathrm{diam}(B_i) < \delta$ for each i. Moreover, we have
$$\sum_i \mathrm{diam}(B_i)^s < 2\varepsilon.$$
Then as $\mathrm{diam}(B_i)^{s'-s} < 1$, we see that
$$\sum_i \mathrm{diam}(B_i)^{s'} = \sum_i \mathrm{diam}(B_i)^{s'-s}\mathrm{diam}(B_i)^s < 2\varepsilon.$$
This shows that for all small enough δ,
$$\mathcal{H}^{s'}_\delta(F) < 2\varepsilon.$$

Therefore $\mathcal{H}^{s'}(F) \le 2\varepsilon$. Since this holds for each $\varepsilon > 0$, we see that

$$\mathcal{H}^{s'}(F) = 0.$$

Now, let $s > 0$ be such that $\mathcal{H}^s(F) > 0$. We want to show that $\mathcal{H}^{s'}(F) = \infty$ for all $s' < s$. This time observe that $\operatorname{diam}(B_i)^{s'-s} > \delta^{-1}$ as long as $\operatorname{diam}(B_i) < \delta$. The rest of the argument is very similar to what we have done earlier in this proof. From here the proof is finished. □

Hausdorff dimension satisfies some useful properties.

Theorem 2.2.5:

Let $F \subset \mathbb{R}^k$ be a Borel set. Then we have

$$\dim_{\mathrm{H}} F \in [0, k].$$

For each set $F' \supset F$, we have

$$\dim_{\mathrm{H}} F' \ge \dim_{\mathrm{H}} F.$$

For any countable collection $F_1, F_2, \dots,$ of Borel sets, we have

$$\dim_{\mathrm{H}}(\cup_i F_i) = \sup\{\dim_{\mathrm{H}} F_i\}_i.$$

Proof. We only show the last point. Let $s > \sup\{F_i\}_i$. Then $\mathcal{H}^s(F_i) = 0$ for each i. Then we see that

$$\mathcal{H}^s(\cup_i F_i) \le \sum_i \mathcal{H}^s(F_i) = 0.$$

This shows that

$$\dim_{\mathrm{H}} \cup_i F_i \le s.$$

Thus we see that

$$\dim_{\mathrm{H}} \cup_i F_i \le \sup\{F_i\}_i.$$

On the other hand, if $s < \sup\{\dim_{\mathrm{H}} F_i\}_i$, then there is at least one index i_0 with $\dim_{\mathrm{H}} F_{i_0} > s$. This implies that

$$\mathcal{H}^s(F_i) = \infty.$$

Therefore we see that

$$\mathcal{H}^s(\cup_i F_i) \ge \mathcal{H}^s(F_{i_0}) = \infty.$$

This proves that

$$\dim_{\mathrm{H}} \cup_i F_i \ge s.$$

Thus we see that

$$\dim_{\mathrm{H}} \cup_i F_i \ge \sup\{F_i\}_i.$$

This concludes the proof. □

As an application, we see that the set of rational numbers has Hausdorff dimension zero because it is a countable union of points and each point has Hausdorff dimension zero. Later, we shall prove a slightly stronger assertion that the uncountable set of Liouville numbers (which are 'almost rationals') also has Hausdorff dimension zero.

2.2.3. Simple examples. It is in general not an easy task to determine the Hausdorff dimension of a given set. Here we only provide some simple examples. Later, we will prove further properties of the Hausdorff measure/dimension and be able to determine the Hausdorff dimension for more sophisticated sets.

Example 2.2.6:Manifolds

Let M be an s-dimensional C^∞-manifold in $\mathbb{R}^k$ where $k \geq 1$. Now, s must be an integer. In this case, one can follow the construction of the Lebesgue measure on $\mathbb{R}$ and deduce that

$$\mathcal{H}^s(M) > 0.$$

Moreover, if M is compact, then $\mathcal{H}^s(M) < \infty$. This shows that

$$\dim_{\mathrm{H}} M = s$$

which is the topological dimension of M. In particular, this fits our intuition that a point should have dimension 0, a curve should have dimension one, etc.

Example 2.2.7:Liouville number

Let $x \in \mathbb{R}$. We say that x is Liouville if for each $\omega > 0$, there are infinitely many coprime integers $p, q > 0$ such that

$$\left| x - \frac{p}{q} \right| < \frac{1}{q^\omega}.$$

Liouville showed that such a number must be either rational or transcendental. We want to consider the set of Liouville numbers in $[0, 1]$, denoted as L. We show that $\dim_{\mathrm{H}} L = 0$. For each $q > 0$, we use A_q^ω for the union of intervals

$$\bigcup_{p \in \{0, \dots, p\}} B_{1/q^\omega}(p/q).$$

We use the cover $\mathcal{C}_q$

$$\bigcup_{n \geq q} A_n^\omega$$

by using the intervals in each A_q^ω. Observe that for $s = 3/\omega$,

$$\sum_{n \geq q} \sum_{p=0}^{n-1} (\frac{1}{n^\omega})^s = \sum_{n \geq q} \frac{n}{n^{s\omega - 1}} = \sum_{n \geq q} \frac{1}{n^2} \leq \frac{C}{q}$$

for some constant $C > 0$. For any q, the cover $\mathcal{C}_q$ will be enough to cover each $x \in L$ because $x \in A_n^\omega$ for infinitely many n and in particular for one such $n > q$. For each $\delta > 0$, we choose $q > 100/\delta$ and conclude that

$$\mathcal{H}^s_{\delta^\omega}(L) \leq C' \delta$$

for some constant $C' > 0$. Therefore, $\mathcal{H}^s(L) = 0$. This shows that

$$\dim_{\mathrm{H}} L \leq s = 3/\omega.$$

Since ω can be chosen to be arbitrarily large, we see that $\dim_{\mathrm{H}} L = 0$.

Remark 2.2.8:

It is a more delicate problem to compute the Hausdorff dimension of the following set for some fixed $\omega > 2$,

$$W_\omega = \{x \in [0,1] : |x - p/q| < 1/q^\omega\}.$$

It is known to be $2/\omega$. By refining the arguments in Example 2.2.7, it is possible to show that

$$\dim_{\mathrm{H}} W_\omega \leq \frac{2}{\omega}.$$

The more difficult direction was proved by Besicovitch and Jarnik separately. See [6].

2.3. Box counting dimension

As computing the Hausdorff dimension for a given set is difficult, it is interesting to find some ways to estimate it. One way is to use the notion of the box-counting dimension. Other than serving as a bound for the Hausdorff dimension, the box-counting dimension is interesting in its own right.

Definition 2.3.1:Box counting number

Let $F \subset \mathbb{R}^k$ be a set. Let $\delta > 0$. The number $N_\delta(F)$ is defined to be the smallest possible amount of δ-balls whose union covers F. If there is no such finite cover, we set $N_\delta(F) = \infty$.

Definition 2.3.2:Box counting dimension

Let $F \subset \mathbb{R}^k$ be a compact set. Then we define

$$\overline{\dim}_{\mathrm{B}} F = \limsup \frac{\log N_\delta(F)}{\log \delta^{-1}},$$

$$\overline{\dim}_{\mathrm{B}} F = \liminf \frac{\log N_\delta(F)}{\log \delta^{-1}}.$$

If the above two numbers are equal, we can further set them to be $\dim_{\mathrm{B}} F$. They are called the upper box dimension, lower box dimension and the box-counting dimension of F.

From the definition, it is possible to show the following result.

Theorem 2.3.3:Some properties of box dimension

- If $F \subset \mathbb{R}^k$, then $\overline{\dim}_{\mathrm{B}} F = \overline{\dim}_{\mathrm{B}} \overline{F}, \underline{\dim}_{\mathrm{B}} F = \underline{\dim}_{\mathrm{B}} \overline{F}$.
- If $F \subset F' \subset \mathbb{R}^k$, then $\overline{\dim}_{\mathrm{B}} F \leq \overline{\dim}_{\mathrm{B}} F', \underline{\dim}_{\mathrm{B}} F = \underline{\dim}_{\mathrm{B}} \overline{F}$.
- If $F_1, \ldots, F_n$ are compact, then

$$\overline{\dim}_{\mathrm{B}} \bigcup_i F_i = \max\{\overline{\dim}_{\mathrm{B}} F_i\}_{i=1,\ldots,n}.$$

- If $F \subset \mathbb{R}^k$, then $\dim_{\mathrm{H}} F \leq \underline{\dim}_{\mathrm{B}} F$.

Remark 2.3.4:

The third point is not true for lower box dimensions. It is possible to find examples F_1, F_2 so that

$$\underline{\dim}_{\mathrm{B}}(F_1 \cup F_2) > \max\{\underline{\dim}_{\mathrm{B}} F_1, \underline{\dim}_{\mathrm{B}} F_2\}.$$

PROOF. We only prove the last point and leave the rest to the reader. Let $\underline{\dim}_{\mathrm{B}} F = s$. From the definition, for each $\varepsilon > 0$, we know that it is possible to find arbitrarily small $\delta > 0$ such that

$$N_\delta(F) < \left(\frac{1}{\delta}\right)^{s+\varepsilon}.$$

We can then use any minimal cover and find that

$$\mathcal{H}^{s+\varepsilon}_\delta(F) < 1.$$

This implies that $\dim_{\mathrm{H}} F \leq s + \varepsilon$. Thus we see that $\dim_{\mathrm{H}} F \leq s$. □

2.4. Dimensions of Cartesian product sets

Given a compact set $F \subset \mathbb{R}^k$. Suppose that we know its (Hausdorff/Box/Assouad) dimension $\dim F$. We can ask what is the dimension of $F \times F \subset \mathbb{R}^{2k}$. One intuition is that the dimension should be doubled. We will see that this intuition is true in most of the cases and false in some cases. For the study of the Hausdorff dimension, we will need some Fourier analytic tools. This will be carried out in the next chapter. Here, we focus on box-counting and Assouad dimensions.

Theorem 2.4.1:

Let $F \subset \mathbb{R}^k$ be compact. Then $\underline{\dim}_{\mathrm{B}}, \overline{\dim}_{\mathrm{B}}(F \times F) = 2(\underline{\dim}_{\mathrm{B}}, \overline{\dim}_{\mathrm{B}} F)$.

PROOF. For each $\delta > 0$. For each δ-covering of F, we can create a δ-covering of $F \times F$ whose δ-boxes are the Cartesian products of those boxes for the covering of F. This shows that

$$\inf N_\delta(F \times F) \leq (\inf N_\delta(F))^2.$$

Can we find a more efficient δ-covering for $F \times F$? In order to answer this question, we find a δ-separated set F_δ in F. Due the the covering property, we see that such F_δ must be

$$\asymp \inf N_\delta(F).$$

We can then consider the finite set $F_\delta \times F_\delta$. This finite set is δ-separated as well. Thus we see that

$$\inf N_\delta(F \times F) \gg (\inf N_\delta(F))^2.$$

From here the result follows. □

Now let us generalise the problem by considering $E \times F$ where E, F are two different compact sets in $\mathbb{R}^k$. A fraction of the above proof still works and shows that

$$\overline{\dim}_{\mathrm{B}} E + \underline{\dim}_{\mathrm{B}} F \leq \overline{\dim}_{\mathrm{B}}(E \times F) \leq \overline{\dim}_{\mathrm{B}} E + \overline{\dim}_{\mathrm{B}} F$$

and

$$\underline{\dim}_{\mathrm{B}} E + \underline{\dim}_{\mathrm{B}} F \leq \underline{\dim}_{\mathrm{B}}(E \times F) \leq \overline{\dim}_{\mathrm{B}} E + \underline{\dim}_{\mathrm{B}} F.$$

All the above inequalities can be strict. Let us see one example of some of the strict inequalities.

Example 2.4.2:

We start with a scheme for constructing compact sets in $[0,1]$. Consider a one-sided infinite sequence $\omega \in \{0,1\}^{\mathbb{N}}$. For each $j \geq 0$, we write ω_j as the j-th digit. For $i = 0$, we read ω_0. If $\omega_0 = 0$ we split $[0,1]$ into equal half and keep both of them. If $\omega_0 = 1$ we split $[0,1]$ into equal half and keep only the left interval. For $i = 1$, performs the same construction for each of the half intervals from the previous step (there could be one or two). This procedure goes on indefinitely and we have a limit compact set K_ω.
We choose a fast increasing sequence, say, $i_1 = 0, i_k = 2^{2^k}, k \geq 2$. We choose ω such that $\omega_i = 1$ whenever $i_{2k-1} < i \leq i_{2k}$ and 0 otherwise. We then find $E = K_\omega$. Similarly, we choose ω' that is the sequence by flipping all digits of ω. We obtain $F = K_{\omega'}$.
It can be checked that $\underline{\dim}_{\mathrm{B}} E, F = 0, \overline{\dim}_{\mathrm{B}} E, F = 1$ and $\underline{\dim}_{\mathrm{B}} E \times F = \overline{\dim}_{\mathrm{B}} E \times F = 1$.

2.4.1. Examples. It is relatively simpler to compute $\overline{\dim}_{\mathrm{B}}, \underline{\dim}_{\mathrm{B}}$ for subsets. First, we see that

$$\overline{\dim}_{\mathrm{B}} \mathbb{Q} \cap [0,1] = 1.$$

This is because the closure of $\mathbb{Q}$ is $\mathbb{R}$. We also know that $\dim_{\mathrm{H}} \mathbb{Q} = 0$ and thus in general Hausdorff dimension of a set can be strictly smaller than its box-counting dimension. Next, we compute the box-counting dimension for some fractals.

Example 2.4.3:The middle-third Cantor set

Consider the middle-third Cantor set K as in Example 1.3.1. We can use the scales $1/3^n$ for integers $n \geq 1$. It is possible to see that for each $n \geq 1$, F can be covered by 2^n many disjoint unions of intervals of length $1/3^n$. This means that this covering is the smallest possible. Therefore we see that

$$N_{1/3^n}(K) = 2^n.$$

From here, it is not hard to show that $\dim_{\mathrm{B}} K = \log 2 / \log 3$.

Example 2.4.4:Kakeya sets in $\mathbb{R}^2$

A Kakeya set in $\mathbb{R}^k$ is a set K that contains unit line segments in every direction. Namely, for each $\theta \in \mathbb{S}^{k-1}$, there is a $b \in \mathbb{R}^k$ such that

$$l_{\theta,b} = \{b + t\theta : t \in [0,1]\} \subset K.$$

It is a difficult problem to determine $\dim_{\mathrm{B}}(\overline{\dim}_{\mathrm{B}}, \underline{\dim}_{\mathrm{B}})K$ for Kakeya sets. It is largely believed that

$$\underline{\dim}_{\mathrm{B}} K = k.$$

For $k \geq 3$, this is still unsolved. We now verify this for $k = 2$. We state the result as a theorem.

Theorem 2.4.5:

Let $K \subset \mathbb{R}^2$ be a Kakeya set. Then

$$\underline{\dim}_{\mathrm{B}} K = 2.$$

Remark 2.4.6:

On the other hand, it is possible to find Kakeya sets $K \subset \mathbb{R}^2$ with zero Lebesgue measure. Thus the situation for Kakeya sets is not very trivial.

Proof. Let $\delta > 0$ be a small number. We first consider a set of directions which is δ-separated. For this, we mean a subset $S_\delta \subset S^1$ such that each pair $\theta, \theta' \in S_\delta$ has a distance at least δ apart. It is true that $\# S_\delta \ll 1/\delta$. This means that $\# S_\delta \leq C/\delta$ for some $C > 0$. On the other direction, we can also find S_δ so that

$$\# S_\delta \asymp 1/\delta.$$

This means that $c/\delta \leq \# S_\delta \leq C/\delta$ for some $c, C > 0$. We will refine this choice later in the proof. For now, we will dive into the proof to see the geometric ideas.

For each $\theta \in S_\delta$, there is a $b_\theta \in \mathbb{R}^2$ such that

$$l^\delta_{\theta, b_\theta} \subset K^\delta.$$

Here for a set $A \subset \mathbb{R}^2$, A^δ is the δ-neighbourhood of A, namely, the points in $\mathbb{R}^2$ that are δ-close to A. We now give a lower bound for

$$N_\delta(\cup_{\theta \in S_\delta} l^\delta_{\theta, b_\theta}).$$

For convenience, we will write l_θ for l_{θ, b_θ}. For each $\theta, \theta' \in S_\delta$, we want to examine the intersection

$$l^\delta_\theta \cap l^\delta_{\theta'}.$$

It can of course be empty. However, we want to have an upper bound for the size of the intersection. We estimate the Lebesgue measure,

$$\lambda(l^\delta_\theta \cap l^\delta_{\theta'}) \ll \frac{\delta^2}{d(\theta, \theta')}$$

where $d(\theta, \theta')$ is the arc distance between $\theta, \theta' \in S^1$. To see this, we can first check that $l^\delta_\theta \cap l^\delta_{\theta'}$ can be contained in a rectangle of sizes $\asymp \delta, \delta/d(\theta, \theta')$. From here we can use

$$\lambda(\cup_{\theta \in S_\delta} l_\theta) \geq \sum_\theta \int_{l^\delta_\theta} dx - \sum_{\theta \neq \theta'} \int_{l^\delta_\theta \cap l^\delta_{\theta'}} dx.$$

To see this, for each $x \in \mathbb{R}^2$, observe that $\sum_\theta l^\delta_\theta(x) = k \geq 1$, then

$$\sum_{\theta \neq \theta'} l^\delta_\theta(x) l^\delta_{\theta'}(x) \geq k(k-1) \geq k - 1.$$

where we used l^δ_θ for indicator functions. The first sum in (2.4.1) is

$$\asymp 1.$$

We now check the second sum,

$$\sum_{\theta\neq\theta'}\frac{\delta^2}{d(\theta,\theta')}=\delta^2\sum_{\theta}\sum_{\theta'}\frac{1}{d(\theta,\theta')}\ll\delta^2\sum_{\theta}\log\frac{1}{\delta}\ll\log\frac{1}{\delta}.$$

Unfortunately, the second sum is too large compared to the first sum. We now need to refine our choice of δ-tubes. Instead of using a δ-separated direction set S_δ, we can choose $\varepsilon>0$ and use $S_{\delta^{1-\varepsilon}}$. Then we will see that the first sum in (2.4.1) becomes

$$\asymp\delta^{\varepsilon}.$$

The second sum is now

$$\ll\delta^{2\varepsilon}\log\frac{1}{\delta}=o(\delta^{\varepsilon}).$$

From here it is clear that

$$N_\delta(\cup_{\theta\in S_\delta}l^{\delta}_{\theta,b_\theta})\gg\delta^{2-\varepsilon}.$$

This shows that

$$\underline{\dim}_{\mathrm{B}}K\geq 2-\varepsilon.$$

Since this holds for all $\varepsilon>0$, we see that

$$\underline{\dim}_{\mathrm{B}}K=2.$$

□

Example 2.4.7:Takagi graphs

We saw in Section 1.1.2 a class of functions with rough graphs (Weierstrass functions). We can compute their box dimensions. However, for those graphs, there are some technical difficulties that are not very pleasant to deal with. For this reason, we consider here a simplified problem.
Let $T:[0,1]\to[0,1/2]$ be the tent function. Namely, $T(x)=x$ for $x\in[0,1/2]$ and $1-x$ for $x\in[1/2,1]$. We then extend T to $\mathbb{R}$ in a $\mathbb{Z}$-periodic manner. We still use T for the extended function. Next, for numbers $0<a<1<b$ we consider

$$T_{a,b}(x)=\sum_{j\geq 0}a^nT(b^nx).$$

The graph $G_{a,b}$ of $T_{a,b}$ over $[0,1]$ is fractal-like if $ab>1$. We want to compute $\dim_{\mathrm{B}}G_{a,b}$ when b is an integer. This will be the content of the next theorem.

Theorem 2.4.8:Takagi graphs

Let $0<a<1<b$ be such that $ab>1$ and $b\in\mathbb{N}$. Then

$$2>\dim_{\mathrm{B}}G_{a,b}=2-\frac{\log(1/a)}{\log b}>1.$$

PROOF. Let $n\geq 0$. Consider the finite sum

$$T^n_{a,b}(x)=\sum_{j=0}^{n}a^nT(b^nx).$$

Its graph is a union of $2b^n$ many line segments whose projections on the X-axis have length $1/(2b^n)$. To get a finer analysis, we decompose the unit interval into a disjoint union of pieces of length $1/b^n$. There are exactly b^n many of them. Over each such interval, say, I, $T^n_{a,b}$ is a union of two line segments. We now study the slopes of those line segments. To do this, we introduce a convenient coding system for those intervals of length $1/b^n$. Consider the leftmost point of each interval. Let I, x be such an interval and a point. It is a rational number with a b-expansion with at most n digits. This finite sequence of digits in $\{0,\dots,b-1\}^n$ can be used as the code for I. We write it as ω_I.

Let I, x be as above. For each point $y\in I$, each $j<n$, we consider the fractional part of b^jy. It is determined by the sequence $(\omega_I)^n_{j+1}$ which is the segment of ω_I from the $(j+1)$-th to the last place. We only need to know whether the fractional part of b^jy is smaller or larger than $1/2$. Thus, we can make a translation of the code ω_I as follows.

For each digit $(\omega_I)_j\leq (b-2)/2$, we replace it with $'-'$. For each digit $(\omega_I)_j\geq b/2$, we replace it with $'+'$. If b is odd, then there is a further possibility $(b-1)/2$ for which we replace it with $'?'$. After this, we associate this new sequence to y and write it as ω_y.

Observe that $(T^n_{a,b})'(y)=\sum_{j=0}^n a^nb^nT'(b^ny)$ where T' is 1 on $(0,1/2)$ and -1 on $(1/2,1)$ and it is $\mathbb{Z}$-periodic. We can then write

$$(T^n_{a,b})'(y)=\sum_{j=0}^{n-1}a^jb^j\omega_{y,j}\pm(ab)^n.$$

The $\pm$ is determined according to $\{b^ny\}<$ or $>1/2$. We interpret $\omega_{y,j}$ as being 1 or -1 for $+,-$ resp. For each $'?'$, we cannot decide the sign. However, in any case, we have the following inequality

$$(*)\qquad \left|(T^n_{a,b})'(y)-\sum_{j=0,\omega_{y,j}\in\{\pm\}}^{n}a^nb^n\omega_{y,j}\right|\ll(ab)^n.$$

As I ranges over all choices of intervals of length $1/b^n$, ω_I ranges over all possible length n expansions with base b. Suppose that b is even. Then the sequence ω_y for each $y\in(0,1)$ does not have $'?'$. Therefore the value $(T^n_{a,b})'(y)$ is determined as a sum like $\pm(ab)^j$. As ω_I ranges over all possible length n expansions with base b, the possible ω_y ranges over sequences of $\pm$ with length n. Such correlation is equidistributed in the sense that each $\pm$ sequence of length n corresponds to exactly $b^n/2^n$ many b-sequences. We want to study the distribution of the values $\sum_{j=0}^n\pm(ab)^j$.

For $ab>1$, we can consider the following multi-set

$$U_n=\left\{(ab)^{-n}\sum_{j=0}^n\pm(ab)^n\right\}\subset[-ab/(ab-1),ab/(ab-1)].$$

We can also construct μ_n as the counting probability measure on U_n (with possible multiplicities). Then μ_n has a limit μ which is a probability measure on $[-ab/(ab-1),ab/(ab-1)]$. (This follows from the weak compactness. This step is N.E. although it also follows from some elementary arguments.) We claim that this measure μ is not concentrated on 0 in the sense that there is some $r>0$ so that $\mu([-r,r])<1/2$.

For the claim, if it is not true, then $\mu(\{0\}) \geq 1/2$. As we will learn later (Theorem 4.2.9), μ is (non-trivially) self-similar and therefore it cannot have atomic support. For now, we just take this result as a blackbox. This shows the claim. (This claim is obvious if $ab \geq 2$. Try to prove it in an elementary way!)

From this claim, we see that there is some $r > 0$, for that for each large enough n,

$$\mu_n([-r, r]) < 1/2.$$

This shows that the multi-set

$$(ab)^n U_n \cap [-r(ab)^{n+1}/(ab-1), r(ab)^{n+1}/(ab-1)]$$

has at most $(1/2)(2^n)$ many of elements. From here, we know that the b-sequences so that

$$\left|\sum_{j=0}^{n-1} a^j b^j \omega_{y,j}\right| > r\frac{(ab)^{n+1}}{ab-1}$$

must be at least $b^n/2$ in quantity. Such a b-sequence corresponds to an interval I of length $1/b^n$. On this interval, $(T^n_{a,b})'$ is

$$\sum_{j=0}^{n-1} a^j b^j \omega_{y,j} \pm (ab)^n,$$

where the $\pm$ is $-$ on the left half and $+$ on the right half. Let $I = [x_L, x_R]$. By integration over I, we then see that (the $\pm(ab)^n$ part does not contribute),

$$|T^n_{a,b}(x_R) - T^n_{a,b}(x_L)| = \left|\frac{1}{b^n}\sum_{j=0}^{n-1} a^j b^j \omega_{y,j}\right| \gg a^n.$$

From the construction, we have

$$T_{a,b} = T^n_{a,b} + T^{>n}_{a,b}$$

where

$$T^{>n}_{a,b}(x) = \sum_{j>n} a^n T(b^n x).$$

Observe that for each interval $I = [x_L, x_R]$ as discussed above, $T^{>n}_{a,b}(x_L) = T^{>n}_{a,b}(x_R) = 0$. Therefore we see that for at least $b^n/2$ many such intervals I,

$$|T_{a,b}(x_R) - T_{a,b}(x_L)| = |T^n_{a,b}(x_R) - T^n_{a,b}(x_L)| \gg a^n.$$

We can now describe the covering property of $G_{a,b}$. We use boxes of size $1/b^n$. For each interval I so that the above holds, we need

$$\gg (ab)^n$$

many such boxes to cover $G_{a,b}$ over I. Therefore we can already see that

$$N_{1/b^n}(G_{a,b}) \gg \frac{b^n}{2}(ab)^n \gg a^n b^{2n}.$$

On the other hand, from (*) we see that for each I,

$$\sup_{x,x'\in I} |T^n_{a,b}(x) - T^n_{a,b}(x')| \ll (ab)^n \frac{1}{b^n} = a^n$$

where $\ll$ does not depend on the choice of I. It is also clear that

$$\max T^{>n}_{a,b} \ll a^n.$$

From here we see that to cover $G_{a,b}$ over any I, we need

$$\ll (ab)^n$$

many of $1/b^n$-boxes. From here we see that

$$N_{1/b^n}(G_{a,b}) \ll b^n (ab)^n = a^n b^{2n}.$$

Thus we have the following box-counting estimate

$$N_{1/b^n}(G_{a,b}) \asymp a^n b^{2n}.$$

This implies that

$$\dim_{\mathrm{B}} G_{a,b} = 2 - \frac{\log(1/a)}{\log b}.$$

We finished the proof for even b. For odd b, some modifications are needed. First, we only consider the $'\pm, ?'$ sequences that are without the symbol $'?'$. There are $(b-1)^n$ many b-sequences that give birth to such $'?'$-free sequences. Out of those $(b-1)^n$ many b-sequences, at least half of them will provide us with the estimate (recall that x_L, x_R are the left, right endpoints)

$$|T_{a,b}(x_R) - T_{a,b}(x_L)| \gg a^n.$$

From here we obtain that

$$a^n b^n (b-1)^n \ll N_{1/b^n}(G_{a,b}) \ll a^n b^{2n}.$$

Thus we see that

$$1 + \frac{\log(b-1)}{\log b} - \frac{\log(1/a)}{\log b} \underline{\dim}_{\mathrm{B}} G_{a,b} \leq \overline{\dim}_{\mathrm{B}} G_{a,b} \leq 2 - \frac{\log(1/a)}{\log b}.$$

To prove the theorem, we must consider the symbol $'?'$ more carefully. The above argument for even b still holds here as long as the symbol $'?'$ only appears at small values j. More precisely, for each $\varepsilon > 0$, we see that for some large constant K, as long as ω_y is $'?'$-free for digits from the $(j-K)$-th to the last place, then

$$\left| (T^n_{a,b})'(y) - \sum_{j=0, \omega_{y,j} \in \{\pm\}}^{n} a^n b^n \omega_{y,j} \right| \leq \varepsilon (ab)^n.$$

There are much more b-sequences with this weaker $'?'$-free property. More precisely, there are

$$b^{n-k}(b-1)^K$$

many of them. From here we see that

$$\underline{\dim}_{\mathrm{B}} G_{a,b} \geq \lim_{n\to\infty} \left(\frac{2n-K}{n} + \frac{K}{n}\frac{\log(b-1)}{\log b} - \frac{\log(1/a)}{\log b} \right) = 2 - \frac{\log(1/a)}{\log b}.$$

This finishes the proof. □

2.5. Assouad/lower dimensions (not lectured), N.E.

After we introduced Hausdorff and box-counting dimensions, it became clear that there are different notions of dimensions that one can explore. Usually, a notion of dimension is introduced when people are considering a certain problem involving certain geometric or measure-theoretic problems of sets or measures. After some time, the notion of dimension became to be standalone as an independent subject. In this last section, we introduce one such example.

Hausdorff dimension and box-counting dimension are in some sense a global property. To see this, let us consider a disjoint union E of a point and an interval on $\mathbb{R}^2$. Then it is possible to check that

$$\dim_{\mathrm{H}} E = \dim_{\mathrm{B}} E = 1.$$

Thus we picked up the interval part of E. The information on the point part is then missing. In order to rectify this, we can introduce the following notions.

Definition 2.5.1:Assouad and lower dimensions

Let $F \subset \mathbb{R}^k$ be a set. We define the Assouad dimension of F to be

$$\dim_{\mathrm{A}} F = \inf\{\alpha > 0 : 0 < r < R < 1, \sup_{x \in F} N_r(B_R(x) \cap F) \ll (R/r)^{\alpha}\}.$$

We define the lower dimension of F to be

$$\dim_{\mathrm{L}} F = \sup\{\alpha > 0 : 0 < r < R < 1, \inf_{x \in F} N_r(B_R(x) \cap F) \gg (R/r)^{\alpha}\}.$$

Thus, for the set E considered earlier in this section, we deduce that

$$\dim_{\mathrm{A}} E = 1, \dim_{\mathrm{L}} E = 0.$$

Originally, the Assouad dimension was introduced for considering an embedding problem of metric spaces. Later on, this notion of dimension was reintroduced by different people working on different problems, e.g. Furstenberg's notion of $*$-dimension.

Thus the Assouad dimension wants to pick up the densest part of a set and the lower dimension wants to pick up the sparsest part. If a set has an equal Assouad/lower dimension, then it is quite homogeneous in some sense. A union of a point of an interval is not very homogeneous in this sense while a single interval is quite homogeneous. We now compute $\dim_{\mathrm{A}} K, \dim_{\mathrm{L}} K$ for the middle third Cantor set.

Example 2.5.2:The middle third Cantor set

Let $K \subset [0, 1]$ be the middle third Cantor set. Let $0 < r < R < 1$. Let $x \in K$ and consider the R-ball $B_R(x)$. There is no loss of generality to consider r, R being powers of $1/3$. For each $x \in K$, the ball $B_R(x)$ will intersect at most two intervals of length R in the construction of the middle third Cantor set. For each of those intervals of length R, we need exactly $(R/r)^{\log 2/\log 3}$ many intervals of length r to cover it. Thus we see that

$$\sup_{x\in K} N_r(B_R(x)\cap K) \asymp \left(\frac{R}{r}\right)^{\log 2/\log 3}.$$

This shows that $\dim_{\mathrm{A}} K = \log 2/\log 3$.
On the other direction, observe that $B_R(x)$ will contain at least one interval of length $R/3$ in the construction of the middle third Cantor set. From here, it can be checked that

$$\dim_{\mathrm{L}} K = \log 2/\log 3.$$

Thus, we see that the middle third Cantor set is quite homogeneous. Later on, we will introduce the notion of AD-regularity and generalise the above example.

CHAPTER 3

The geometry of Fractals and Fractal measures

3.1. Regularity of measures

3.1.1. AD-regularity. Consider the middle third Cantor set K, we know that it has equal Assouad and lower dimensions. Intuitively, this means that the densest and spareest parts of K are of the same size. In this section, we are going to introduce a yet more precise notion of homogeneity of this kind.

Definition 3.1.1:Support of measure

Let $k \geq 1$ be an integer. Let μ be a Borel measure on $\mathbb{R}^k$. Its support, $\mathrm{supp}(\mu)$, is the collection of points

$$\mathrm{supp}(\mu) = \{x \in \mathbb{R}^k : \forall r > 0, \mu(B_r(x)) > 0\}.$$

Lemma 3.1.2:

$\mathrm{supp}(\mu)$ is a closed set.

Proof. If $x \notin \mathrm{supp}(\mu)$, we can find a number $r > 0$ such that $\mu(B_r(x)) = 0$. Then for each $y \in B_r(x)$ and $r' < r/10$, $B_{r'}(y) \subset B_r(x)$. This implies that $\mu(B_{r'}(y)) = 0$. Therefore $y \notin \mathrm{supp}(\mu)$. This implies that $\mathrm{supp}(\mu)^c$ is open. Therefore $\mathrm{supp}(\mu)$ is closed. $\square$

Definition 3.1.3:Alfors-David regularity

Let $k \geq 1$ be an integer. Let μ be a Borel measure on $\mathbb{R}^k$. We say that μ is AD-regular, if there is a number $s \in [0, k]$ such that for each $x \in \mathrm{supp}(\mu)$,

$$\mu(B_r(x)) \asymp r^s$$

where implicit constant in $\asymp$ is uniform across all $x \in \mathrm{supp}(\mu)$. We say that a closed set $A \subset \mathbb{R}^k$ is AD-regular if it is the support of an AD-regular measure. We will also mention s-regularity if s needs to be explicit.

Let us see some examples.

Example 3.1.4:

Let $M \subset \mathbb{R}^k$ be a smooth manifold. Then it is s-AD-regular. The AD-regular measure that supported on M is the Lebesgue measure on M and s is the topological dimension of M.

Example 3.1.5:

Let K be the middle third Cantor set. We can construct a probability measure on K as follows. First, we give the unit interval $[0,1]$ a measure one. Then each sub-interval of length $1/3$ has measure $1/2$ (there are exactly two of them). We can iterate this construction indefinitely and use Carathéodory's extension theorem to obtain a measure μ_K on K. Notice that this construction is similar to that of the Lebesgue measure on $[0,1]$. We will call μ_K to be the natural Cantor-Lebesgue measure on K. It can be also checked that $\mu_K = \mathcal{H}^{\log 2/\log 3}_{|K}$ and it is $(\log 2/\log 3)$-AD regular.

AD-regularity is a very strong condition. It gives the intuition that a measure should look homogeneous everywhere in its support and as a result, its support must be quite homogeneous. Recall that a sense of homogeneity can be captured by the equality of Assouad/lower dimensions. We now prove the following result.

Theorem 3.1.6:N.E.

Let $F \subset \mathbb{R}^k$ be a compact s-AD-regular set where $s \in [0,k]$. Then $\dim_{\mathrm{A}} F = \dim_{\mathrm{L}} F = s$.

Proof. Let μ be an AD-regular measure whose support is F. Let $0<r<R<1$. Let $x \in A$. Then the ball $B_R(x)$ has a μ measure which is $\asymp R^s$. We now cover $B_R(x)\cap F$ by using r balls. We can do so in a naive way. We first cover $B_R(x)$ into a union of r-balls in such a way that the overlapping multiplicity is bounded by some constant depending on k only. After we do this, we can then select those r-balls that intersect F and obtain a covering $\mathcal{C}$ for $B_R(x)\cap F$. Observe that, as $\mathcal{C}$ has bounded multiplicity,

$$\sum_{B_r\in\mathcal{C}} \mu(B_r) \ll \mu(B_R(x)\cap F).$$

This proves

$$\#\mathcal{C} \ll (R/r)^s.$$

The reversed inequality is true for any finite cover of $B_R(x)\cap F$. From here, we see that

$$\dim_{\mathrm{A}} F = s.$$

We omit the proof for $\dim_{\mathrm{L}} F = s$ which is similar to what we have done above. $\square$

3.1.2. Frostman exponent (the upper regularity). As we pointed out before, the AD-regularity is an extremely strong condition. In most cases, we would only be interested in non-concentration conditions. That is for a certain Borel measure μ, we want to know whether or not a large amount of mass is concentrated in a small area. In order to make sense of the previous intuition, we first check some examples.

Example 3.1.7:Dirac's δ-'function'

Consider $\mathbb{R}$. Dirac used the notion of δ-function to indicate an object δ that behaves like the zero function on $\mathbb{R}\setminus 0$. At 0, the δ function has a singularity '$\delta(0)=\infty$' with the condition that for each continuous function $f \in C(\mathbb{R})$,

$$\int f(x)\delta(x)dx = f(0).$$

Clearly, there is no function that behaves like the object δ. Nowadays, we understand that δ is in fact a functional that maps functions to numbers. In our context, we can realise δ as a Borel measure as follows,

$$\delta(A) = 1$$

if $0 \in A$ and

$$\delta(A) = 0$$

if $0 \notin A$. In this way, we see that for each continuous function f, it is indeed the case that

$$\int f(x)d\delta(x) = f(0).$$

The notational difference is that now $d\delta$ is a measure to be integrated against. To understand the rationale behind the choice of notations, observe that for each function $g \in L^1(\mathbb{R})$ with respect to the Lebesgue measure, we can define a Borel measure dg such that for each continuous function f,

$$\int f(x)dg(x) = \int f(x)g(x)d(x).$$

Thus, δ is a measure that gives all its mass to the point $\{0\}$. Intuitively, we understand that mass concentrates on extremely small areas.

Dirac's δ-measure is an extreme example. On the other side, we have the Lebesgue measure λ. For each r-ball B_r, we have

$$\lambda(B_r) \ll r^1.$$

Of course, λ is in fact 1-regular and we have $\asymp$ in the place of $\ll$. Here, the exponent 1 is the largest possible such number. To see this, we prove the following result.

Theorem 3.1.8:

Let $s>1$. Let μ be a compactly supported Borel probability measure on $\mathbb{R}$. Suppose that $\mu(B_r(x)) \ll r^s$ for μ almost all x where the implicit constant in $\ll$ is allowed to vary with x. Then μ is the zero measure.

Proof. We assume that $\operatorname{supp}(\mu) \subset [0,1]$. Next, we can use Egorov's theorem to find for each $\varepsilon > 0$, a compact set $E \subset [0,1]$ such that $\mu(E) > 1-\varepsilon$ and such that $\mu(B_r(x)) \ll r^s$ uniformly across $x \in E$. Now, we can choose $r>0$ to be any positive number and cover E with a disjoint union of r-balls with

$$\mu(B_r) \leq Cr^s.$$

To see this, we first decompose $[0,1]$ into a disjoint union of $r/2$-balls. We then select those that intersect E. For each such a $r/2$-ball $B_{r/2}$, we can find $x \in B_{r/2} \cap E$. Then we see that $B_{r/2} \subset B_r(x)$. From here the above inequality follows.

However, there are at most $\ll 1/r$ many such r-balls because they come from a decomposition of the unit interval. Thus we see that

$$\mu(E) \ll r^s r^{-1} = r^{s-1}.$$

As $s-1>0$, we see that $\mu(E)=0$. Thus $\mu([0,1]) \leq \varepsilon$. Thus we see that $\mu([0,1]) = 0$. □

Having seen two extreme examples, we now introduce the notion of Frostman exponent. It is the upper bound in the AD-regularity condition.

Definition 3.1.9:Frostman exponent

Let μ be a compactly supported Borel measure on $\mathbb{R}^k$. We say that it has a Frostman exponent s (or it is an s-Frostman measure) if

$$\mu(B_r) \ll r^s$$

holds for all r-balls where the implicit constant in $\ll$ is not required to be uniform across all $x \in \mathbb{R}^k$. If $\ll$ holds uniformly, we say that μ is uniformly s-Frostman.

Thus all s-AD regular measures have Frostman exponent s. The Frostman exponent is useful in the study of the Hausdorff dimension. We now prove the following result which extends Theorem 3.1.8.

Theorem 3.1.10:The Mass Distribution Principle

Let μ be a s-Frostman probability measure on $\mathbb{R}^k$ where $s \in [0,k]$. Then $\dim_{\mathrm{H}} \operatorname{supp}(\mu) \geq s$.

Proof. By Egorov, for each $\eta > 0$, there is a compact set $E_\eta \subset \operatorname{supp}(\mu)$ with $\mu(E_\eta) < \eta$ such that on $\operatorname{supp}(\mu) \setminus E_\eta$, the s-Frostman property holds uniformly. We restrict μ on $\operatorname{supp}(\mu) \setminus E_\eta$ and renormalise it to be a probability measure. Thus, there is no loss of generality if we assume that the s-Frostman property simply holds uniformly across all x.

Suppose that $\dim_{\mathrm{H}} \operatorname{supp}(\mu) < s$. Let $\delta, \varepsilon > 0$. We cover $\operatorname{supp}(\mu)$ with a collection $\mathcal{C}_\delta$ of balls with diameters at most δ. As $\dim_{\mathrm{H}} \operatorname{supp}(\mu) < s$, as long as δ is small enough, we can find such a cover with the property that

$$\sum_{B \in \mathcal{C}_\delta} |B|^s < \varepsilon.$$

Here we use $|B|$ for the diameter of B. Since μ is s-Frostman, we see that

$$\mu(\operatorname{supp}(\mu)) \ll \sum_{B \in \mathcal{C}_\delta} |B|^s < \varepsilon.$$

This implies that μ is the zero measure, a contradiction. □

On the other hand, it is possible to show a reversed version of the above result.

Theorem 3.1.11:Frostman's lemma

Let $E \subset \mathbb{R}^k$ be compact. Suppose that $\mathcal{H}^s(E) > 0$, then there exists a uniformly s-Frostman measure such that $\mu(E) > 0$.

Remark 3.1.12:

This result holds for all Borel sets rather than just compact sets. We will not cover this more involved result. In our special case for compact sets, we can restrict μ on E and obtain an s-Frostman measure whose support is E. Moreover, we can in fact find a μ whose Frostman property holds uniformly.

For the proof, we will develop some useful ideas that will appear repeatedly.

Definition 3.1.13:Dyadic decomposition

We can decompose $\mathbb{R}^k$ into a disjoint union of cubes of side-length 1. To be specific, the corners of those cubes are precisely $\mathbb{Z}^k$. We denote this collection of cubes as $\mathcal{D}_0$. Next, for each $D \in \mathcal{D}_0$, we can decompose it into 2^k many disjoint cubes of side-length $1/2$. We denote the collection of all the smaller cubes as $\mathcal{D}_1$. In this way, for each $l \geq 0$, we obtain a disjoint collection of cubes of side-length $1/2^l$, denoted by $\mathcal{D}_l$. Clearly, for each point $x \in \mathbb{R}^k$, we can find at least one nested list of cubes $D_0 \supset D_1 \dots$ so that

$$x = \bigcap_{i \geq 0} D_i.$$

There are points with more than one such nested list. Those points are dyadic rational points, i.e. $\mathbb{Z}^k/2^l, l \geq 0$.
Let $E \subset [0,1]^k$ be a closed set. Then we can represent E as a tree T. The root of T is $D_0 = [0,1]^k$. For each $D_1 \subset D_0$ with $D_1 \in \mathcal{D}_1$, we include it as a child of D_0 if $D_1 \cap E \neq \emptyset$. This process goes on indefinitely. As a result, we obtain an infinite tree T so that each node has at most 2^k children. For each path of T with infinite length, there is one and only one $x \in E$ that corresponds to the nested list of dyadic cubes this path represents. This association may not be injective.

Proof. We assume w.o.l.g. that $E \subset [0,1]^k$. Next, we consider the tree T that represents E. For each $l \geq 0$, we consider the space of non-negative weights $\mathcal{W}_l$ on T_l (the vertices of T on the level l) satisfies the following

- Each $\mu \in \mathcal{W}_l$ is a map from T_l to $[0,1]$ where T_l is the collection of nodes at level l (the root is at level 0).
- Each $\mu \in \mathcal{W}_l$ is s-regular up to level l in the sense that $\mu([t]) \leq |t|^s$ where $t \in T_r$ for some $s \leq l$, $|t| = (1/2)^r$, $[t]$ is the collection of all infinite paths that passes t and $[t] \cap T_l$ is the collection of offsprings of t in T_l.

We can view $\mathcal{W}_l$ as a subset of $\prod_{t \in T_l}[0,1] = [0,1]^{\#T_l}$. This subset is compact and non-empty because of the existence of the zero weight. We can order weights in a component-wise manner and we can find at least one maximal weight. Let this weight be μ_l. Notice that μ_l defines a positive measure on T_l. Our aim is to take a limit $\lim_{l\to\infty}\mu_l$ as measures (take a subsequence of $\mu_l, l \geq 0$ if necessary). Observe that the total mass is $\mu_l(T_l) = \mu_l([t_0]) \leq 1$. Due to the weak-compactness of the space of bounded measures, such a limit μ_∞ exists and satisfy

$$\mu_\infty([t]) \leq |t|^s$$

for all node $t \in T$. However, there is no guarantee that μ_∞ is not the zero measure. In order to rectify this, we need to use the condition that $\mathcal{H}^s(E) > 0$.

For each $l \geq 1$, observe that each path $t_0t_1 \ldots t_l$ in T contains at least one $0 \leq r \leq l$ with

$$\mu_l([t_r]) = |t_r|^s.$$

For otherwise, we can add $\mu(t_l)$ by some small number and obtain a strictly larger weight than μ_l. This is not possible. Thus, for each path of length $l+1$, there is a smallest r with $\mu_l([t_r]) = |t_r|^s$. Therefore there is a collection of nodes C_l in $\bigcup_{r\leq l} T_r$ such that

- For each $t, t' \in C_l$, t is not an offspring of t'.
- For each $t \in T_l$, t is an offspring of some $t' \in C_l$.

This strategy is often named the 'stopping time argument'. It is extremely useful in areas like Harmonic Analysis. Consider the dyadic cubes represented by C_l. Denote this collection of cubes by $\mathcal{C}_l$. Then $\mathcal{C}_l$ is a disjoint collection of dyadic cubes that covers E. For such a covering, we see that

$$\sum_{D\in\mathcal{C}_l} \operatorname{diam}(D)^s \asymp \sum_{t\in C_l} |t|^s = \mu_l(T_l).$$

From the condition that $\mathcal{H}^s(E) > 0$, we know that for all $\delta > 0$,

$$\mathcal{H}^s_\delta(E) = \inf_{C_\delta \text{covers } E} \left\{ \sum_{B\in C_\delta} \operatorname{diam}(B)^s \right\} > 0.$$

To see this, observe that $\mathcal{H}^s_\delta(E) > 0$ for all small enough δ, say, $0 < \delta < \delta_0 < 1$. For $\delta \geq \delta_0$, we know that to get a possibly smaller sum of form $\sum_C \operatorname{diam}(C)^s$ we need to let at least one C to be larger than δ_0. Otherwise, our sum is considered already in $\mathcal{H}^s_{\delta_0}(E)$. From here we see that in particular $\mathcal{H}^s_{100}(E) > 0$. Thus we see that

$$\mu_l(T_l) \asymp \sum_{D\in\mathcal{C}_l} \operatorname{diam}(D)^s \geq \mathcal{H}^s_{100}(E).$$

Since the above holds for all l and the rightmost term is a constant, we see that the limit measure μ_∞ is not zero.

Finally, we can push forward μ_∞ from T to $[0,1]^k$ via the natural map that assigns each infinite path (infinite nested list of dyadic cubes) in T the unique point $x \in E$ that corresponds to the infinite nested list of dyadic cubes. The compactness of E is crucial in this procedure. A more straightforward way of doing this is to first observe that each μ_l defines a measure on $[0,1]^l$ by giving weights to dyadic cubes of side $1/2^l$. Then by passing to the limit, we obtain a Borel measure on $[0,1]^k$ which is the push-forwarded measure of μ_∞ that we mentioned earlier. Let μ be this measure.

Clearly μ is a non-zero measure with $\mu([0,1]^k) \leq 1$. Next, for each dyadic cube D of an arbitrary size, we have

$$\mu(D) \leq c \cdot \operatorname{diam}(D)^s$$

for some constant $c > 0$ that depends on k only. Since any ball of radius r is contained in a dyadic cube of size at most $2r$, we obtain the s-Frostman property for μ.

Our last point to resolve is the fact that $\mu(E) > 0$. To show this, we need to use the fact that μ is a finite Borel measure on $[0,1]^k$. Then μ is regular in the sense that for each Borel set $B \subset [0,1]^k$,

$$\mu(B) = \inf\{\mu(O) : O \supset B, O \text{ open}\} = \sup\{\mu(K) : K \subset B, K \text{ compact}\}.$$

We can use the middle part of the above. Observe that if $O \supset E$ is an open set, then there is some $\delta > 0$ so that $O \supset E^\delta$ the δ-neighbourhood of E (this is because E is compact). E^δ contains the union of cubes correspond to T_l for all large enough $l \geq 0$. Thus in particular we see that

$$\mu(O) \geq \mu(E^\delta) \geq \lim_{l\to\infty} \mu_l(T_l) > \mathcal{H}^s_{100}(E).$$

This implies that

$$\mu(E) = \inf\{\mu(O) : O \supset E, O \text{ open}\} > \mathcal{H}^s_{100}(E) > 0.$$

The proof is now finished. □

Example 3.1.14:

From Theorem 3.1.10 (Mass Distribution Principle) we see that $\dim_{\mathrm{H}} K \geq \log 2/\log 3$ where K is the middle third Cantor set. On the other direction, we know that $\dim_{\mathrm{B}} K = \log 2/\log 3$ and therefore

$$\dim_{\mathrm{H}} K \leq \log 2/\log 3.$$

Thus we see that $\dim_{\mathrm{H}} K = \log 2/\log 3$. In fact, for each s-AD-regular set F, we have

$$\dim_{\mathrm{H}} F = s.$$

3.2. Supplementary material: Fourier analysis of measures (N.E.)

A significant portion of the study of fractal geometry is to understand the statistics of small-scale structures of a certain fractal set. In mathematics, there are two general tools for achieving this: Fourier Analysis and Ergodic Theory. In this section, we introduce the basics of Fourier analysis and use it to prove nice results on the geometry of fractal sets and measures.

Definition 3.2.1:Fourier transform

Let μ be a probability measure on $\mathbb{R}^k$. We define its Fourier transform $\mathcal{F}(\mu)$ to be

$$\hat{\mu}(\xi) = \int e^{-2\pi i(\xi,x)} d\mu(x)$$

where $\xi \in \mathbb{R}^k$ and $(.,.)$ is the standard Euclidean bilinear form.

Theorem 3.2.2:Basic facts of Fourier transform of measures (To be extended)

- $\hat{\mu}$ is a continuous function and $\|\hat{\mu}\|_\infty = 1$.
- $\hat{\mu}(0) = 1$.
- For each positive L^1-function (integrable), the measure $f dx$ has a Fourier transform equal to the Fourier transform of the function f.
- For two measures μ, ν the convolution $\mu * \nu$ has Fourier transform $\hat{\mu}\hat{\nu}$. Here $\mu * \nu(E) = \int \mu(E-x)d\nu(x) = \int \nu(E-x)d\mu(x)$ defines a Borel probability measure.

Example 3.2.3:

Let K be the middle third Cantor set and μ_K be the natural Cantor-Lebesgue measure. We want to compute $\hat{\mu}_K$. Observe that μ_K satisfies the following property (the self-similarity),

$$\mu_K = \frac{1}{2}T_1(\mu_K) + \frac{1}{2}T_2(\mu_K)$$

where $T_1 : x \to x/3$ and $T_2 : x \to (x+2)/3$ are maps $[0,1] \to [0,1]$ which also push-forward measures. Then we see that

$$\begin{aligned}
\hat{\mu}_K(\xi) &= \int e^{-2\pi i \xi x} d\mu_K(x) \\
&= \frac{1}{2}\int e^{-2\pi i \xi x} dT_1(\mu_K)(x) + \frac{1}{2}\int e^{-2\pi i \xi x} dT_2(\mu_K)(x) \\
&= \frac{1}{2}\int e^{-2\pi i \xi (x/3)} d\mu_K(x) + \frac{1}{2}\int e^{-2\pi i \xi (x/3+2/3)} d\mu_K(x) \\
&= \frac{1}{2}\hat{\mu}_K(\xi/3) + \frac{1}{2}\hat{\mu}_K(\xi/3)e^{-2\pi i(2\xi/3)} \\
&= \frac{1+e^{-2\pi i(2\xi/3)}}{2}\hat{\mu}_K(\xi/3) \\
&= \cdots = \prod_{j\geq 1} \frac{1+e^{-2\pi i(2\xi/3^j)}}{2}.
\end{aligned}$$

For the above limit, we have used the fact that

$$\lim_{j\to\infty} \hat{\mu}_K(\xi/3^j) = 1$$

for all $\xi \in \mathbb{R}$. We conclude that $|\hat{\mu}_K(3^l)| > c$ for some constant $c > 0$ and all $l \geq 0$.
(N.E.)This is an exceptional phenomenon. In fact, consider the following infinite product (with a in the place of 3 in μ_K),

$$\hat{\lambda}_a(\xi) = \prod_{j\geq 1} \frac{1+e^{-2\pi i(2\xi/a^j)}}{2}.$$

For now, we just treat $\hat{\lambda}_a$ as a name and do not treat $\hat{\lambda}_a$ as the Fourier transform of λ_a. It can be proved that for most of numbers a, $\hat{\lambda}_a(\xi) \to 0$ as $|\xi| \to \infty$. In fact, we know that there is a set $E \in (0,1)$ of Hausdorff dimension zero such that for all $a \in (0,1) \setminus E$, there is a constant $e_a > 0$ so that

$$|\hat{\lambda}_a(\xi)| = O(|\xi|^{-e_a}).$$

On the less probabilistic side, we know that if $|\hat{\lambda}_a(\xi)| \not\to 0$ as $|\xi| \to \infty$, then a must belong to a certain class of algebraic numbers (Pisot numbers).

3.2.1. The L^2-theory. There is a strong connection between the Hausdorff dimension and the Fourier transform. In order to explore this connection we use the following standard result in Hilbert's theory of L^2-spaces.

Theorem 3.2.4:Plancherel theorem

Let $f, g \in L^2(\mathbb{R}^k)$ w.r.t. the Lebesgue measure. Then we have

$$\int \hat{f}(\xi)\overline{\hat{g}(\xi)}d\xi = \int f(x)\overline{g(x)}dx.$$

Definition 3.2.5:Smooth approximation of a measure

Let μ be a compactly support Borel probability measure on $\mathbb{R}^k$. Let $\delta > 0$. Let

$$\mu_\delta = I_\delta * \mu$$

where I_δ is a smooth function with values in $[0,1]$ that is $= 1$ on the ball $B_\delta(0)$ and vanishes outside of $B_{2\delta}(0)$. To find such a function, we first find a smooth function I that is $= 1$ on the ball $B_1(0)$ and vanishes outside of $B_2(0)$. Then we define $I_\delta(x) = \delta^{-k} I(x/\delta)$ for $x \in \mathbb{R}^k$. Thus μ_δ is an approximation of μ up to the scale δ. Notice that μ_δ is a smooth function. In a more concrete way, for each $x \in \mathbb{R}^k$, we have

$$\mu_\delta(x) = \int I_\delta(x-y)d\mu(y) \in [\delta^{-k}\mu(B_\delta(x)), \delta^{-k}\mu(B_{2\delta}(x))].$$

For example, if μ is s-AD regular, then we see that for $x \in \mathrm{supp}(\mu)$,

$$\mu_\delta(x) \asymp \delta^{s-k}$$

and in general (uniformly across $x \in \mathbb{R}^k$),

$$\mu_\delta(x) \ll \delta^{s-k}.$$

Example 3.2.6:A first encounter of Fourier argument

Continue the context from the above definition. From Theorem 3.2.4, we see that

$$\int |\mu_\delta(x)|^2 dx = \int |\hat{\mu}_\delta(\xi)|^2 d\xi = \int |\hat{\mu}(x)|^2 |\hat{I}_\delta(x)|^2 d\xi. \tag{3.1}$$

We now study the function $\hat{I}_\delta$. Observe that

$$\hat{I}_\delta(\xi) = \hat{I}(\delta\xi).$$

As I is a compactly supported smooth function, we know that

$$\hat{I}(\xi) \to 0$$

very fast as $|\xi| \to \infty$. More precisely, for each $n \geq 1$,

$$|\hat{I}(\xi)| \ll |\xi|^{-n}.$$

This can be proved via the integration by parts. In fact, even more is true, a result of (N.E.) Paley-Wiener says that the decay should be exponential. Thus, we see that $\hat{I}$ is roughly supported on the unit ball at the origin. Therefore $\hat{I}_\delta$ is roughly supported on the ball $B_{1/\delta}(0)$ and decays fast outside of it. This correspondence between the two balls $B_\delta(0)$ and $B_{1/\delta}(0)$ via Fourier transform is a very useful fact. Intuitively, we can say that:

"The Fourier transform of $B_\delta(0)$ is roughly $B_{1/\delta}(0)$."

We can continue the chain (3.1). Observe that for each $\varepsilon > 0$,

$$\int_{|\xi|\leq 1/\delta} |\hat{\mu}(x)|^2 d\xi \ll \int |\hat{\mu}(x)|^2 |\hat{I}_\delta(x)|^2 d\xi \ll \int_{|\xi|\leq (1/\delta)^{1-\varepsilon}} |\hat{\mu}(x)|^2 d\xi.$$

For example, if μ is s-Frostman, then

$$\int |\mu_\delta(x)|^2 dx \ll \int_{\mathrm{supp}(\mu)^\delta} \delta^{2s-2k} dx \ll \lambda(\mathrm{supp}(\mu)^\delta)\delta^{2s-2k}.$$

If μ is s-AD-regular, then more is true. We have

$$\lambda(\mathrm{supp}(\mu)^\delta) \asymp \delta^{k-s}.$$

Therefore we see that if μ is s-AD-regular,

$$\int_{|\xi|\leq 1/\delta} |\hat{\mu}(x)|^2 d\xi \ll \frac{1}{\delta^{k-s}}.$$

From here, we get that the exponent $k-s$ controls the growth of the L^2 integral of $\hat{\mu}$. As will be proved later, this exponent is in some sense optimal.

3.2.2. $L^p - L^q$ duality. The L^2-theory plays a central role in the study of Fourier analysis because of Theorem 3.2.4. It is possible to extend this theorem further. However, the result is no longer as nice.

Theorem 3.2.7:Hausdorff-Young N.E.

Let $p \in (1,2]$ and $q > 0$ be such that $p^{-1} + q^{-1} = 1$. Then for each $f \in L^q$, we have

$$\left(\int |\hat{f}(x)|^q dx\right)^{1/q} = \|\hat{f}\|_q \leq \|f\|_p = \left(\int |f(x)|^p dx\right)^{1/p}.$$

For $p = 1$ we write $q = \infty$ and

$$\|\hat{f}\|_\infty \leq \|f\|_1.$$

The special case $p = 1, q = \infty$ is almost trivial to show. In fact, observe that

$$|\hat{f}(\xi)| = \left| \int e^{-2\pi i(\xi,x)} f(x)dx \right| \leq \int |f(x)|dx = \|f\|_1.$$

In particular, let f be a function such that $\hat{f} \in L^1$, then

$$\|f\|_\infty = \|\hat{\hat{f}}\|_\infty \leq \|\hat{f}\|_1.$$

3.2.3. Fourier transform and the absolute continuity of measures. In fractal geometry, a question that is often asked is whether or not some fractal measure μ is not only a measure but also a function in a certain sense.

Definition 3.2.8:The absolute continuity of measures

Let μ be a compactly supported Borel probability measure on $\mathbb{R}^k$. We say that it is absolutely continuous w.r.t. the Lebesgue measure if there is some $f \in L^1$ so that f is the density function of μ. We say that μ is L^p for $p \geq 1$ if $f \in L^p$. We say that μ is continuous, smooth, analytic if f is continuous, smooth, analytic.

The following result is useful in testing against the absolute continuity of μ.

Theorem 3.2.9:Riemann-Lebesgue N.E.

Let μ be a compactly supported Borel probability measure on $\mathbb{R}^k$. If μ is absolutely continuous, then $\hat{\mu}(\xi) \to 0$ as $\xi \to \infty$.

Corollary 3.2.10:

The middle third Cantor measure μ_K is not absolutely continuous.

The following result is useful in testing the absolute continuity of μ.

Theorem 3.2.11:

Let μ be a compactly supported Borel probability measure on $\mathbb{R}^k$. Consider its Fourier transform $\hat{\mu}$. It is a continuous function. If $\hat{\mu} \in L^2$ then μ is absolutely continuous with an L^2 density function. If $\hat{\mu} \in L^1$ then μ is continuous.

Proof. If $\hat{\mu} \in L^2$, we see that the preimage $\mathcal{F}^{-1}(\hat{\mu})$ exists as an L^2-function (Hint: Plancherel's Theorem). We denote this function as f_μ and we prove that f_μ is the density function of μ. In fact, for each $\xi \in \mathbb{R}^k$ we see that by definition,

$$\int f_\mu(x)e^{-2\pi i(\xi,x)}dx = \int e^{-2\pi i(\xi,x)}d\mu(x).$$

Since $e^{-2\pi i(\xi,x)}$ for $\xi \in \mathbb{R}^k$ is $\|.\|_\infty$ dense inside the space of continuous function on any compact set K, we conclude that for all continuous function ϕ on K,

$$\int f_\mu(x)\phi(x)dx = \int \phi(x)d\mu(x).$$

As K can be chosen arbitrarily, we see that the above holds for each compactly supported and continuous function ϕ. From here we conclude that the two measures $f_\mu dx$ and μ agree on bounded open sets. Thus, as Borel measures, they are the same. This is what we wanted to show.

If $\hat{\mu} \in L^1$, we consider μ_δ for $\delta > 0$. Since μ_δ is smooth and compactly supported, we see that $\hat{\mu}_\delta$ is also smooth and has fast decay. Therefore, we see that (Fourier inversion formula)

$$\mu_\delta(x) = \int \hat{\mu}_\delta(\xi) e^{2\pi i(\xi,x)} d\xi = \int \hat{\mu}(\xi)\hat{I}_\delta(\xi) e^{2\pi i(\xi,x)} d\xi.$$

As $\hat{\mu} \in L^1$, we can set $\delta \to 0$ and obtain

$$\lim_{\delta \to 0} \mu_\delta(x) = \lim_{\delta \to 0} \int \hat{\mu}(\xi)\hat{I}_\delta(\xi) e^{2\pi i(\xi,x)} d\xi.$$

We show that the RHS limit exists at all x and the limit is a continuous function. Indeed, because $\hat{\mu} \in L^1$, for each $\varepsilon > 0$, we can find some large number $R > 0$ such that

$$\int_{|\xi|>R} |\hat{\mu}(\xi)| d\xi < \varepsilon.$$

Observe that

$$\int \hat{\mu}(\xi)\hat{I}_\delta(\xi) e^{2\pi i(\xi,x)} d\xi = \int_{|\xi|\le R} \hat{\mu}(\xi)\hat{I}_\delta(\xi) e^{2\pi i(\xi,x)} d\xi + \int_{|\xi|>R} \hat{\mu}(\xi)\hat{I}_\delta(\xi) e^{2\pi i(\xi,x)} d\xi.$$

The last term is bounded in absolute value by ε. For the first term in the sum, observe that $\hat{I}_\delta(\xi) \to 1$ uniformly for all $|\xi| \le R$. We see that

$$\begin{aligned}
&\left| \int \hat{\mu}(\xi)\hat{I}_\delta(\xi) e^{2\pi i(\xi,x)} d\xi - \int \hat{\mu}(x) e^{2\pi i(\xi,x)} d\xi \right| \\
&\le \left| \int_{\le R} \hat{\mu}(\xi)\hat{I}_\delta(\xi) e^{2\pi i(\xi,x)} d\xi - \int_{\le R} \hat{\mu}(x) e^{2\pi i(\xi,x)} d\xi \right| \\
&+ \left| \int_{>R} \hat{\mu}(\xi)\hat{I}_\delta(\xi) e^{2\pi i(\xi,x)} d\xi - \int_{>R} \hat{\mu}(x) e^{2\pi i(\xi,x)} d\xi \right|.
\end{aligned}$$

The last term is at most 2ε and the first term in the sum tends to 0 as $\delta \to 0$. As this holds for all $\varepsilon > 0$, we conclude that

$$f_\mu(x) = \lim_{\delta \to 0} \int \hat{\mu}(\xi)\hat{I}_\delta(\xi) e^{2\pi i(\xi,x)} d\xi = \int \hat{\mu}(x) e^{2\pi i(\xi,x)} dx.$$

In the above argument, different choices of x do not make any difference therefore the limit is in fact uniform. We see that the limit f_μ is a continuous function. We have to check that f_μ is the density function of μ. To do this, we can integral against compactly supported functions as we did in the L^2 case. The crucial fact we need at this stage is that $\hat{f}_\mu = \hat{\mu}$. We omit further details. This finishes the proof. □

3.3. Energy and Hausdorff dimension

In Example 3.2.6, we illustrate the growth exponent of the L^2 integral of an AD-regular measure μ. In this section, we will apply this method to more general measures. In order to do this, we want to get finer information on a measure.

Definition 3.3.1:The s-energy integral

Let μ be a Borel measure on $\mathbb{R}^k$. Let $s \in (0, k)$. The s-energy integral of μ is

$$I_s(\mu) = \int \int \frac{d\mu(x)d\mu(y)}{|x-y|^s}.$$

In other words, we can write the double integral in the form of a convolution

$$I_s(\mu) = \int K_s * \mu(x)d\mu(x)$$

where $K_s(.) = |.|^{-s}$ is the s-Riesz kernel. Here the convolution is defined as the integral

$$K_s * \mu(x) = \int \frac{d\mu(y)}{|x-y|^s}.$$

We do not claim that the above integral is finite in a pointwise sense. However, notice that if the double integral form of $I_s(\mu)$ is finite, then the above convolution defines an $L^1(\mu)$ function. Another strategy that avoids this hassle is to use smooth approximation μ_δ for μ. This is because

$$I_s(\mu_\delta) = \int \int \frac{d\mu_\delta(x)d\mu_\delta(y)}{|x-y|^s} = \int K_s * \mu_\delta(x)d\mu_\delta(x)$$

is well defined for all $\delta > 0$. We can then use the Fourier transform and write the above integral as

$$I_s(\mu_\delta) = \int K_s \hat{*} \mu_\delta(\xi)\overline{\hat{\mu}_\delta(\xi)}d\xi.$$

For the convolution term inside the integral, we have the following general result.

Theorem 3.3.2:

Let $s \in (0, k)$ and $K_s : x \to |x|^{-s}$ as a function on $\mathbb{R}^k$. Let ϕ_1, ϕ_2 be smooth and compactly supported functions. Then

$$\int K_s * \phi_1(x)\overline{\phi_2(x)}dx = c_{k,s} \int K_{k-s}(\xi)\hat{\phi}_1(\xi)\overline{\hat{\phi}_2(\xi)}d\xi$$

for a constant $c_{k,s} > 0$.

This theorem says that as a distribution $\hat{K}_s$ is equal to K_{k-s} and as a distribution, $K_s \hat{*} \phi$ is equal to $K_{k-s}\hat{\phi}$. If K_s, K_{k-s} would be also smooth and compactly supported, then this is simply $\hat{K}_s = c_{k,s}K_{k-s}$. All the efforts we put into this theorem are to rectify the fact that K_s, K_{k-s} are not smooth and compactly supported. At this stage, it is beneficial to introduce the notion of distributions. This is not the first time we have encountered such an object. Recall Example 3.1.7.

Definition 3.3.3:Schwartz function

Consider $\mathbb{R}^k$ for some $k \geq 1$. Let $f : \mathbb{R}^k \to \mathbb{C}$ be a smooth function on $\mathbb{R}^k$. We say that f decays superpolynomially if for each $n \geq 0$,

$$|f(x)| \ll |x|^{-n}$$

where $\ll$ depends on the choice of n. Next, we define $\mathcal{S}(\mathbb{R}^k)$ (or $\mathbb{S}$ if k is clear from the context), the set of Schwartz functions in $\mathbb{R}^k$, to be

$$\mathcal{S}(\mathbb{R}^k) =$$
$$\{f \in C^\infty(\mathbb{R}^k) : \text{all partial derivatives of } f \text{ decays superpolynomially}\}.$$

The space $\mathcal{S}$ is a topological space whose topology is defined via the uniform convergence for each of the partial derivatives.

Theorem 3.3.4:N.E.

Let $f \in \mathcal{S}$. Then $\hat{f} \in \mathcal{S}$.

Definition 3.3.5:Tempered distribution

Assume the context of Definition 3.3.3. A tempered distribution L is a continuous and linear map $L : \mathcal{S}(\mathbb{R}^k) \to \mathbb{C}$. We define its Fourier transform $\hat{L}$ to be the linear map such that

$$\hat{L}(\hat{\phi}) = L(\phi)$$

for each $\phi \in \mathcal{S}$.

Remark 3.3.6:Why 'tempered'?

As each smooth and compactly supported function is a Schwartz function, we see that a tempered distribution L is also defined as a continuous and linear map $C_c^\infty \to \mathbb{C}$. The latter space is called the space of distributions. In most cases, we do not really differentiate the two notions. However, we remark that the space of distributions is strictly larger than the space of tempered distributions.

Observe that for each measurable function f, the measure $f d\mu$ is well defined. Suppose that $f d\mu$ is finite on compact sets. Then $f : \phi \to \int \phi f d\mu$ defines a tempered distribution. It may happen that $\hat{f}$ is not defined as a function but as a tempered distribution. For example, the following lemma shows that $\hat{K}_s = c_{k,s}\hat{K}_{k-s}$ as tempered distributions.

Lemma 3.3.7:

Let $s \in (0, k)$. Then for each smooth and compactly supported function ϕ,

$$\int K_s(x)\phi(x)dx = c_{k,s} \int K_{k-s}(\xi)\hat{\phi}(\xi)d\xi$$

where $c_{k,s} > 0$ is a constant that depends on k, s.

PROOF. We first show that the LHS integral is finite. Observe that

$$\begin{aligned}&\int K_s(x)\phi(x)dx\\ =&\int_{|x|<1} K_s(x)\phi(x)dx + \int_{|x|\geq 1} K_s(x)\phi(x)dx\end{aligned}$$

The second part is finite due to the compactness of the support of ϕ. For the first integral, the singular part of K_s is at $x = 0$. The result follows from the following fact ($s < k$)

$$\int_{|x|<1} \frac{1}{|x|^s} dx < \infty.$$

Similarly, the RHS also converges.

Suppose that $s > k/2$, then

$$K_s \mathbf{1}_{|x|\leq 1} \in L^1, K_s \mathbf{1}_{|x|>1} \in L^2.$$

Therefore $\hat{K}_s$ exists as a sum of $L^\infty + L^2$. Thus we can write

$$\int K_s(x)\phi(x)dx = \int \hat{K}_s(\xi)\hat{\phi}(\xi)d\xi.$$

Let $\delta > 0$ be a number. Then

$$\begin{aligned}\int K_s(x)\phi(x)dx &= \int K_s(x/\delta)\phi(x/\delta)d(x/\delta)\\ &= \delta^s \int K_s(x)\delta^{-k}\phi(x/\delta)dx\\ &= \delta^s \int \hat{K}_s(\xi)\hat{\phi}(\delta\xi)d\xi\\ &= \delta^{s-k} \int \hat{K}_s(\xi/\delta)\hat{\phi}(\xi)d\xi.\end{aligned}$$

Thus we have $\hat{K}_s(\xi) = \delta^{s-k}\hat{K}_s(\xi/\delta)$. Similarly by replacing the scaling by δ with a rotation centred at the origin, it is possible to show that $\hat{K}_s(\xi) = \hat{K}_s(|\xi|)$. From here we conclude that

$$\hat{K}_s = c_{k,s} K_{k-s}$$

for some constant $c_{k,s}$. The above equality is inside $L^\infty + L^2$. The constant $c_{k,s}$ can be determined by choosing a well-known function ϕ for which we also know $\hat{\phi}$. A handy choice is the Gaussian $\phi(x) = e^{-\pi|x|^2}$. Although it is not compactly supported, its fast decay at infinity will make all the above arguments valid. We omit further details regarding the value of $c_{k,s}$. We note that it is continuous with respect to s and $c_{k,k/2} = 1$.

Suppose that $s < k/2$, then we can apply the above argument and show that

$$\int K_{k-s}\hat{\phi} = c_{k,k-s} \int K_s \phi$$

where we used the fact that

$$\hat{\hat{\phi}}(x) = \phi(-x).$$

Thus we see that

$$\int K_s \phi = \frac{1}{c_{k,k-s}} \int K_{k-s} \hat{\phi}.$$

We proved the theorem for all $s \in (0,k) \setminus \{k/2\}$. Then case $s = k/2$ follows by a limiting argument and the fact that $c_{k,k/2} = 1$. □

PROOF OF THEOREM 3.3.2. label=default

$$\begin{aligned} c_{k,s} \int K_{k-s}(\xi)\hat{\phi}_1(\xi)\overline{\hat{\phi}_2(\xi)}d\xi &= \int K_s(x)\hat{\hat{\phi}}_1 * \overline{\hat{\hat{\phi}}_2}(x)dx \\ = \int K_s(x) \int \phi_1(-\theta - x)\overline{\phi_2(\theta)}d\theta dx \\ = \int K_s(-x) \int \phi_1(-\theta - x)\overline{\phi_2(\theta)}d\theta dx \\ = \int K_s(x)\phi_1(-\theta - x)\overline{\phi_2(\theta)}dxd\theta \\ = \int (K_s * \phi_1)(x)\overline{\phi_2(x)}dx \end{aligned}$$

□

Thus we see that

$$I_s(\mu_\delta) = c_{k,s} \int K_{k-s}(\xi)|\hat{\mu}_\delta(\xi)|^2 d\xi = c_{k,s} \int \frac{1}{|\xi|^{k-s}} |\hat{\mu}(\xi)|^2 |\hat{I}_\delta(\xi)|^2 d\xi.$$

By letting $\delta \to 0$ we have the following result.

Theorem 3.3.8:

Let μ be a Borel probability measure. Then

$$I_s(\mu) = \int |\hat{\mu}(\xi)|^2 \frac{1}{|\xi|^{k-s}} d\xi$$

as long as the RHS is finite. The RHS can be ∞. If this is the case, then $I_s(\mu) = \infty$.

We can now prove a more general version of Example 3.2.6.

Theorem 3.3.9:

Let μ be a uniformly s-Frostman measure. Then for all $t \in (0,s)$,

$$I_t(\mu) = \int\int \frac{d\mu(x)d\mu(y)}{|x-y|^t} < \infty.$$

Therefore we have

$$\int |\hat{\mu}(\xi)|^2 \frac{1}{|\xi|^{k-t}} d\xi < \infty.$$

Remark 3.3.10:

We saw in Example 3.2.6 that for each s-AD regular measure we have

$$\int_{|\xi|\leq 1/\delta} |\hat{\mu}(\xi)|^2 \ll \frac{1}{\delta^{k-s}}.$$

In particular, this implies that for each $t < s$,

$$\int |\hat{\mu}(\xi)|^2 \frac{1}{|\xi|^{k-t}} d\xi < \infty.$$

Proof. Observe that for each $y \in \mathbb{R}^k$ such that $\mu(\{y\}) = 0$,

$$\int \frac{d\mu(x)}{|x-y|^t} = \sum_{l\in\mathbb{Z}} \int_{B_{2^l}(y)\setminus B_{2^{l+1}}(y)} \frac{d\mu(x)}{|x-y|^t} \ll \sum_{l\in\mathbb{Z}} 2^{-tl}\mu(B_{2^{l+1}}(y)).$$

This sum ends for some fixed $L > 0$ for all y. We can thus write

$$\sum_{l\leq L} 2^{-tl}\mu(B_{2^{l+1}}(y)) \ll \sum_{l\leq L} 2^{-tl}2^{sl} < C$$

for a constant $C > 0$. Then we see that

$$I_t(\mu) \ll 1.$$

This is what we wanted to show. □

We can now relate the energy integral with the Hausdorff dimension.

Theorem 3.3.11:

Let E be a compact set in $\mathbb{R}^k$. Suppose that $\dim_{\mathrm{H}} E = s$. Then for each $t < s$, there is a Borel probability measure on E such that

$$I_t(\mu) < \infty.$$

Conversely, if μ is a Borel probability measure with $I_s(\mu) < \infty$, then $\dim_{\mathrm{H}} \operatorname{supp}(\mu) \geq s$.

Proof. We see that $\mathcal{H}^t(E) > 0$ for each $t < s$. Then we use Theorem 3.1.11 and find an t-Frostman measure μ with $\mu(E) > 0$. We can restrict μ on E and renormalise it to get an t-Frostman probability measure (still be written as μ) with $\mu(E) = 1$. By Theorem 3.3.9, we see that

$$I_{t'}(\mu) < \infty.$$

Since $t' < t < s$ are arbitrary the result follows.

Conversely, suppose that $I_s(\mu) < \infty$. Then for μ.a.e. y, we have

$$\int \frac{d\mu(x)}{|x-y|^s} < \infty.$$

From here, we see that for $r > 0$,

$$\mu(B_r(y)) = \int_{|x-y|\leq r} d\mu(x) \ll r^s \int \frac{d\mu(x)}{|x-y|^s} \ll r^s.$$

From an Egorov argument, for each $\varepsilon > 0$, we can find a compact set A_ε so that

$$\mu(E_\varepsilon) > 1 - \varepsilon$$

and $B_r(y) \ll r^{-s}$ uniformly for each $y \in E_\varepsilon$. Thus $\mu|_{E_\varepsilon}$ is s-Frostman and $\dim_{\mathrm{H}} E_\varepsilon \geq s$. Therefore $\dim_{\mathrm{H}} \operatorname{supp}(\mu) \geq s$. □

3.4. Fourier decay properties (N.E.)

We saw that for a Borel probability measure on $\mathbb{R}^k$, if $I_s(\mu) < \infty$ then

$$\int |\hat{\mu}(\xi)|^2 \frac{1}{|\xi|^{k-s}} d\xi < \infty.$$

In this section, we generalise this condition. To do this, we first observe that the above condition quantifies in a squared averaged sense how $|\hat{\mu}(\xi)| \to 0$ as $|\xi| \to \infty$. Various notions around this idea are called 'Fourier decay properties.' We now introduce a stronger notion of the Fourier decay property.

Definition 3.4.1:Pointwise decay

Let μ be a Borel probability measure on $\mathbb{R}^k$. We say that μ has Fourier decay if

$$\lim_{|\xi| \to \infty} \hat{\mu}(\xi) = 0.$$

We say that μ has polynomial Fourier decay if

$$\dim_F \mu := \sup\{s > 0 : |\hat{\mu}(\xi)| \ll |\xi|^{-s/2}\} > 0.$$

We name $\dim_F \mu$ as the Fourier dimension of μ.

Theorem 3.4.2:

Let μ be a Borel probability measure. If $\dim_F \mu = s > 0$, then $I_t(\mu) < \infty$ for each $t < s$.

Proof. This is clear. □

It is not hard to find measures with polynomial Fourier decay.

Theorem 3.4.3:N.E.

Let $M \subset \mathbb{R}^k$ be an analytic manifold not containing any lines, then any compactly supported and smooth measure μ on M has polynomial Fourier decay. In particular, for each analytic hypersurface M with non-vanishing Gaussian curvature, any compactly supported and smooth measure μ on M satisfies

$$|\hat{\mu}(\xi)| \ll |\xi|^{-(k-1)/2}.$$

Remark 3.4.4:

We do not prove this result here. It is a standard but deep result in Harmonic Analysis and Analytic Number Theory. The result holds for more general manifolds. See refs.

From this result of manifolds, we can now formulate the notion of Salem sets.

Definition 3.4.5:Salem set

Let $S \subset \mathbb{R}^k$ be a Borel set. We define its Fourier dimension to be

$$\dim_F S = \min\{\sup\{s > 0 : \exists \mu, \dim_F \mu \geq s\}, 1\}$$

where μ ranges over Borel probability measures such that $\mu(S) = 1$. We say that S is Salem if

$$\dim_F S = \dim_{\mathrm{H}} S.$$

Show that $\dim_F S \leq \dim_{\mathrm{H}} S$.

Example 3.4.6:Salem set

We can use Theorem 3.4.3 to see that the unit sphere in $\mathbb{R}^k$ is a Salem set for each $k \geq 2$. On the other hand, hyperplanes are not Salem as they have zero Fourier dimension. The torus

$$\mathbb{T}^2 = \partial\{B_{1/2}(a) : a = (a_x, a_y, 0), a_x^2 + a_y^2 = 1\}$$

is not Salem. The Gaussian curvature vanishes somewhere on $\mathbb{T}^2$. (N.E.) For a manifold M with a co-dimension greater than one, the study of the Fourier transform of smooth measures on M is extremely complicated. There are no satisfactory (i.e. simple) criteria telling us whether or not such a manifold is Salem.

After pointwise Fourier decay properties, we return to the averaged version.

Definition 3.4.7:Fourier l^p dimension

Let μ be a Borel probability measure on $\mathbb{R}^k$. Let $p \in (0, \infty]$, we define

$$\dim_{l^p} \mu = \sup\left\{s > 0 : \int_{|\xi|\leq R} |\hat{\mu}(\xi)|^p d\xi \ll |\xi|^{k-s}\right\}.$$

We have used l^p rather than L^p to help us remember that the l^p is taken in Fourier space (i.e. the space of ξ) rather than the physical space (i.e. the space of x).

Theorem 3.4.8:

Let μ be a Borel probability measure. If $\dim_{l^2} \mu = s > 0$, then $I_t(\mu) < \infty$ for each $t < s$.

Proof. This is clear. □

We now make sense of the idea that the squared Fourier integral in Example 3.2.6 is optimal.

Theorem 3.4.9:

Let μ be a s-AD regular measure. Then $\dim_{l^2} \mu = s$.

PROOF. In Example 3.2.6 we saw that

$$\dim_{l^2} \mu \geq s.$$

We have to show that this cannot be improved. Suppose that $\dim_{l^2} \mu > s$. Then we see that $I_{s'}(\mu) < \infty$ for some $s' > s$. Then we see that $\dim_{\mathrm{H}} \operatorname{supp}(\mu) \geq s' > s$. This contradicts with the fact that μ is s-AD regular and therefore its support has Hausdorff dimension s. ☐

3.5. Hausdorff dimension of Cartesian product sets

We start our study of the geometry of fractal sets and measures with a discussion on the Hausdorff dimension of Cartesian product sets.

Theorem 3.5.1:

Let $E, F \subset \mathbb{R}^k$ be compact sets. Then we have

$$\dim_{\mathrm{H}} E \times F \geq \dim_{\mathrm{H}} E + \dim_{\mathrm{H}} F.$$

Remark 3.5.2:

This result holds for all Borel sets.

PROOF. Let $\dim_{\mathrm{H}} E = s, \dim_{\mathrm{H}} F = t$. We can use Frostman's lemma 3.1.11 (and the remark that follows) to find Borel probability measures μ_E, μ_F that are s- and t-Frostman resp., and $\mu_E(E) = \mu_F(F) = 1$. We consider the product measure $\mu_E \times \mu_F$. It is a Borel probability measure on $\mathbb{R}^{2k}$ and $\operatorname{supp}(\mu_E \times \mu_F) = E \times F$. For each $x \in E \times F$ and $\delta > 0$, we see that

$$\mu_E \times \mu_F(B_\delta(x)) \ll \mu_E(B_\delta(x_E))\mu_F(B_\delta(x_F)) \ll \delta^{s+t}.$$

where $x = (x_E, x_F)$ is the coordinates of x. Therefore we see that $\mu_E \times \mu_F$ is an $(s+t)$-Frostman measure on $\mathbb{R}^{2k}$. From the Mass Distribution Principle (Theorem 3.1.10) we see that $\dim_{\mathrm{H}} E \times F \geq s + t$. This finishes the proof. ☐

This inequality can be strict. We provide an example. For this, we will need some results that will be proved later. Nonetheless, here is the best place for the example to appear. We make no delay.

Example 3.5.3:

Recall the sets E, F from Example 2.4.2. Observe that $E+F = \{x+y : x \in E, y \in F\} \supset (0,1)$. This is the image of $E \times F$ under the map $(x,y) \to x+y$. This map is Lipshitz. From Theorem 3.6.3 we see that $\dim_{\mathrm{H}} E \times F \geq 1$. However, $\dim_{\mathrm{H}} E = \dim_{\mathrm{H}} F = 0$.

On the other hand, for regular enough sets, we do expect that an equality.

Theorem 3.5.4:

Let $E, F \subset \mathbb{R}^k$ be compact sets. Then we have

$$\dim_{\mathrm{H}} E \times F \leq \dim_{\mathrm{H}} E + \overline{\dim}_{\mathrm{B}} F.$$

Proof. Let $\dim_{\mathrm{H}} E = s, \overline{\dim}_{\mathrm{B}} F = t$. Let $\varepsilon > 0$. Let $\delta > 0$ be a small number. If δ is small enough, we can find a $C_{\leq\delta}(E)$ covering for E with

$$\sum_{B\in C_{\leq\delta}(E)} \operatorname{diam}(B)^{s+\varepsilon} < 1.$$

Since E is compact, we can assume that $C_{\leq\delta}(E)$ is finite. For each $B \in C_{\leq\delta}(E)$ we consider the set

$$B \times F \subset E \times F.$$

We can cover $B \times F$ by first covering F with balls of size $\operatorname{diam}(B)$. The upper box dimension tells us that the number of such balls that we need is

$$\ll (\operatorname{diam}(B))^{-(t+\varepsilon/2)}.$$

We can perform this covering strategy until we obtain a covering C for $E \times F$. For this covering,

$$\sum_{B\in C} \operatorname{diam}(B)^{s+t+1.5\varepsilon} \ll \sum_{B\in C_{\leq\delta}(E)} \operatorname{diam}(B)^{s+\varepsilon}\operatorname{diam}(B)^{t+\varepsilon/2}\operatorname{diam}(B)^{-(t+\varepsilon/2)} \ll 1.$$

This implies that $\mathcal{H}^{s+t+1.5\varepsilon}(E \times F) < \infty$. Thus

$$\dim_{\mathrm{H}} E \times F \leq s + t + 1.5\varepsilon.$$

Since ε can be arbitrarily small, the result follows. □

3.6. Projections

3.6.1. Image of fractals under Lipshitz maps. Given a fractal set $F \subset \mathbb{R}^k$ and a map f from f to some metric space. It is of great interest to study the structure of the image $f(F)$. Of course, if f can be arbitrary, there is no hope to extract any information from $f(F)$. We will only study the case when f is induced from a continuous function on the ambient space $\mathbb{R}^k$.

3.6.1.1. *Warm up: Euclidean Isometry.* Consider the case when $f : \mathbb{R}^k \to \mathbb{R}^k$ is an Euclidean isometry, i.e. $f(|x - y|) = |f(x) - f(y)|$ for all pairs $x, y \in \mathbb{R}^k$.

Theorem 3.6.1:

Let $F \subset \mathbb{R}^k$ be a Borel set and f be an Euclidean isometry, then

$$\dim F = \dim f(F)$$

for dim being the Hausdorff dimension, lower box dimension, and upper box dimension.

Proof. We only prove this Theorem for the Hausdorff dimension. Consider the sets $F, f(F)$. The map f establishes an one-one correspondence of coverings of F and of $f(F)$. As f does not change the diameter of any set, we can follow the definition of Hausforff dimension and prove that $\dim_{\mathrm{H}} F = \dim_{\mathrm{H}} f(F)$. □

In fact, so much more is true. If μ is absolutely continuous w.r.t the Lebesgue measure, then the same holds for $f(\mu)$. The core idea is that for all dimensions/regularities/Fourier transforms that have been developed are all invariant under Euclidean isometries.

3.6.1.2. *General Lipshitz maps.* We now allow f to be more wild.

Definition 3.6.2:Lipshitz map

A function $f : \mathbb{R}^k \to \mathbb{R}^l$ for $k, l \geq 1$ is Lipshitz continuous if there is a constant $C > 0$ so that

$$\frac{|f(x) - f(y)|}{|x - y|} \leq C$$

for all $x, y \in \mathbb{R}^k$. We say that f is bi-Lipshitz if f^{-1} exists and it is also Lipshitz.

Theorem 3.6.3:

Let $F \subset \mathbb{R}^k$ be a Borel set and f be a Lipshitz map, then

$$\dim F \geq \dim f(F)$$

for dim being the Hausdorff dimension, lower box dimension, and upper box dimension.

Proof. We prove it for the upper/lower box dimensions. The result for the Hausdorff dimension follows similarly. For this, observe that a covering of F can be pushed to be a covering of $f(F)$ by taking the images of each set in the covering. As f is Lipshitz, it will not increase the size of the sets significantly. Thus a *delta*-covering will be sent to a $\ll \delta$-covering. We can enlarge each of the images of the sets in the covering if necessary and obtain a δ-covering of $f(F)$. In this way, we prove the result for the upper and lower box dimensions. □

3.6.2. Projections.

In this section, we introduce several special Lipshitz maps that carry geometric intuitions.

Definition 3.6.4:Linear projection

A function $L : \mathbb{R}^k \to \mathbb{R}^l$ for $k, l \geq 1$ is a linear projection if it is an affine linear map. Namely, for some non-zero $l \times k$ matrix A and a vector $b \in \mathbb{R}^l$,

$$L(x) = Ax + b.$$

The space $\ker(A)$ is called the direction of the projection L. We can then identify $\mathrm{im}(L)$ in $\mathbb{R}^k$ as the orthogonal complement of $\ker(A)$. Notice that $\dim \mathrm{im}(L) \leq k$. We can then consider $L_{K,I}$ where $K = \ker(A)$ and $I = \mathrm{im}(L)$ to be the linear map with the corresponding kernel and image. Notice that for each set $F \subset \mathbb{R}^k$, $L_{K,I}(F)$ and $L(F)$ defers by an invertible linear transformation.

Example 3.6.5:

Consider $\mathbb{R}^2$. Let $\theta \in \mathbb{S}^1 = \{(x, y) : |(x, y)| = 1\}$. Then there is a linear map L_θ that maps θ to $(0, 0)$. We can choose L_θ to be such that it is an isometry on $\theta^\perp$. We see that L_θ has the geometric meaning that projects points towards $\theta^\perp$ along the direction pointing along θ.

Definition 3.6.6:Radial projection

Consider $\mathbb{R}^k$. Let $a \in \mathbb{R}^k$. Consider the map

$$R_a : x \to \frac{x-a}{|x-a|}$$

that maps $\mathbb{R}^k \setminus \{0\}$ to the unit sphere $S^{k-1} \subset \mathbb{R}^k$. This maps is called the radial projection around a.

Remark 3.6.7:

The geometric meaning of R_a is as clear as linear projections. In fact, in everyday life, our eye system receives images from our surroundings via radial projections (at least approximately).

Linear projections and radial projections have the property that any fibre is linear, i,e. the inverse image of any point is a linear space. There are other projections with non-linear fibres. We illustrate two such projections.

Definition 3.6.8:Multiplicative projection

Consider the map $m : \mathbb{R}^k \to \mathbb{R}$ defined as

$$m((x_1, \dots, x_k)) = x_1 \dots x_k.$$

This map m is not Lipshitz. However, for each compact set $K \subset \mathbb{R}^k$, the restricted map $m|K$ is Lipshitz.

Fibres of multiplicative projections are not linear. In fact, they are hyperbolas except that the inverse image of 0 is a union of hyperplanes.

Definition 3.6.9:Distance projection

Consider the map $\Delta : \mathbb{R}^k \to \mathbb{R}$ defined as

$$\Delta((x_1, \dots, x_k)) = |(x_1, \dots, x_k)|.$$

More generally, for each $a \in \mathbb{R}^k$, we define

$$\Delta_a(x_1, \dots, x_k) = |(x_1, \dots, x_k) - a|.$$

Fibres of distance projections are spheres.

3.6.3. Marstrand's projection theorem. Linear projections and radial projections are Lipshitz. Thus under projections, dimensions of Borel cannot increase. See Theorem 3.6.3. Here a natural question to ask is how a set loses dimension after projections. First, observe that it can happen that there are strict dimension drops. For example, consider a line segment in $\mathbb{R}^2$. We can linearly project this line segment onto one-dimensional spaces. It can be easily checked that for all but one of the directions, the projected image is still a line segment and there is no dimension drop. There is exactly one direction (i.e. the direction along the line segment) for which the projected image is a single point. Then we have a dimension drop.

We shall see very soon that dimension drop is a rather exceptional phenomenon and the situation is not too far away from the line segment case we just discussed.

Theorem 3.6.10:Marstrand's projection theorem (special)

Let $F \subset \mathbb{R}^k$ be a compact set. Let H be a $k-1$ hyperplane. Let L_ω be the orthogonal projection from $\mathbb{R}^k$ onto H where $\omega \perp H$ is a unit vector (there are two choices and we write $H = H_\omega$). Then for Lebesgue almost all $\omega \in \mathbb{S}^{k-1}$,

$$\dim_{\mathrm{H}} L_\omega(F) = \max\{k-1, \dim_{\mathrm{H}} F\}.$$

Moreover, if $\dim_{\mathrm{H}} F > k-1$, then for almost all ω, $L_\omega(F)$ has a positive Lebesgue measure.

Remark 3.6.11:

We use the Fourier transform in the proof. It is also possible to prove this without using Fourier transform although the proof would be much more complicated.

Proof. Let $t < \dim_{\mathrm{H}} F$. Then there is a probability measure μ such that $I_t(\mu) < \infty$ and $\mu(F) = 1$. For each $v \in \mathbb{S}^{k-1}$, we consider the projected measure $L_v(\mu)$. It is defined to be

$$L_v(\mu)(E) = \mu(L_v^{-1}(E))$$

for each Borel $E \subset L_v(\mathbb{R}^k)$. The measure $L_v(\mu)$ is a Borel probability measure and $L_v(\mu)(L_v(F)) = 1$. We now observe that as $I_t(\mu) < \infty$,

$$\begin{aligned}\infty &> \int |\hat{\mu}(\xi)|^2 \frac{1}{|\xi|^{k-t}} d\xi \\ &= c_k \int\int |\hat{\mu}(r_\omega)|^2 \frac{1}{|r_\omega|^{k-t}} |r_\omega| dH_\omega(r_\omega) d\omega\end{aligned}$$

where $d\omega$ is the Lebesgue measure on the unit sphere $\mathbb{S}^{k-1}$, dH_ω is the Lebesgue measure on the hyperplane H_ω and $c_k > 0$ is a constant depending on k. Thus $c_k|r_\omega|dH_\omega d\omega$ is the Lebesgue measure on $\mathbb{R}^k$ under map $(r_\omega, \omega) \to r_\omega \in H_\omega \subset \mathbb{R}^k$. That is to say, for each Borel function

$$c_k \int\int f(r_\omega)|r_\omega| dH_\omega(r_\omega) d\omega = \int f(\xi) d\xi.$$

We see that for $d\omega$-a.e ω, we have

$$\int |\hat{\mu}(r_\omega)|^2 \frac{1}{|r_\omega|^{k-t}} |r_\omega| dH_\omega(r_\omega) < \infty.$$

From here we see that for such ω,

$$\int |\hat{\mu}(r_\omega)|^2 \frac{1}{|r_\omega|^{k-t-1}} dH_\omega(r_\omega) < \infty.$$

Now we consider the function $r_\omega \to \hat{\mu}(r_\omega)$. Observe that

$$L_\omega \hat{(}\mu)(\xi) = \int e^{-2\pi i(\xi,x)} dL_\omega(\mu)(x) = \int e^{-2\pi i(\xi, L_\omega(x))} d\mu(x) = \hat{\mu}(\xi_\omega)$$

where ξ_ω is the orthogonal projection of ξ on H_ω. From here we see that for each $r_\omega \in H_\omega$,

$$L_\omega\hat{(}\mu)(r_\omega) = \hat{\mu}(r_\omega).$$

Therefore the restricted function $\hat{\mu}|H_\omega$ is the Fourier transform of $L_\omega(\mu)$. Thus we see that

$$I_{k-t+1}(L_\omega\hat{(}\mu)) = \int |L_\omega\hat{(}\mu)(r_\omega)|^2 \frac{1}{|r_\omega|^{k-t-1}} dH_\omega(r_\omega) < \infty.$$

If $k-t-1>0$, then we see that $\dim_{\mathrm{H}} \operatorname{supp}(L_\omega(\mu)) \geq t$. If $k-t-1 \leq 0$, then we have an even stronger conclusion that $L_\omega(\mu)$ is absolutely continuous with an L^2-density function. In this case $\operatorname{supp}(L_\omega(\mu))$ has positive Lebesgue measure and therefore with Hausdorff dimension $k-1$. Observe that

$$\operatorname{supp}(L_\omega(\mu)) = L_\omega(F).$$

This finishes the proof. □

Marstrand's projection theorem holds for general linear projections other than the codimension one case we just considered. In order to formulate the general version, we need the notion of Grassmannian. We will not discuss further details other than its definition.

Definition 3.6.12:Grassmannian N.E.

Let $1 \leq l \leq k$ be integers. The Grassmannian $G_l(k)$ is the set of l-dimensional linear subspaces in $\mathbb{R}^k$.

Remark 3.6.13:N.E.

$G_l(k)$ is in fact a manifold of dimension $l(k-l)$. In addition, it has many other structures as an algebraic variety, a homogeneous space, etc.

Theorem 3.6.14:Marstand's projection theorem (general)

Let $F \subset \mathbb{R}^k$ be a compact set. Let $1 \leq l \leq k$ be an integer. For each $\gamma \in G_l(l)$, we use L_γ to denote the orthogonal projection onto the linear space l. Then for Lebesgue almost all $\gamma \in G_l(k)$,

$$\dim_{\mathrm{H}} L_\gamma(F) = \max\{l, \dim_{\mathrm{H}} F\}.$$

Moreover, if $\dim_{\mathrm{H}} F > l$, then for almost all γ, $L_\gamma(F)$ has a positive Lebesgue measure.

The proof of this result is very similar to the special case (Theorem 3.6.10). The most crucial ingredient is to write the Lebesgue measure on $\mathbb{R}^k$ as $c_{k,l}|r_\gamma|^{k-l}dH_\gamma(r_\gamma)d\gamma$. Such a relation is called 'a disintegration of a measure'. Common places where such a relation can be found include measure-theoretic ergodic theory (ergodic decomposition), homogeneous spaces (Weyl's integration formula), etc. We do not cover further details of this topic here.

3.6.4. non-Fourier arguments for projections. Having seen Marstrand's projection theorem, it is natural to ask for more general results that hold also for other types of projections. However, the Fourier analytic proof we introduced for linear projections depends crucially on the so-called trace formula. We did not explicitly make use of it. In our context, we used the special fact that the Fourier transform of linearly projected measure equals the restriction of the Fourier transform of the original measure on a suitable linear subspace. Such a relation reflects the fact that $\mathbb{R}^k$ is a linear space with the natural translation action $T_a : x \to x + a$ for fixed $a \in \mathbb{R}^k$. The Representation Theory of Lie Groups has much more to say about this. We do not cover further details.

We now discuss a different approach to study projections. This approach is more flexible than our Fourier analytic method. We first prove Marstrand's projection theorem again.

Theorem 3.6.15:

Let $F \subset \mathbb{R}^k$ be a compact set. Let H be a $k-1$ hyperplane. Let L_ω be the orthogonal projection from $\mathbb{R}^k$ onto H where $\omega \perp H$ is a unit vector (there are two choices and we write $H = H_\omega$). Then for Lebesgue almost all $\omega \in \mathbb{S}^{k-1}$,
$$\dim_{\mathrm{H}} L_\omega(F) = \max\{k-1, \dim_{\mathrm{H}} F\}.$$

Remark 3.6.16:

If $\dim_{\mathrm{H}} F > k-1$, we know that $L_\omega(F)$ have a positive Lebesgue measure for a typical ω. This is difficult to show without Fourier analysis.

Proof. We prove the case with $k = 2$. We reduce the situation a bit. First, instead of considering $L_\omega(F)$ we consider the following set
$$P_\theta(F) = \{x + \theta y : (x, y) \in F\}.$$
We see that $P_\theta(F)$ and $L_\omega(F)$ are scaling copies of each other with properly chosen $\theta \in (0, \infty)$ and $\omega \in \mathbb{S}^1$. We can also restrict ourselves to $\theta \in [1, 2]$. This will not cause any loss of generality.

Since $\dim_{\mathrm{H}} F = s$, we can find a uniformly s^--Frostman measure on F for any $s^- < s$. Suppose that $s < 1$. We need to estimate
$$I_{s^-}(\mu_\theta)$$
where μ_θ is the pushed-forward measure of μ under the map $(x, y) \to x + \theta y$. Consider the integral
$$\int_{[1,2]} I_{s^-}(\mu_\theta) d\theta.$$
For this integral to make sense we need $\theta \to I_{s^-}(\mu_\theta)$ to be Borel measurable. This fact is simple to establish and we omit its proof. Observe that
$$\begin{aligned}\int I_{s^-}(\mu_\theta) d\theta &= \int\int\int \frac{d\mu_\theta(x) d\mu_\theta(y)}{|x-y|^{s^-}} d\theta \\ &= \int\int\int \frac{d\mu(a) d\mu(b) d\theta}{|(a_x + \theta a_y) - (b_x + \theta b_y)|^{s^-}}.\end{aligned}$$

For the last line, we used the fact that for each Borel function $f : \mathbb{R}^2 \to \mathbb{R}$,

$$\int\int d\mu_\theta(x)d\mu_\theta(y)f(x,y) = \int\int d\mu(a)d\mu(b)f(a_x + \theta a_y, b_x + \theta b_y).$$

This can be viewed as the definition of μ_θ. We can now use Fubini's theorem to see that

$$\begin{aligned}&\int\int\int \frac{d\mu(a)d\mu(b)d\theta}{|(a_x+\theta a_y)-(b_x+\theta b_y)|^{s^-}}\\ &= \int\int d\mu(a)d\mu(b)\int \frac{d\theta}{|(a_x+\theta a_y)-(b_x+\theta b_y)|^{s^-}}\\ &\ll \int\int d\mu(a)d\mu(b)\sum_{k\in\mathbb{Z}} 2^{-ks^-}|A_k|\end{aligned}$$

where $|A_k|$ is the Lebesgue measure of the set

$$A_k = \{\theta : |(a_x+\theta a_y)-(b_x+\theta b_y)| \in [2^k, 2^{k+1})\}.$$

Consider the function

$$g(\theta) = (a_x+\theta a_y)-(b_x+\theta b_y).$$

We have $g'(\theta) = (a_y - b_y)$. Thus we see that

$$|A_k| \ll \frac{2^k}{|a-b|}.$$

From here we have

$$\sum_{k\in\mathbb{Z}} 2^{ks^-}|A_k| \ll \frac{1}{|a-b|}\sum_k 2^{(1-s^-)k}.$$

Since $s < 1$, we see that the sum over k is $\ll |a-b|^{1-s^-}$. We now have

$$\int I_{s^-}(\mu_\theta)d\theta \ll \int\int \frac{d\mu(a)d\mu(b)}{|a-b|^{s^-}} < \infty$$

because μ is s^--Frostman. From here we see that for Lebesgue almost all θ,

$$I_{s^-}(\mu_\theta) < \infty.$$

This implies that $\dim_{\mathrm{H}} P_\theta(F) \geq s^-$.

Now, if $s \geq 1$, we can repeat the above argument with s^- being replaced by 1^-, a number smaller than one. We then obtain the result that for Lebesgue almost all θ,

$$I_{1^-}(\mu_\theta) < \infty.$$

We have now proved that for Lebesgue almost all θ,

$$\dim_{\mathrm{H}} P_\theta(F) \geq \min\{1, \dim_{\mathrm{H}} F\}.$$

The result follows because the above inequality also holds with $\geq$ being replaced with $\leq$. □

Recall Definitions 3.6.6 and 3.6.9. For $F \subset \mathbb{R}^k$, we introduced families $R_a(F), \Delta_a(F)$ with parameter $a \in \mathbb{R}^k$. From Marstrand's projection theorem we can intuitively think that dimension drop rarely happens. We will now prove the following result.

Theorem 3.6.17:

Let $F \subset \mathbb{R}^2$ be compact. Suppose that $\dim_{\mathrm{H}} F > 1$. Then for Lebesgue a.e. $a \in \mathbb{R}^2$,
$$\Delta_a(F), R_a(F)$$
has a positive Lebesgue measure. More generally, for Lebesgue almost all a, we have
$$\dim_{\mathrm{H}} \Delta_a(F), \dim_{\mathrm{H}} R_a(F) = \min\{1, \dim_{\mathrm{H}} F\}.$$

We will assume that F has zero Lebesgue measure otherwise the result holds. In this case, we only consider $a \notin F$. We show that for Lebesgue almost all such choices the theorem holds. Since F is compact, we will further localise our problem by considering $a \in B$ where B is a ball which is disjoint with F. Then $d(F, B) > 0$. Without loss of generality, we assume that $d(B, F) > 1$. We will need the following lemmas that capture the geometry that we need for proving Theorem 3.6.17.

Lemma 3.6.18:two circles lemma

Let $1/2 < r_1 \leq r_2 < N$ where $N > 0$ is a positive number. Let $\delta \in (0, 1/10)$. Let $x, y \in \mathbb{R}^2$ such that $d(x, y) < 1/2$. Consider the circles $C_1 = C_{r_1}(x), C_2 = C_{r_2}(y)$ centred at x, y resp. Then consider the δ-neighbourhoods C_1^δ, C_2^δ. Then we have
$$\lambda(C_1^\delta \cap C_2^\delta) \ll \frac{\delta^2}{|d(x, y) - |r_1 - r_2||},$$
where the implicit constant depends on N.

Proof. $C_1^\delta \cap C_2^\delta$ has large intersection when $d(x, y)$ is close to $r_1 + r_2$ or $|r_1 - r_2|$. Our conditions say that $r_1 + r_2$ is much larger than $d(x, y)$. Therefore, we only need to consider the second case. The worst case scenario is when $r_1 = r_2$ and $x = y$. In this case $C_1^\delta \cap C_2^\delta$ is seen to have Lebesgue measure as large as $\asymp \delta$. If $d(x, y) \neq |r_1 - r_2|$ then the two circles C_1, C_2 do not share tangent lines. In other words, the tangent lines of C_1, C_2 are their intersection are not the same. The angle formed by those tangent lines is $\gg |d(x - y) - |r_1 - r_2||$. Therefore we see that
$$\lambda(C_1^\delta \cap C_2^\delta) \ll \delta \times \frac{\delta}{|d(x - y) - |r_1 - r_2||}.$$
We use this inequality when $|d(x - y) - |r_1 - r_2|| \geq \delta$ and we use trivial upper bound δ otherwise. This proves this lemma. $\square$

Lemma 3.6.19:two lines lemma

Let $x, y \in \mathbb{R}^2$. Let $a \in \mathbb{R}^2$ be such that a is far away from the line segment xy. More precisely, we assume that $N > d(a, xy) > 1/2$ for some number $N > 0$. Let $\delta \in (0, 1/10)$. Consider the two lines l_{ax}, l_{ay}. Consider the δ-neighbourhoods $l_{ax}^\delta, l_{ay}^\delta$. Then we have
$$\lambda(l_{ax}^\delta \cap l_{ay}^\delta) \ll \frac{\delta^2}{\angle xay}.$$
The implicit constant depends on N.

PROOF. The conclusion that

$$\lambda(l^\delta_{ax}\cap l^\delta_{ay})\ll\frac{\delta^2}{\angle xay}$$

follows from what we did in the proof of Theorem 2.4.5. □

PROOF OF THEOREM 3.6.17. We only prove the conclusion for R_a. The conclusion for Δ_a follows via similar lines. We only consider the case when $\dim_{\mathrm{H}} F\leq 1$. The conclusion for the rest cases follows similarly.

Let $s=\dim_{\mathrm{H}} F\leq 1$. Let μ be an s-Frostman measure supported on F. Let $\delta>0$. We consider the δ-scaled smooth approximation μ_δ. For some $t\in(0,s)$. We estimate $I_t(R_a(\mu_\delta))$ for $a\in B$ where B is a ball such that $d(B,F)>1$. We can further assume that the diameter of $B\cup F$ is at most $N>0$. The strategy is to estimate the integral

$$\int I_t(R_a(\mu_\delta))da.$$

As $I_t(R_a(\mu_\delta))>0$ for all $a\in B$ and $I_t(R_a(\mu_\delta))\to I_t(R_a(\mu))$ as $\delta\to 0$. The limit can be of course ∞. We know that

$$I_t(R_a(\mu))=\limsup_{\delta\to 0} I_t(R_a(\mu_\delta)).$$

Then from the monotone convergence theorem, we see that

$$\int I_t(R_a(\mu))da=\lim_{\delta\to 0}\int I_t(R_a(\mu_\delta))da.$$

Thus if we can show that as $\delta\to 0$

$$\int I_t(R_a(\mu_\delta))da\ll 1,$$

then we see that

$$\int I_t(R_a(\mu))da<\infty.$$

This will show that $I_t(R_a(\mu))<\infty$ for Lebesgue almost all $a\in B$. Therefore $\dim_{\mathrm{H}}\operatorname{supp}(R_a(\mu))\geq t$. As $t<s$ can be arbitrarily chosen, we see that $\dim_{\mathrm{H}}\operatorname{supp}(R_a(\mu))\geq s$. This proves the theorem.

Now we estimate

$$\begin{aligned}&\int I_t(R_a(\mu_\delta))da\\ =&\int\int\int\frac{dR_a(\mu_\delta)(x)dR_a(\mu_\delta)(y)}{|x-y|^t}da\\ \asymp&\int\int\int\frac{1}{\delta^2}\frac{\mu(l^\delta_{a,x})\mu(l^\delta_{a,y})}{|x-y|^t}dxdyda\end{aligned}$$

Here, $l_{a,x}$ is the line starting from a with direction $x\in S^1$ (the target set of R_a). We can continue the above computation

$$\begin{aligned}&\int\int\int\frac{1}{\delta^2}\frac{\mu(l^\delta_{a,x})\mu(l^\delta_{a,y})}{|x-y|^t}dxdyda\\ =&\int\int\int\int\int\frac{1}{\delta^2}\frac{\mathbf{1}_{l^\delta_{a,x}}(g)\mathbf{1}_{l^\delta_{a,y}}(h)}{|x-y|^t}dxdydad\mu(h)d\mu(g)\end{aligned}$$

We know that a is far away from the line segment gh. We can integrate the variables x, y. As long as the angle $\angle gac > \delta$, we have

$$\int\int \frac{1}{\delta^2} \frac{\mathbf{1}_{l^\delta_{a,x}}(g)\mathbf{1}_{l^\delta_{a,y}}(h)}{|x-y|^t} dxdy \ll \frac{1}{\delta^2} \frac{\delta}{d(a,g)} \frac{\delta}{d(a,h)} (1/\angle gah)^t.$$

If $\angle gac < \delta$, this integral is

$$\ll \frac{1}{\delta^2} \int_{|x|,|y|<\delta} \frac{1}{|x-y|^t} dxdy \ll \delta^{-t}.$$

We thus see that

$$\int\int \frac{1}{\delta^2} \frac{\mathbf{1}_{l^\delta_{a,x}}(g)\mathbf{1}_{l^\delta_{a,y}}(h)}{|x-y|^t} dxdy \ll \frac{1}{\max\{\angle gah, \delta\}^t}.$$

As $a \in B$ and B is far away from F, we have as long as $t < 1$,

$$\int \frac{1}{\max\{\angle gah, \delta\}^t} da \ll \frac{1}{|g-h|^t}.$$

To see this, without loss of generality, we assume that $l_{gh} \cap B \neq \emptyset$. Then we have for $d < 0.1$.

$$\angle gah \gg d(a, l_{gh})|g-h|.$$

For $d \geq 0.1$ we have

$$\angle gah \gg |g-h|.$$

Then we see that

$$\int \frac{1}{\max\{\angle gah, \delta\}^t} da \ll \int \frac{|g-h|^{-t}}{d(a, l_{gh})^t} da = |g-h|^{-t} \int \frac{1}{d(a, l_{gh})^t} da.$$

The last integral is finite because it is $\asymp$ to the one dimensional integral $\int_0^1 1/r^t dr$. From here we have

$$\int\int\int\int\int \frac{1}{\delta^2} \frac{\mathbf{1}_{l^\delta_{a,x}}(g)\mathbf{1}_{l^\delta_{a,y}}(h)}{|x-y|^t} dxdydad\mu(h)d\mu(g)$$
$$\ll \int\int \frac{d\mu(g)d\mu(h)}{|g-h|^t} = I_t(\mu) < \infty.$$

For the last line, we used the fact that μ is s-Frostman and for $t < s$, we have $I_t(\mu) < \infty$. □

3.7. Intersections

We have studied a certain projection problem for fractal sets. Now we study the 'dual problem'. To have some ideas, consider a projection problem for a finite set $E \subset \mathbb{Z}^2$. We want to consider the set

$$P(E) = \{x + y : (x, y) \in E\}.$$

For each $z \in \mathbb{Z}$, we have the fibre $P^{-1}(z) = \{(x, y) \in \mathbb{Z}^2 : x + y = z\}$. Let $\#_z(E)$ be the cardinality of $P^{-1}(z) \cap E$. We have the following relation

$$\#E = \sum_{z \in P(E)} \#_z(E).$$

Thus consider $\#E$ as being fixed (as a large number like 1 million). If $\#P(E)$ is large, then $\#_z(E)$ is not too large for most of $z \in P(E)$. Conversely, if all $\#_z(E)$ is not large, then $\#P(E)$ is forced to be large.[1] The study of $P(E)$ is similar to our projection problem and the study $\#_z(E)$ is then considered to be an intersection problem. We now want to transfer this study from finite sets to fractal sets.

We first discuss some intuitive ideas. Let $F \subset \mathbb{R}^2$ be a compact fractal. For each $x \in \mathbb{R}$, we use F_x to denote the x-slice of F

$$F_x = \{(x, y) \in F\}.$$

Suppose that $\dim_{\mathrm{H}} F = s$, then from the discrete consideration above, we have the idea that an average F_x should have $\dim_{\mathrm{H}} F = \max\{0, s-1\}$. This intuition looks solid. However, we remark that it is not true in general. For example, for any function over $[0,1]$ with a graph F with Hausdorff dimension bigger than one, we see that F_x is a single point for all x and will always have dimension zero. It is not hard to find such graphs. In fact, the Takagi graphs have this property. We do not prove this fact here. Another example of such a graph is a typical graph of the one-dimensional Brownian motion. We do not prove this fact either.

We prove the following result.

Theorem 3.7.1:

Let $E \subset \mathbb{R}^{k+l}$ be a compact set. Suppose that $\dim_{\mathrm{H}} E = s > 0$. Then for Lebesgue almost all $x \in \mathbb{R}^k$, the slice

$$E_x = \{(x, y) \in E\} \subset \mathbb{R}^{k+l}$$

satisfies $\dim_{\mathrm{H}} E_x \le \max\{0, s-k\}$.

Remark 3.7.2:

This result holds for slices with different orientations.

PROOF. Suppose the opposite. Let $L \subset \mathbb{R}^k$ be such that $x \in L$ implies that

$$\dim_{\mathrm{H}} E_x > \max\{0, s-k\} = \rho.$$

Then it is possible to find a ρ-Frostman measure μ_x on E_x. Without loss of generality, we can assume that those ρ-Frostman properties hold uniformly across $x \in L$. We can then define the measure

$$\mu = \int_L \mu_x dx$$

via the following relation that for each Borel $A \subset \mathbb{R}^{k+l}$

$$\mu(A) = \int_L \mu_x(A) dx.$$

Notice that $\mu_x(A) = \mu_x(A \cap E_x)$. Observe that $\mu(E) = 1$. Therefore μ is a Borel probability measure on E. For each r-ball, B_r, say, we have

$$\mu(B_r) = \int_L \mu_x(B_r) dx = \int_L \mu_x(B_r \cap E_x) dx \ll r^\rho \int_L \mathbf{1}_{x: B_r \cap E_x \neq \emptyset}(x) dx \ll r^{\rho+k}.$$

[1] If you like this problem for finite sets. Then perhaps you can have a look at the mathematical subject: Additive Combinatorics.

Thus μ is a $\rho + k$-Frostman measure. Since $\mu(E) = 1$, we see from the Mass Distribution Principle that $\dim_{\mathrm{H}} E \geq \rho + k > \max\{k, s\}$. This is not possible because $\dim_{\mathrm{H}} E = s$. □

It is very tempted to believe that $\dim_{\mathrm{H}} E_x \leq \max\{0, s-1\}$ for all x. This is not the case. A trivial example would be a line segment that is perpendicular to the x-axis.

Next, we can generalise this intersection result.

Theorem 3.7.3:

Let E, F be compact sets in $\mathbb{R}^k$. Let $\dim_{\mathrm{H}} E = s, \dim_{\mathrm{H}} F = t$. Then for Lebesgue almost all $a \in \mathbb{R}^k$,

$$\dim_{\mathrm{H}} E \cap (F + a) \leq \max\{0, \dim_{\mathrm{H}} E \times F - k\}.$$

Remark 3.7.4:

If $\dim_{\mathrm{H}} E = \underline{\dim}_{\mathrm{B}} E$ or $\dim_{\mathrm{H}} F = \underline{\dim}_{\mathrm{B}} F$ then

$$\dim_{\mathrm{H}} E \times F = s + t$$

and we have for almost all $a \in \mathbb{R}^k$,

$$\dim_{\mathrm{H}} E \cap (F + a) \leq \max\{0, s + t - k\}.$$

Proof. Consider $E \times F \subset \mathbb{R}^k \times \mathbb{R}^k$. The diagonal $\Delta = \{(x, x)\}_{x \in \mathbb{R}^k}$ has the property that

$$\Delta \cap E \times F$$

is $E \cap E$ after some non-degenerate (invertible) linear map. More generally, for each $a \in \mathbb{R}^k$, $\Delta_a = \{(x, x + a)\}_{x \in \mathbb{R}^k}$ has the property that

$$\Delta \cap E \times F$$

is basically $E \cap (F + a)$. Thus the result follows from the Marstrand slicing theorem. □

CHAPTER 4

Linear IFS

4.1. IFSs and their attractors

An important class of fractal sets can be constructed via iterated function systems.[1] Let $f_1, \ldots, f_n$ be functions $\mathbb{R}^k \to \mathbb{R}^k$. We write $\Lambda = \{f_1, \ldots, f_n\}$. Let $K \subset \mathbb{R}^k$ be a compact set. We call K to be an attractor of Λ if

$$K = \bigcup_{i=1}^{n} f_i(K).$$

We need to have a convenient sufficient condition to test the existence of at least one such K. This is due to Hutchinson.

Theorem 4.1.1:Hutchinson's theorem

Let $\Lambda = \{f_1, \ldots, f_n\}$ be contracting. Namely, they are Lipshitz with constants all strictly smaller than 1. Namely, $|f_i(x) - f_i(y)| < \rho|x - y|$ for all i, x, y s.t. $x \neq y$. Then there is a unique attractor K for Λ. We write it as K_Λ.

To prove this result, we need to notion of Hausdorff distance.

Definition 4.1.2:Hausdorff distance

Let $X, Y \subset \mathbb{R}^k$ be compact subsets. We define

$$d_H(X, Y) = \inf\{\delta > 0 : X \in Y^\delta, Y \in X^\delta\}.$$

Thus, if $d_H(X, Y) = \delta > 0$, then for each $x \in X$, the ball $B_\delta(x) \cap Y \neq \emptyset$.

Theorem 4.1.3:

Hausdorff distance defines a metric on the space of compact subsets of $\mathbb{R}^k$.

Proof. First observe that if $d_H(X, Y) = 0$, then $X = Y$. Next, it is obvious that $d_H(X, Y) \geq 0$ for all compact X, Y. We now check the triangle inequality, i.e.

$$d_H(X, Y) + d_H(Y, Z) \geq d_H(X, Y)$$

for all compact X, Y, Z. Let $\varepsilon > 0$. Observe that for each $x \in X$, it is possible to find $y \in Y$ so that $|x - y| < d_H(X, Y) + \varepsilon$. For this $y \in Y$, it is possible to find a $Z \in Z$ with

$$|y - z| < d_H(Y, Z) + \varepsilon.$$

[1] In fact, in some daily contexts, fractals and IFSs are equivalent.

Thus we see that

$$|x - z| \leq |x - y| + |y - z| < d_H(X, Y) + d_H(Y, Z) + 2\varepsilon.$$

From here we conclude that

$$d_H(X, Z) \leq d_H(X, Y) + d_H(Y, Z) + 2\varepsilon.$$

This proves the triangle inequality as ε is arbitrary. $\square$

Theorem 4.1.4:

The metric space $(\mathcal{K}, d_H)$ is complete where $\mathcal{K}$ is the space of non-empty compact sets in $\mathbb{R}^k$.

Proof. Let $K_i, i \geq 1$ be a Cauchy sequence. We need to identify a compact set K so that

$$\lim_{i\to\infty} d_H(K_i, K) = 0.$$

For each $\delta > 0$, there is an $N > 0$, such that for each $i, j \geq N$,

$$d_H(K_i, K_j) < \delta.$$

We can choose $\delta_l = 1/2^l$ for $l \geq 1$ and we choose N_l accordingly. Consider the sequence $F_l = \overline{K_{N_l}^{2\delta_l}}$ for $l \geq 1$. Then F_l is a decreasing sequence of compact sets. In particular,

$$K = \bigcap_{l\geq 1} F_l \neq \emptyset.$$

We already see that $d_H(K_{N_l}, K) \to 0$ as $l \to \infty$. Use the Cauchy property, we can upgrade this to

$$\lim_{i\to\infty} d_H(K_i, K) = 0.$$

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

Proof of Theorem 4.1.5. Let K be a compact set. Consider $\Lambda(K) = \cup_{i=1}^n f_i(K)$ and $\Lambda^2(K) = \Lambda(\Lambda(K))$. Let $\delta = d_H(K, \Lambda(K))$. Let $\varepsilon > 0$. Then for each $x \in K$, there is some f_i and some $y \in K$ so that $d(x, f_i(y)) < \delta + \varepsilon$. Let $z \in \Lambda(K)$ be arbitrary. There is some $x' \in K$ and j so that $z = f_j(x')$. For this x', we can find y' and i' so that

$$d(x', f_{i'}(y')) < \delta + \varepsilon.$$

We can then see that

$$d(z = f_j(x'), f_j \circ f_{i'}(y')) < \rho d(x', f_{i'}(y')) < \rho(\delta + \varepsilon).$$

For small enough ε, we conclude that

$$d_H(\Lambda(K), \Lambda^2(K)) < \rho^+ d_H(K, \Lambda(K))$$

where $\rho^+ < 1$. This implies that the sequence of compact sets $\Lambda^l(K), l \geq 1$ is a Cauchy sequence. There is a limit K_Λ. This limit satisfies

$$d_H(K_\Lambda, \Lambda(K_\Lambda)) = 0.$$

This proves the existence.

For the uniqueness, consider the sequence $\Lambda^l(\{x\})$ for each singleton $\{x\}$. This sequence converges to some attractor K_x. For a different singleton $\{y\}$, we have a possibly different attractor K_y. The argument above can provide us with the following inequality,

$$d_H(\Lambda^l(x), \Lambda^l(y)) < (\rho^+)^l d(x, y).$$

Therefore we see that $K_x = K_y$. From here the uniqueness follows. □

There is a version of Hutchinson's theorem that also holds in the space of Borel measures. In fact, it is possible to define a Hausdorff distance in the space of Borel probability measures and perform the argument as in the proof of Hutchinson's theorem. We omit the proof but only show its statement.

Theorem 4.1.5:Hutchinson's theorem for measures

Let $\Lambda = \{f_1, \ldots, f_n\}$ be contracting. Namely, they are Lipshitz with constants all strictly smaller than 1. Namely, $|f_i(x) - f_i(y)| < \rho|x - y|$ for all i, x, y s.t. $x \neq y$.
Let $p_1, \ldots, p_n$ be non-negative numbers with sum one. Then there is a unique Borel probability measure μ for Λ. We write it as μ_Λ. In fact, $\mathrm{supp}(\mu_\Lambda) \subset K_\Lambda$ and the inclusion is equal if $p_1, \ldots, p_n > 0$. The measure μ_Λ satisfy the following property

$$\mu_\Lambda = \sum_{i=1}^{n} p_i f_i(\mu_\Lambda).$$

4.2. Self-similar sets/measures

For a system Λ with similarity maps (self-similar system), i.e. Euclidean isometries composed with scaling, we call Λ to be a self-similar system and K_Λ, μ_Λ are called to be a self-similar set and a self-similar measure. By Hutchinson's theorem, if all the scaling ratios are strictly smaller than one, we have the existence and uniqueness of the attractor.

We can now rediscover some of the fractal examples we saw earlier.

Example 4.2.1:The middle-third Cantor set/measure

Let $\Lambda = \{f_1(x) = x/3, f_2(x) = (x + 2)/3\}$. Then K_Λ is the middle-third Cantor set. If we give the two maps the probability weight $(1/2, 1/2)$ then μ_Λ is the AD-regular measure on K_Λ.

4.2.1. Dimension theory for self-similar sets. Given a self-similar system Λ, we consider the attractor K_Λ. How can we compute $\dim K_\Lambda$? It is in general an extremely difficult problem. Luckily, the following result simplifies the situation a bit.

Theorem 4.2.2:Falconer's implicit theorem

Given a self-similar system Λ, its attractor K_Λ satisfies

$$\dim_{\mathrm{H}} K_\Lambda = \dim_{\mathrm{B}} K_\Lambda.$$

We will prove this result after some preparations. First, we introduce the notion of the self-similarity dimension. It serves as a natural upper bound for the dimensions of self-similar sets.

Definition 4.2.3:Self-similarity dimension

Let Λ be a self-similar system. Let $\{r_1, \dots, r_n\}$ be its scaling ratios (counting multiplicities). They are all smaller than one. The self-similarity dimension is defined as

$$\dim_s K_\Lambda = t$$

where t is a unique positive solution to the equation

$$\sum_i r_i^t = 1.$$

To see why a solution exists, observe that

$$\lambda : t \to \sum_i r_i^t$$

satisfies

$$\lambda(0) = n \geq 1$$

and it is monotonically decreasing. For example, for the middle-third Cantor system Λ, we have $r_1 = r_2 = 1/3$ then we solve

$$2(1/3)^t = 1$$

to get

$$\dim_s K_\Lambda = \frac{\log 2}{\log 3}.$$

Theorem 4.2.4:

Let $\Lambda = \{f_1, \dots, f_n\}$ be a self-similar system in $\mathbb{R}^k$. Then

$$\overline{\dim}_{\mathrm{B}} K_\Lambda \leq \dim_s K_\Lambda.$$

Remark 4.2.5:Exact overlaps

It is possible to have a gap, i.e.

$$\dim_{\mathrm{H}} K_\Lambda \leq \overline{\dim}_{\mathrm{B}} K_\Lambda < \dim_s K_\Lambda.$$

This is the case if any of f_i and f_j are the same. More generally, this also happens if there are two maps from further iterations of Λ are the same. In such a situation, we say that Λ has exact overlaps. It is largely believed that $\dim_{\mathrm{H}} K_\Lambda < \dim_s K_\Lambda$ if and only if Λ has exact overlaps. This is a difficult conjecture.

Proof. We consider the case when $k = 1$ in $\mathbb{R}^k$. Thus we are considering self-similar sets in $\mathbb{R}$. For higher dimensions, the proof is similar. Let $C(\Lambda)$ be the convex hull of K_Λ. It is the smallest closed interval that contains K_Λ.

We can now assign a probability weight $(p_1, \ldots, p_n)$ to $\{f_1, \ldots, f_n\}$ so that it satisfy

$$p_i = r_i^t$$

where $t = \dim_s K_\Lambda$.

Let $\delta > 0$. We want to study K_Λ at scale δ. Let $\omega \in \{1, \ldots, n\}^{\mathbb{N}}$ be a sequence in $\{1, \ldots, n\}$. We then define iterated functions for integers $j \geq i \geq 1$,

$$f_{\omega_i^j} = f_{\omega_j} \circ \cdots \circ f_{\omega_i}.$$

The scaling ratio $r_{\omega_i^j}$ for $f_{\omega_i^j}$ is equal to $\prod_{l=i}^{j} r_{\omega_l}$. We can find the smallest number l such that

$$r_{\omega_1^l} \leq \delta.$$

The existence of l is because $r_1, \ldots, r_n$ are smaller than one. Since this argument holds for all $\omega \in \{1, \ldots, n\}^{\mathbb{N}}$, we can find a covering $\mathcal{C}$ for K_Λ with all such

$$f_{\omega_1^l}(C(\Lambda)).$$

Observe that each $B \in \mathcal{C}$ as diameter at most δ and at least $\min\{r_1, \ldots, r_n\}\delta$. Each $B \in \mathcal{C}$ is associated with the probability weight

$$\prod_{j=1}^{l} p_{\omega_j} = \prod_{j=1}^{l} r_{\omega_j}^t \asymp \operatorname{diam}(B)^t.$$

We have

$$\#\mathcal{C} \asymp \frac{1}{\operatorname{diam}(B)^t} \asymp \frac{1}{\delta^t}.$$

We can enlarge each $B \in \mathcal{C}$ so that they all have size δ. Not all $B \in \mathcal{C}$ are disjoint, we nonetheless have

$$N_\delta(K_\lambda) \ll \frac{1}{\delta^t}.$$

This proves that $\overline{\dim}_{\mathrm{B}} K_\Lambda \leq \dim_s K_\Lambda$. $\square$

From the proof, we see that it is almost the case that

$$\overline{\dim}_{\mathrm{B}} K_\Lambda = \dim_s K_\Lambda.$$

If the covering $\mathcal{C}$ would be disjoint, then we will have the above equality. Some conditions lead us to this conclusion.

Definition 4.2.6:Separation Conditions

Let Λ be a self-similar system. We say that Λ or K_Λ has the Open Set Condition (OSC) if there is an open set O such that for different $f, g \in \Lambda$

$$f(O) \cap g(O) = \emptyset$$

and $O \supset \cup_{f \in \Lambda} f(O) = \Lambda(O)$.

We say that Λ or K_Λ has the Strong Separation Condition (SSC) if for different $f, g \in \Lambda$,

$$f(K_\Lambda) \cap g(K_\Lambda) = \emptyset.$$

The SSC is stronger than the OSC. To see that, we consider a K_Λ with the SSC. Suppose that it does not have the OSC. Then for each open set O with $\Lambda(O) \subset O$, there is some pair of functions $f, g \in \Lambda$ so that

$$f(O) \cap g(O) \neq \emptyset.$$

We can apply this conclusion to K_Λ^δ as $\delta \to 0$. Since Λ is finite, there is at least one pair f, g so that

$$f(K_\Lambda^{\delta_i}) \cap g(K_\Lambda^{\delta_i}) \neq \emptyset$$

for some $\delta_i \to 0$. Since K_Λ is compact, we see that

$$f(K_\Lambda) \cap g(K_\Lambda) \neq \emptyset.$$

This contradicts the SSC.

Theorem 4.2.7:

Let Λ be a self-similar system with the OSC. Then $\dim_{\mathrm{H}} K_\Lambda = \dim_s K_\Lambda$. In particular, $\dim_{\mathrm{B}} K_\Lambda = \dim_s K_\Lambda$.

Remark 4.2.8:A glimpse into the modern era

We see that if Λ satisfies a certain separation condition then $\dim_{\mathrm{H}} K_\Lambda = \dim_s K_\Lambda$. On the other extreme, if there are exact overlaps, then $\dim_{\mathrm{H}} K_\Lambda < \dim_s K_\Lambda$. What happens in between is delicate. Nowadays, we know that the conclusion of this result holds for self-similar systems with the Exponential Separation Condition (ESC). We assume that Λ is homogeneous on $\mathbb{R}$, i.e. it has only one scaling/contraction ratio. Consider the set $\Lambda^N(\{0\})$ for $N \geq 1$. Let Δ_N denote the smallest gap among the numbers in $\Lambda^N(\{0\})$ (as a multiset as there might be exact overlaps). A deep result of Hochman says that if $\dim_{\mathrm{H}} K_\Lambda < \dim_s K_\Lambda$, then $\lim_{N\to\infty} \log|\Delta_N|/N \to \infty$. Therefore, the gaps decay super-exponentially. To appreciate this result, consider the situation when $\Lambda_t = \{x/2, (x+1)/2, (x+t)/2\}$. Then Δ_N is the value of a polynomial P of degree at most N with coefficients in $\{0, \pm 1, \pm t, \pm(1-t)\}$ at $1/2$. Therefore if t is an integer and Δ_N is not zero, then it is at least $1/2^N$. From here, we see that the system Λ_t for t being integers cannot have super-exponentially decaying gaps unless there are exact overlaps.

Proof. As in the proof of Theorem 4.2.4, for each small δ, we obtain $\mathcal{C}$ with sets of size $\asymp \delta$. In our situation, we will construct $\mathcal{C}$ with the initial set being an open set O for the OSC. Notice that O might not be connected. Together with the covering, we also assign each $B \in \mathcal{C}$ a weight which corresponds to its size (power the self-similarity dimension). We denote this measure as μ_δ. It is obtained by assigning each open set in $\mathcal{C}$ the corresponding weight and making it to be the scaled Lebesgue measure on this open set. Then $\lim_{\delta\to 0} \mu$ will converge to a probability measure on K_Λ. The limit is in fact μ_Λ for the chosen probability weight on the symbols Λ.

We will apply the Mass Distribution Principle. Let $x \in K_\Lambda$ and let $r > 0$ be a small number. We want to estimate $\mu_\Lambda(B_r(x))$. We do this by considering $\mu_\delta(B_r(x))$. Let $\delta = r$. We then construct $\mathcal{C}$ according to this δ. Since $x \in K_\Lambda$, at least one $B \in \mathcal{C}$ will intersect $B_r(x)$. Then there is a constant $c > 0$ depending on the initial choice of the open set O such that

$$B_{cr}(x) \supset B.$$

There might be more than one $B \in \mathcal{C}$ that intersect $B_{cr}(x)$. All of them are contained in $B_{c^2r}(x)$. Observe that the initial open set O contains a non-trivial ball of size $c' > 0$. This implies that each $B \in \mathcal{C}$ contains a non-trivial ball of size $\gg \delta c' \gg \delta$. The ball $B_{c^2r}(x)$ can contain at most $\ll (r/\delta)^k \ll 1$ many balls of size $\gg \delta$. Therefore we see that $B_r(x)$ can intersect $\ll 1$ many $B \in \mathcal{C}$. Those $B \in \mathcal{C}$ carry all the mass that could be captured by $B_r(x)$. This implies that

$$\mu_\Lambda(B_r(x)) \ll \delta^t.$$

Then the result follows from the Mass Distribution Principle. $\square$

We can now prove Theorem 4.2.2.

Proof of Theorem 4.2.2. Given K_Λ with $\overline{\dim}_{\mathrm{B}} K_\Lambda = s$. We see that there are arbitrarily small δ so that (for $s^- < s$)

$$N_\delta(K_\Lambda) \gg \delta^{-(s^-)}.$$

From here, we can find similar copies of K_Λ of scale δ that covers K_Λ. Among those similar copies, we can find

$$\gg \delta^{-(s^-)}$$

disjoint ones. We can now define a new Λ' that only includes those disjoint similar copies of K_Λ. We then find $K_{\Lambda'} \subset K_\Lambda$. Moreover, $K_{\Lambda'}$ satisfies the OSC. We then see that

$$\dim_{\mathrm{H}} K_{\Lambda'} = \dim_s K_{\Lambda'}.$$

We list the scaling ratios of Λ' as

$$\{r_1, \dots, r_N\}$$

for some $N \asymp (1/\delta)^t$. Moreover, $r_i \asymp r_j$ for all i, j with $\asymp$ depends on the original Λ only. From here we see that the equation

$$\sum_{i \le N} r_i^{t'} = 1$$

implies that

$$N\delta^{t'} \asymp 1.$$

This is saying that

$$\dim_s K_{\Lambda'} = t' = o_{\delta \to 0}(1) + \frac{\log N}{|\log \delta|}.$$

Then we have for some $\delta \to 0$

$$\dim_{\mathrm{H}} K_\Lambda \ge \dim_{\mathrm{H}} K_{\Lambda'} \ge s^-.$$

This finishes the proof. $\square$

The self-similarity is an extremely strong condition. To illustrate this idea we show the following result shows that self-similar sets with zero dimension are rather special. On the other hand, notice that general sets with zero dimension can still be rather complicated.

Theorem 4.2.9:

Let K_Λ be a self-similar set. Then $\dim_{\mathrm{H}} K_\Lambda > 0$ unless K_Λ is a singleton. Similarly, for each self-similar measure μ_K, there is no atomic support (i.e. $\exists x, \mu_K(\{x\}) > 0$) unless K_Λ is a singleton and μ_K is a Dirac mass.

Proof. Consider $\Lambda^N(K_\Lambda)$ for large number N. Suppose that $\Lambda^N(K_\Lambda)$ contains two disjoint similar copies of K_Λ. Then we can define a subsystem with these two copies and obtain a self-similar set with the OSC. This smaller self-similar set has a positive Hausdorff dimension and the original K_Λ must have a positive Hausdorff dimension as well. Thus the only way to achieve $\dim_{\mathrm{H}} K_\Lambda$ is that or all large N, all similar copies of $\Lambda^N(K_\Lambda)$ have a non-trivial intersection. This implies that $\mathrm{diam}(K_\Lambda) = 0$. Thus K_Λ is a singleton.

For μ_K, we can assume that all $p_f \neq 0$, if $\mu_K(\{x\}) > 0$, we can consider $x_f = f^{-1}(x)$ for $f \in \Lambda$. We have

$$0 < \mu_K(\{x\}) = \sum_f p_f \mu_K(\{x_f\}).$$

There is at least one x_f so that $\mu_K(\{x_f\}) \geq \mu_K(\{x\})$.

If we can find some x so that $\mu_K(\{x\}) > 0$ and all x_f with $\mu_K(\{x_f\}) \geq \mu_K(\{x\})$ satisfies $x_f = x$. Then we can collect some f terms with $x_f = x$ and write

$$\mu_K(\{x\}) = \alpha \mu_K(\{x\}) + \sum_{f : x_f \neq x} p_f \mu_K(\{x_f\}).$$

However, since $\mu_K(\{x_f\}) < \mu_K(\{x\})$ for $x_f \neq x$, we see that

$$\mu_K(\{x\}) = \alpha \mu_K(\{x\}) + \beta\rho$$

where $\alpha + \beta = 1, \alpha, \beta \geq 0, \rho < \mu_K(\{x\})$. Thus the only possibility is that $\beta = 0$. The only way to get this situation is that all the functions $f \in \Lambda$ have a common fix point x. However, this says that $K_\Lambda = \lim_{N\to\infty} \Lambda^N(\{x\})$ in Hausdorff distance and thus K_Λ is a singleton. This forces μ_K to be the Dirac mass on x.

Therefore there is a number $l > 0$ and dfferent points $x_1 = x, \dots, x_l$ such that for each $N > 0$, the preimages of $\{x_1, \dots, x_l\}$ under Λ^N that have mass at least $\mu_K(\{x\})$ are contained in $\{x_1, \dots, x_l\}$. Without loss of generality assume that they have the same mass. If not, we can take one with the maximum mass and get a subset of points with this maximum mass. Then we see that $\{x_1, \dots, x_l\}$ are fixed by taking preimages in Λ^{-1}, i.e. $\Lambda^{-1}(\{x_1, \dots, x_l\}) \subset \{x_1, \dots, x_l\}$. Then $\Lambda^{-N}\{x_1, \dots, x_l\}$ is a decreasing sequence of finite sets and it must be stabilised, i.e. for a non-empty X subset of $\{x_1, \dots, x_l\}$,

$$\Lambda^{-1} X = X.$$

This implies that $X = \Lambda(X)$. Thus $K_\Lambda = X$. Then $\dim_{\mathrm{H}} K_\Lambda = 0$ and this implies that K_Λ is a singleton. $\square$

4.3. Self-affine sets/measures

Include Takagi/Weierstrass functions

Part 2

N.E.Further Topics

CHAPTER 5

Combinatorics and Fractals (under construction)

5.1. Discretisation of fractals

Theorem 5.1.1:Almost AD-regularity

Let $E \subset \mathbb{R}^k$ be a compact set. Suppose that $\dim_{\mathrm{H}} E = s \in (0, k)$. Then for each $s^- < s < s^+$, there is $\mu \in \mathcal{P}(E)$ so that for μ almost all x,

$$s^- \leq \liminf_{r \to 0} \frac{\log \mu(B_r(x))}{\log r} \leq s^+.$$

Proof. As $s^- < s$, by Theorem 3.1.11 we can find an s^--Frostman measure μ with $\mu(E) = 1$. For this measure, we see that uniformly for all $x \in \mathbb{R}^k$,

$$\mu(B_r(x)) \ll r^{s^-}$$

as $r \to 0$. We cannot do better than this in the sense that the set of $x \in E$ with

$$\mu(B_r(x)) \ll r^{s^+}$$

has zero μ measure. From here we see that for μ almost all x,

$$s^- \leq \liminf_{r \to 0} \frac{\log \mu(B_r(x))}{\log r} \leq s^+.$$

□

We see that for general E, it is in general not possible to find an AD-regular measure in $\mathcal{P}(E)$. However, some regularity information can be obtained. The following result is useful in considering various combinatorial problems involving fractal sets.

Theorem 5.1.2:

Let $E \subset \mathbb{R}^k$ be a compact set. Suppose that $\dim_{\mathrm{H}} E = s$. Let $s^- < s$. For for large $n \geq 1$, E contains a 2^{-n}-separated set E_n so that

$$\#E_n \gg \frac{2^{ns^-}}{n}.$$

Moreover, for each $R \in (2^{-n}, 1)$ and any ball B_R of size R, the following non-concentration condition holds,

$$\frac{\#E_n \cap B_R}{\#E_n} \ll nR^{s^-}.$$

Remark 5.1.3:

Thus if R is of size $2^{-\rho n}$ for some $\rho > 0$, we can easily write

$$nR^{s^-}$$

as R^{s^-} for a different s^-.

Proof. Let $\mu \in \mathcal{P}(E)$ be s^--Frostman. Observe that for each $n \geq 1$

$$\sum_{B \in \mathcal{D}_n(E)} \mu(B) \geq 1.$$

We now regulate the size of $\mu(B)$ in the following way. For each $l \geq 0$ let

$$\mathcal{B}_l = \{B \in \mathcal{D}_n(E) : \mu(B) \in (2^{-l-1}, 2^{-l}]\}.$$

Observe that

$$\sum_{l \geq 2kn} \sum_{B \in \mathcal{B}_l} \mu(B) \ll 2^{kn} 2^{-2kn} = 2^{-kn}.$$

Thus as long as n is large enough, we do not need to consider $l \geq 2nk$. Among the choices $l \leq 2nk$, at least one such a choice of l will lead to

$$\sum_{B \in \mathcal{B}_l} \mu(B) \geq \frac{1}{2nk}.$$

Such an l cannot be too small because of the s^--Frostman property of μ. This leads to $l \geq n^{s-}$. We now examine the collection $\mathcal{B}_l$ closely. It is a collection of dyadic cubes of size $1/2^n$ and each cube has μ measure $\asymp 2^{-l}$. Thus we see that

$$2^l \gg \#\mathcal{B}_l \gg \frac{1}{2nk} 2^l.$$

On the other hand, by the s^--Frostman property of μ, we see that for each R-ball B_R of size $R \in (2^{-n}, 1)$,

$$\#\{B \in \mathcal{B}_l, B \cap B_R \neq \emptyset\} 2^{-l} \ll R^{s^-}.$$

This implies that

$$\frac{\#\{B \in \mathcal{B}_l, B \cap B_R \neq \emptyset\}}{\#\mathcal{B}_l} \ll nR^{s^-}.$$

□

5.2. Random projections of discrete sets

In order to gain further insights that were hidden behind the Fourier analytic proof of Marstrand's projection theorem, it is illustrative to consider a discrete version of Marstrand's projection theorem. We will develop a method showing that projections of 'nice' discrete sets tend to be as large as possible. Along this way, it is possible to prove a weaker version of Marstrand's projection theorem with purely combinatorial arguments. Neither Fourier Analysis nor Energy Integrals are needed.

Definition 5.2.1:

Let $\delta > 0$. A δ-separated set A_δ is a finite set in $\mathbb{R}^k$ such that all pairs of points are at least δ away from each other in terms of the Euclidean distance.

Definition 5.2.2:

Let $s, \delta > 0$. A δ-separated set $A_\delta \subset [0,1]^k$ is s-Frostman if for each number $R \in (\delta, 1)$ and any R-ball B,

$$\#A_\delta \cap B \leq \#AR^s.$$

Theorem 5.2.3:Discrete Marstrand's projection theorem

Let $\delta > 0$ and $A_\delta \subset [0,1]^2$ be s-Frostman where $s \in (0,1)$. For each $\theta \in S^1$, let $P_\theta = P_\theta(A_\delta)$ be the image of A_δ under the projection along the direction θ. Consider $N_\delta(P_\theta)$. Then for all $\theta \in S^1$ apart from a set with measure $\ll 1/|\log \delta|$,

$$N_\delta(P_\theta) \gg \delta^{-s}/|\log \delta|.$$

Remark 5.2.4:

With the help of this theorem and Theorem 5.1.2, we can show that for Lebesgue almost all θ, $\underline{\dim}_{\mathrm{B}} P_\theta(E) \geq s$ where E is a compact set with Hausdorff dimension s.

Proof. For each $\theta \in S^1$, consider

$$L_\theta = \int (\sum_{a \in A_\delta} P_\theta(B_\delta(a)))^2 dx$$

where dx is the Lebesgue measure on the image of P_θ (which is $\mathbb{R}$). We now estimate

$$\begin{aligned}
&\int d\theta \int (\sum_{a\in A_\delta} P_\theta(B_\delta(a)))^2 dx \\
&= \sum_{a,b\in A_\delta} \int\int d\theta P_\theta(B_\delta(a))P_\theta(B_\delta(b))dx \\
&\asymp \sum_{a,b\in A_\delta} \delta|\{\theta\in S^1 : P_\theta(B_\delta(a))\cap P_\theta(B_\delta(b))\neq\emptyset\}| \\
&\ll \sum_{a,b\in A_\delta} \frac{\delta^2}{|a-b|} \\
&\ll \delta \sum_{a\in A_\delta} \sum_{l\geq 0}^{2^l\delta\leq 1} \sum_{b\in A_\delta\cap B_{\delta 2^l}(a)} \frac{\delta}{2^l\delta} \\
&\ll \delta \sum_{a\in A_\delta} \sum_{l\geq 0}^{2^l\delta\leq 1} \frac{\#A_\delta(\delta 2^l)^s}{2^l} \\
&\ll \delta \sum_{a\in A_\delta} \#A_\delta\delta^s \ll \#A_\delta^2\delta^{s+1}.
\end{aligned}$$

Thus we see that

$$\{\theta : L_\theta \geq \#A_\delta^2\delta^{s+1}|\log\delta|\} \ll 1/|\log\delta|.$$

Suppose that $L_\theta \leq \#A_\delta^2\delta^{s+1}|\log\delta|$. Then we see that by Cauchy-Schwarz,

$$\delta^{-1}N_\delta(P_\theta)^{-1}(\int \sum_{a\in A_\delta} P_\theta(B_\delta(a))dx)^2 \ll L_\theta \leq \#A_\delta^2\delta^{s+1}|\log\delta|.$$

Observe that

$$\int \sum_{a\in A_\delta} P_\theta(B_\delta(a))dx \asymp \delta\#A_\delta.$$

Therefore we see that

$$N_\delta(P_\theta) \gg \delta^{-s}/|\log\delta|.$$

□

5.3. Erdős' distance problem and Falconer's distance problem

Let $E\subset\mathbb{R}^2$ be a set of N distinct points. Consider the distance set

$$\Delta(E) = \{|a-b| : a,b\in E\}.$$

The following result was posed by Erdős.

Theorem 5.3.1:

Let E be the integer lattice $\{0,\dots,n\}^2\subset\mathbb{R}^2$. There is a constant $c>0$, such that

$$\#\Delta(E) \leq c\frac{n^2}{\sqrt{2\log n}} \ll \frac{\#E}{\sqrt{\log\#E}}.$$

Remark 5.3.2:

Erdős conjectured that this lower bound is in fact optimal. Namely, for any finite set $E \subset \mathbb{R}^2$,

$$\#\Delta(E) \gg \frac{\#E}{\sqrt{\log \#E}}.$$

Recently, the problem of Erdős was almost solved.

Theorem 5.3.3:Guth-Katz

For any finite set $E \subset \mathbb{R}^2$,

$$\#\Delta(E) \gg \frac{\#E}{\log \#E}.$$

We do not prove those results here. (insert ref here) Instead, we now introduce a fractal version of the distance set problem.

Conjecture 5.3.4:Falconer's distance conjecture

Let $E \subset \mathbb{R}^k$ be a compact set with $\dim_{\mathrm{H}} E > k/2$. Then $\Delta(E)$ has a positive Lebesgue measure.

We will focus on the case when $k = 2$. In this case, it is known that the conclusion of Falconer's distance conjecture holds for $\dim_{\mathrm{H}} E > 5/4$. This is a deep result due to refs. On the other hand, if one is looking for sufficient conditions for $\dim \Delta(E) = 1$, then the problem becomes much more combinatorial and we have the following result due to Shmerkin.

Theorem 5.3.5:

Let $E \subset \mathbb{R}^2$ be a compact set with $\dim_{\mathrm{H}} E = \dim_{\mathrm{B}} E > 1$. Then $\dim_{\mathrm{H}} \Delta(E) = 1$.

We do not prove the above result. However, we will prove some weaker results which already contain interesting ideas.

Theorem 5.3.6:Falconer's distance theorem

Let $E \subset \mathbb{R}^2$ be compact with $\dim_{\mathrm{H}} E > 3/2$. Then $\Delta(E)$ contain intervals.

Proof. Let $\dim_{\mathrm{H}} E = s$. For each $s^- < s$, we can find $\mu \in \mathcal{P}(E)$ a uniformly s^--Frostman measure. For this measure, we consider its Fourier transform $\hat{\mu}$. Next, we consider the convolution $\eta = \mu * \mu_-$ where μ_- is the reflection of μ, i.e. $\mu_-(-A) = \mu(A)$ for each Borel set A. Then η is a probability measure supported on the difference set

$$E - E = \{x - y : x, y \in E\}$$

i.e. $\mathrm{supp}(\eta) \subset E - E$. We now study the pinned distance set $\Delta_0(E-E)$ which is equal to $\Delta(E)$. Observe that

$$\hat{\eta}(\xi) = |\hat{\mu}(\xi)|^2.$$

According to Theorem 3.3.9, we see that whenever $t < s$,

$$\int |\eta(\xi)| \frac{1}{|\xi|^{2-t}} d\xi < \infty. \tag{5.1}$$

Now, we consider the circle C_r centred at 0 with radius $r > 0$. Let $\delta > 0$ be a small number and we consider C_r^δ to be the thin annulus around C_r. We treat C_r as the probability Lebesgue measure on C_r. Consider a smooth bump function I_δ supported on $B_\delta(0)$ and $= \delta^2$ on $B_{\delta/2}(0)$. By considering the convolution $C_r * I_\delta$, we can effectively treat C_r^δ as a smooth function that is supported inside the set C_r^δ and $\asymp \delta^{-1}$ inside $C_r^{\delta/2}$. From now on, we write C_r^δ for this smooth function. Consider the following integral

$$f_\delta(r) = \int C_r^\delta(x) d\eta(x).$$

We can use the standard L^2-formula to see that

$$\begin{aligned} f_\delta(r) &= \int \hat{C}_r^\delta(\xi)\hat{\eta}(-\xi)d\xi \\ &= \int \hat{C}_r^\delta(\xi)\hat{\eta}(-\xi)d\xi \\ &= \int \hat{C}_r(\xi)\hat{I}_\delta(\xi)\hat{\eta}(-\xi)d\xi \\ &\ll_r \int |\xi|^{-1/2}|\hat{\mu}(\xi)|^2 d\xi \end{aligned}$$

where we used Theorem 3.4.3. The above asymptotic depends on the choice of r. However, we remark that as long as we constrain $r \in T$ for any fixed compact set T that does not contain the number 0, the above asymptotic is in fact uniform across $r \in T$. Next, for each fixed δ, $f_\delta(r)$ is a smooth function. Thus if we choose $t > 1.5$ then we have the following uniform asymptotic

$$f_\delta(r) \ll_r 1.$$

for all $r \in T$. We therefore see the following pointwise limit for all $r \in T$,

$$\lim_{\delta \to 0} f_\delta(r) = f(r) = \int \hat{C}_r(\xi)\hat{\eta}(-\xi)d\xi.$$

We also know that $\hat{C}_{r'} \to \hat{C}_r$ pointwisely and uniformly on each compact set. Thus we see that $f(r)$ is continuous for $r \in T$. The problem now is that f can be constantly zero. We show that this is not the case for a suitable choice of T. We choose $T = [\varepsilon, \varepsilon^{-1}]$ for small $\varepsilon > 0$. This can be checked by observing

$$\int_T f_\delta(r)dr \gg \eta(C_T)$$

where $C_T = \cup_{r\in T} C_r$. Observe that f is also the L^1-limit of $\lim_{\delta\to 0} f_\delta$. Therefore if f is zero over T, then $\eta(C_T) = 0$. For ε being small enough, this implies that $\mathrm{supp}(\eta) \subset B_\varepsilon(0)$. If this is the case for all such ε, we can only have $\mathrm{supp}(\eta) = \{0\}$. However, this would make (5.1) impossible. We conclude that f is continuous, non-negative, and not constantly zero on some large enough $T = [\varepsilon, \varepsilon^{-1}]$.

Now, of $f(r) > 0$, we see that $C_r^\delta \cap (E - E)$ for all small enough δ and this implies that $r \in \Delta(E)$. This proves the theorem. $\square$

5.4. The Kakeya problem

5.4.1. The bush and the hairbrush methods. We showed in Theorem 2.4.5 that any Kakeya set in $\mathbb{R}^k$ has lower box dimension two for $k = 2$. The situation is much more unclear for $k \geq 3$. We will introduce here two simple combinatorial methods for the Kakeya problem.

Theorem 5.4.1:The bush method

Let $K \subset \mathbb{R}^k$ be a Kakeya set. Then $\underline{\dim}_{\mathrm{B}} K \geq (k+1)/2$.

Remark 5.4.2:

The method for proving this theorem is more important than the statement of this result.

Proof. For each $\delta > 0$, we can first find a δ-separated subset $S_\delta \subset S^{k-1}$. Notice that $\# S_\delta \asymp \delta^{-(k-1)}$. We can then find the corresponding line segments (and make them δ-tubes) for directions in S_δ. As a result, we have a union of $\asymp \delta^{-(k-1)}$ many δ-tubes. As each δ-tube carries $\asymp \delta^{-1}$ many δ-balls, we see that counting multiplicities, we have $\asymp \delta^{-k}$ many such δ-balls.

Let $M > 0$ be a large number. Suppose that there is a δ-ball B so that some $x \in B$ has multiplicity at least M. Namely, there are at least M many δ-balls that contain x. Then there are at least M many line segments passing through B. Since the directions of those line segments are δ-separated, we see that the union of these at least M line segments must be covered with

$$\gg M/\delta$$

many δ-balls. On the other hand, if no x has multiplicity more than M, then clearly, we need $\gg \delta^{-k}/M$ many disjoint δ-balls in order to host those δ-tubes. From here we see that

$$N_\delta(K) \gg \min\{M/\delta, \delta^{-k}/M\}.$$

We can choose $M = \delta^{(k+1/2)}$ and get the result that

$$N_\delta(K) \gg \delta^{-(k+1)/2}.$$

This finishes the proof. $\square$

Theorem 5.4.3:The hairbrush method

Let $K \subset \mathbb{R}^k$ be a Kakeya set. Then $\underline{\dim}_{\mathrm{B}} K \geq (k+2)/2$.

Proof. For each $\delta > 0$, we can first find a 100δ-separated subset $S_\delta \subset S^{k-1}$. Notice that $\# S_\delta \asymp \delta^{-(k-1)}$. We can then find the corresponding line segments (and make them δ-tubes) for directions in S_δ. As a result, we have a union of $\asymp \delta^{-(k-1)}$ many δ-tubes. As each δ-tube carries $\asymp \delta^{-1}$ many δ-balls, we see that counting multiplicities, we have $\asymp \delta^{-k}$ many such δ-balls.

We now seek a different configuration for which our counting method in Theorem 2.4.5 can help. Let l^δ be one of the δ-tubes in consideration. We define $H = H_{l^\delta}$ to be the union of all δ-tubes that intersect l^δ and whose direction is at least 0.1-separated from that of l. We call $m = m_{l^\delta}$ to be the amount of such δ-tubes. We call H to be a hairbrush.

Suppose that there is a hairbrush $H = H_{l_0^\delta}$ with $m > M$ for some number $M > 0$. For each segment l so that l^δ is included in H, we see that the 2-dimensional affine subspace Π_{l,l_0} spanned by l_0, l essentially contains l^δ, l_0^δ. More precisely, $\Pi_{l,l_0}^{10\delta}$ contains l^δ, l_0^δ. As l ranges over the collection of segments that are included in the hairbrush H, we can find many different $\Pi_{l,l_0}^{10\delta}$. We can choose a subcollection of those $\Pi_{l,l_0}^{10\delta}$ to make sure that different Π_{l,l_0} are at least δ-separated away (say, with $> \rho$ distance for some $\rho > 0$) from l_0. We can now use the argument in Theorem 2.4.5 on each such $\Pi_{l,l_0}^{10\delta}$ and add the counting balls together. As a result, we will have $\gg M\delta^{-1}/\log(1/\delta)$ many disjoint δ-balls for this hairbrush H.

On the other hand, if there is no hairbrush with $m \geq M$. We perform the following counting trick. We call $l \in H$ if l is included in the hairbrush H. Then we see that

$$\sum_l \sum_{D: l' \in H_{l^\delta}, D\cap l \cap l' \neq \emptyset} 1 \ll M \frac{1}{\delta^{k-1}}$$

where D ranges over disjoint δ-balls that covers the segment l. We can perform the double sum in a different way

$$\sum_D \sum_{l,l': D\cap l\cap l' \neq \emptyset} 1 = \sum_D m^2(D)$$

where $m(D)$ is the multiplicity for D (not the hairbrush). By Cauchy-Schwarz we see that

$$\sum_D m(D)^2 \gg \frac{(\sum_D m(D))^2}{\#\{D : m(D) \neq 0\}}$$

Notice that $\#\{D : m(D) \neq 0\}$ is what we want to count. From here we see that

$$\#\{D : m(D) \neq 0\} \gg \frac{(\sum_D m(D))^2}{M\frac{1}{\delta^{k-1}}} \asymp \frac{1}{M}\frac{1}{\delta^{k+1}}.$$

We thus obtained that

$$N_\delta(K) \gg \min\left\{\frac{1}{M}\frac{1}{\delta^{k+1}}, M\delta^{-1}/\log(1/\delta)\right\}.$$

By choosing a specific M according to δ, we see that

$$N_\delta(K) \gg \frac{1}{\delta^{(k+2)/2}}\frac{1}{\log^{1/2}(1/\delta)}.$$

This finishes the proof. □

5.4.2. A variation of the Kakeya problem. There is no reason to just consider segments with all directions as in the traditional Kakeya problem. There are many other natural questions to ask about. For example, we can change segments to circles and ask for a compact set K to contain circles with all radius in $[1, 2]$. Such a set will also have a full dimension. This was proved by Wolff ref.

Here, we want to look at yet another variation of Kakeya set.

Definition 5.4.4:Dipole Kakeya set

Let $D \subset \mathbb{R}^k$ be compact. We call D to be a dipole Kakeya set if for each $\theta \in S^{k-1}$, there is some $t_\theta \in \mathbb{R}^k$ such that

$$\{t_\theta \pm \theta\} \subset D.$$

Remark 5.4.5:

The difference between D and a Kakeya set is that D contains the pair of endpoints for segments in each direction while a Kakeya set asks for the full segment to be included.

We do not know how small can D be, even for $k = 2$. A good guess is that

$$\underline{\dim}_{\mathrm{B}} D \geq \frac{k-1}{2}\frac{k}{k-1}. \tag{5.2}$$

For $k = 2$, we expect that $\underline{\dim}_{\mathrm{B}} D \geq 2/3$. Trivial counting arguments can show that

$$\underline{\dim}_{\mathrm{B}} D \geq \frac{k-1}{2}.$$

Therefore the expected optimal lower bound for $\underline{\dim}_{\mathrm{B}} D$ is only slightly larger than the trivial bound. The conjectural lower bound (5.2) is in fact attained.

Theorem 5.4.6:

There exist dipole Kakeya sets D in $\mathbb{R}^k$ with

$$\underline{\dim}_{\mathrm{B}} D = \frac{k-1}{2}\frac{k}{k-1}.$$

Proof. See [10]. □

For $k = 2$ we can show the following improvement from the trivial bound $1/2$.

Theorem 5.4.7:

Let D in $\mathbb{R}^2$ be a dipole Kakeya set. Then

$$\underline{\dim}_{\mathrm{B}} D \geq \frac{4}{7} > \frac{1}{2}.$$

Proof. See [10]. (to be explained here) □

For the Hausdorff dimension, the situation is completely different to the story of Kakeya set.

Theorem 5.4.8:

There exist dipole Kakeya sets in $\mathbb{R}^k$ with zero Hausdorff dimension.

Proof. See [10]. □

CHAPTER 6

Additive Combinatorics and dimensions of self-similar sets

Proofs in this chapter are not complete and may contain so many mistakes. I tried to illustrate the ideas clearly. For perfect proofs, see [**3**, **5**, **9**].

6.1. super-exponential gaps in self-similar sets

We know that for a self-similar set K_Λ,

$$\dim_{\mathrm{H}} K_\Lambda \leq \dim_s K_\Lambda.$$

See also Remark 4.2.5. We have the following conjecture.

Conjecture 6.1.1:The exact overlap conjecture

Let K_Λ be a self-similar system. Then $\dim_{\mathrm{H}} K_\Lambda < \dim_s K_\Lambda$ if and only if there are exact overlaps in Λ.

This problem is open for self-similar sets in $\mathbb{R}^k$ for all $k \geq 1$. For self-similar sets in $\mathbb{R}$, there are partial results. For those results, we need some additive combinatorics. To get an idea of why additive combinatorics is useful here. We consider the situation when Λ has a unique scaling ratio $0 < r < 1$. Without loss of generality, we assume that K_Λ has convex hull $[0, 1]$.

Let $N > 0$ be an integer. Then $\Lambda^N(K_\Lambda)$ is a union of r^N scaled copies of K_Λ. Then we can decompose $[0, 1]$ into disjoint union of intervals of length r^N. For each such interval I_N, there might be more than one copies of K_Λ that intersect I_N. Let those copies to be $\mathcal{I}_N$. Each copy can be obtained by

$$r^N(K_\Lambda) + a$$

for some $a \in \mathbb{R}$. Let A_N be the collection of such a for $\mathcal{I}_N$. Then at I_N, $K_\Lambda \cap I_N$ is basically the sumset

$$A + r^N K_\Lambda.$$

Let us assume (wrongly) that there is some $\varepsilon > 0$ so that for all $D > 1$, the covering number

$$N_{r^{DN}}(A + r^N K_\Lambda) \gg N_{r^{DN}}(A)^\varepsilon N_{r^{DN}}(r^N K_\Lambda) \gg N_{r^{DN}}(A)^\varepsilon N_{r^{(D-1)N}}(K_\Lambda) \tag{6.1}$$

$$= N_{r^{DN}}(A)^\varepsilon r^{(D-1)N(s+o(1))} \tag{6.2}$$

where $s = \dim_{\mathrm{H}} K_\Lambda = \dim_{\mathrm{B}} K_\Lambda$. This is where additive combinatorics becomes useful since we want to study the structure of the sumset $A + r^N K_\Lambda$.

Since no exact overlaps are in Λ, we see that for some large enough D, $N_{r^{DN}}(A)^{\varepsilon} = \#A^{\varepsilon}$ for all such A. We can then add the above asymptotic to see that

$$N_{r^{DN}}(K_{\Lambda}) \gg (\sum_{A} \#A^{\varepsilon}) r^{(D-1)N(s+o(1))}.$$

For $\sum_A \#A^{\varepsilon}$, observe that

$$\sum_{A} \#A = (\#\Lambda)^N = r^{N \dim_s K_{\Lambda}}.$$

The amount of such A is $r^{N(s+o(1))}$. Thus a typical A will have

$$\#A \asymp r^{N(\dim_s K_{\Lambda} - s - o(1))}. \tag{6.3}$$

This is non-trivial as we assumed that $\dim_s K_{\Lambda} > \dim_{\mathrm{H}} K_{\Lambda} = s$. Then we see that

$$\sum_{A} \#A^{\varepsilon} \gg r^{N(s+o(1))} r^{\varepsilon N(\dim_s K_{\Lambda} - s - o(1))}.$$

This forces (for some other $\varepsilon > 0$)

$$N_{r^{DN}}(K_{\Lambda}) \gg (\sum_{A} \#A^{\varepsilon}) r^{(D-1)N(s+o(1))} \gg r^{DN(s+o(1))} r^{\varepsilon N}$$

If D is bounded by M. This implies that

$$\dim_{\mathrm{B}} K_{\Lambda} \geq s + \frac{\varepsilon}{M}.$$

This is not possible. Thus the only way to get out of this situation is that D must tend to ∞ as $N \to \infty$. However, D is chosen so that the N-level copies of K_{Λ} are $\asymp r^{DN}$-separated translations, we arrive at the following fake result because we assumed something wrong.

THEOREM (The not yet proved Super-exponential gap theorem). Let K_{Λ} be a self-similar system with a uniform scaling ratio. Then $\dim_{\mathrm{H}} K_{\Lambda} < \dim_s K_{\Lambda}$ only if there are super-exponential overlaps in Λ. Namely, for $N \to \infty$, $G_N = \min\{\Delta(\Lambda^N(\{0\}))\}$ where $\Delta(.)$ is the distance set satisfies

$$\frac{\log |G_N|}{N} \to \infty.$$

We will explore various methods for treating the sum set $A + r^N K_{\Lambda}$. Essentially, we have two discrete sets A, K where A is an arbitrary set and K is a well-structured set (being regular). Then from $\#(A+K) \ll \#A^{\varepsilon}\#K$, it is possible to infer that $\#A$ must be already quite small a some quantitative way. However, we can think of A being not that small from (6.3). This can help us rectify the wrong assumption (6.1) and establish the super-exponential gap theorem.

Although the super-exponential gap theorem does not solve the exact overlap conjecture, it greatly improves our understanding towards the structure of self-similar sets. For example, we have the following result.

Theorem 6.1.2:

Assuming the super-exponential gap theorem, if Λ is a self-similar system on $\mathbb{R}$ defined with a uniform scaling ratio and with algebraic parameters, then the exact overlap conjecture holds, i.e. if $\dim_{\mathrm{H}} K_{\Lambda} < \dim_s K_{\Lambda}$, then there are exact overlaps in Λ.

PROOF. Consider $\Lambda^N(\{0\})$. It is the value of polynomials with certain algebraic coefficients at the algebraic number r. We can assume that the translation parameters are algebraic integers.

For each $x \in \Lambda^N(\{0\})$, there is a polynomial P with coefficients in $T = \{a_1, \dots, a_N\}$ which are the translation parameters of the linear maps in Λ such that

$$\deg P = N$$

and $x = P(r)$. Thus for each $x, y \in \Lambda^N(\{0\})$, $x - y$ is the value of a degree N polynomial $Q_{x,y}$ with coefficients in $T - T$ at r. Let $x - y\alpha$. It is an algebraic number. Consider the conjugates $\alpha_1 = \alpha, \dots, \alpha_L$. There is a upper bound for each α_i. Namely, there is a constant $C > 0$ depending on a_i, r and their conjugates such that

$$\max_i \alpha_i < C^N.$$

However, $\prod_i \alpha_i = p/q \in \mathbb{Q}$ for some $q \leq Q^N$ where Q is an integer such that Qr is an algebraic integer. Thus we see that

$$x - y = \alpha \gg \frac{1}{C^{NL} Q^{LN}}.$$

Since L, C, Q are constants, we see that Λ does not have super-exponential gaps other than the exact overlaps. Then the result follows from the super-exponential gap theorem. $\square$

6.2. Additive combinatorics basics

Here we use $|A|$ to denote the size of A. It is more convenient to write than the other notation $\#A$.

Given an abelian group G, let $A, B \subset G$ be finite sets. It is of great interest to study the sumset $A + B$. For our study, we will concentrate on the size growth of $|A + B|/|A|$ or $|A + B|/|B|$. On the one hand, it is clear that

$$\max\{|A|, |B|\} \leq |A + B| \leq |A||B|.$$

We are interested in the case when $|A + B|$ is not much larger than $|A|$. Consider first the case when $B = \{e, b\}$ for $b \neq e \in G$. Then $|A + B| = |A|$ if for each $a \in A$, the element $a + b \in A$ as well. This can only happen if A is isomorphic to a cyclic group, in other words, A is a coset of a cyclic subgroup of G. This is a rather trivial example. The intuition is that if $|A + B|$ is not much larger than $|A|$ then A itself must resemble a coset. Given that $|A + B| < c|A|$ for some constant $c > 0$ and a large set B with $|B|$ being much larger than c, it can be checked that a significant proportion of A is a union of a small number of large cosets of G.[1] This will then make $|A + B + B|$ not too large compared with $|A|$. The following important result was due to Plünnecke and simplified by Ruzsa.

Theorem 6.2.1:PR-inequality

Let A, B be finite subsets of an infinite abelian group G. Suppose that $|A + B| \leq c|A|$ for a number $c > 0$. Then for each n,

$$|nB| < c^n |A|.$$

[1]See Freiman's theorem for a more detailed statement.

PROOF. Menger's theorem in graph theory and a tensor product trick. Omit for now. See [8, Chapter 1]. □

We apply this result in the context of fractal geometry. Let A, B be finite sets with δ-sep. Such sets are often obtained by finite scale approximation of fractal sets. We often assume that $\delta = 1/2^N$ for integers N and $A, B \subset \mathbb{Z}2^{-N}$. Suppose that $|B| > 1$, then for any integer $T > 0$,

$$2^T \leq |2^T B| \leq c^{2^T} |A|.$$

This implies that $|A| \geq 2^T c^{-2^T}$. This is not an impressive result as for large T, the lower bound is less than one unless c is extremely close to one. The idea is to apply this kind of argument for multiple scales rather than one and accumulate the increments c for those different scales.

A convenient way to perform the multi-scale analysis is to use the concept of the coding tree. Let us consider $A, B \subset [0, 1]$. Then we can construct the binary trees T_A, T_B. If $A, B \subset \mathbb{Z}2^{-N}$, then we only need the first N-levels of T_A, T_B.

Let $N = mT$ for integers N, m, T. We group T many levels of the trees together. Let B be such that for each $s \in \{1, \ldots, m\}$. There are only two nodes in $T_{B,sT}$ for each subtree sandwiched in levels $\{(s-1)T + 1, \ldots, sT\}$. Moreover, each of the two nodes, viewed as dyadic intervals of length 2^{-sT}, are at least $2^{-sT} 2^{n_s}$-apart and at most $2^{-sT} 2^{n_s+1}$ apart for some integer $n_s \in [0, \ldots, T-1]$. We call such a T_B to be a binary toy tree with thickness T.

We now perform the iterated sum $2^{2T} B$. First, we can obtain a binary coding for B by writing $\omega \in \{0, 1\}^m$ to indicate the choice of left/right branches.

Define the following map ($n_0 = 0$, $-n_s + n_{s-1} - 1 \leq T$),

$$\Sigma : B_\Sigma = \prod_{s=1}^{T} \{0, \ldots, \min\{2^{T - n_s + n_{s-1} - 1}\}\} \to 2^{2T} B$$

in the following way.

First for $(t_1, \ldots, t_m) \in B_\Sigma$, we write out a $m \times 2^{2T}$ table of symbols in $\{0, 1\}$ by writing in row i with t_i many 1's followed by zeros. Then each column represents a number in B by considering its binary code. There are 2^{2T} many columns and this table represents an element in $2^{2T} B$ by taking the sum of the numbers represented by each of the columns. In this way, we obtain $|B_\Sigma|$ many possible numbers in $2^{2T} B$. For two different $(t_1, \ldots, t_m), (t'_1, \ldots, t'_m)$. We find their first different digit, say $t_i > t'_i$. Then no matter how to choose the rest of the digits, we have

$$\Sigma(t_1, \ldots, t_m) - \Sigma(t'_1, \ldots, t'_m) \geq 2^{-i(T - n_i)} - 2 \sum_{i' \geq i+1} 2^{T - n_{i'} + n_{i'-1} - 1} 2^{-i'T}.$$

This is strictly bigger than zero and therefore bigger or equal to 2^{-N}. This implies that

$$|2^{2T} B| \geq 2^{mT + \sum_{s=1}^{m} (-n_s + n_{s-1}) - m} \geq 2^{(m-1)T - m}.$$

Then we can use PR-inequality to see that

$$|A| \geq 2^{(m-1)T - m} c^{-2^{2T}}.$$

To get a more transparent expression, we write $|A| = 2^{\alpha N}, |B| = 2^{\beta N}$ and

$$c = 2^{\delta N}.$$

Thus α, β, δ can be interpreted as dimensions. From here we see that

$$\alpha \geq \frac{m(T-1)-m}{N} + \delta 2^{2T}.$$

As $N = mT$ with some fixed T but large m, we have

$$\alpha > 1^{-} + \delta 2^{2T}.$$

This result is much less trivial than its single-scale analogy.

Theorem 6.2.2:A toy result

Let $T > 1$ be an integer. Let $N = mT$ with $m > 1$ be an integer. Let $A, B \subset \mathbb{Z}2^{-N} \cap [0,1]$ with $|A| = 2^{\alpha N}, |B| = 2^{\beta N}$ and

$$c = 2^{\delta N}$$

such that $0 < \alpha, \beta < 1, |A+B| = c|A|$. Moreover, T_B is a binary toy tree (thickness T). Then for each $\alpha^{+} > \alpha$, all large enough $m > 0$,

$$\delta > \frac{1-\alpha^{+}}{2^{2T}}.$$

Remark 6.2.3:

T_B being a binary toy tree says that $\beta = \frac{2}{T}$. We then obtain a non-trivial 'dimension' increase of $A+B$ compared with A of amount

$$\delta \gg \beta T / 2^{2T}.$$

This is no longer trivial for any T.

6.3. Bourgain's sumset estimate

We want to extend Theorem 6.2.2. The reason is that in applications, we often do have have any information available for B except for its size. Therefore it is desirable to weaken the structural constraint for B. This will be done in this section.

First, we want to extract from B a suitable subset B' with that $T_{B'}$ are not too much different from being a binary toy tree with some thickness T which is a fixed large constant that will be determined upon other parameters. The tree T_B may be complicated. We first want to regularise it by taking a nicer subtree without a significant loss of the number of nodes at the last level $N = mT$.

We decompose the levels $0, \dots, mT$ into disjoint layers $\{0, \dots, T-1\}, \{T, \dots, 2T-1\}, \dots, \{(m-1)T, \dots, mT\}$. Consider the last group of levels. The tree T_B at these levels can be decomposed into a disjoint set of subtrees of height T. We can label each of those subtree with a unique number r between 0 and T which indicates the size of its last level set of nodes is in $[2^r, 2^{r+1}-1]$. Next, if a subtree has only one node or two nodes and they are adjacent (i.e. they represent adjacent dyadic intervals), then we override the label of this subtree to be 'NULL'. Finally, there is some number r' between 0 and T that represents the minimal distance between two last-level nodes (roughly $2^{r'}$). We then consider paired label (r, r'). By the pigeonhole principle, there is a label L_{m-1} so that all subtrees with this label contribute $\geq |B|/(T+1)^2$ many last-level nodes. Among those nodes, we can then choose about half of the last-level nodes to make sure that there are no adjacent nodes. After this, there are still $\geq |B|/(3(T+1)^2)$ last level nodes left.

We then ignore all the other subtrees in this layer. We can then perform the same procedure in the layer $\{(m-2)T, \ldots, (m-1)T-1\}$ as well as all the layers above. We then obtain a subtree T'_B with the following property:

- The size of the last level nodes $|T'_{B,N}|$ is at least $|B|/(3(T+1)^2)^m$. We write $|B| = 2^{mT\beta}$ for $\beta > 0$ to indicate the 'dimension' of B. Then we see that (assuming $T > 3$)
$$|T'_{B,N}| \geq 2^{mT\beta} 2^{-\log(3(T+1)^{2m})} = 2^{mT(\beta - 2\log(T+1)/T)}.$$
Thus if T is large, then the above tells us that we still maintain the original 'dimension' of B. The choice of T will then depend on β.
- For each layer $L_j = \{jT, \ldots, (j+1)T-1\}$, all subtrees in it have roughly the same amount of the last level nodes and the same minimal separating distance between two pair of last level nodes.

The tree T'_B gives us a subset $B' \subset B$ with $T'_B = T_{B'}$. If all layers (of thickness T) of $T_{B'}$ are not NULL (i.e. the subtrees sandwiched in this layer are not labeled as being NULL), then we can find a binary toy subtree of $T_{B'}$. The problem now is that only some fractions of the layers are not NULL. This will indicate that $|B'|$ is small. For example, if only an η-portion of the layers of $T_{B'}$ are not NULL, then
$$|B'| \ll 2^{\eta N}.$$
Thus η cannot be much smaller than $\beta > 0$. Therefore, we must maintain a positive proportional layers in $T_{B'}$ that are not NULL. The problem now is to analyse the 'dimension increment' even though the non-NULL layers are not consecutive.

We cannot afford to regularise A as it would cause a dimension drop $\log T/T$ and we can only have a dimension gain $\approx 1/2^{2T}$. We then assume that A is already regular in the sense that inside each layer, all subtrees have essentially the same number of last-level nodes. (We do not assume any separation for A.) We say that A has the uniform branching property.

For each layer $\{sT+1, \ldots, (s+1)T\}$, the branching number of A (i.e. the number of last level nodes for sandwiched subtrees in this layer) is roughly 2^{r_s} for some $n_s \in [0, T]$. Consider the sumset $A + B'$ at scale 2^{-sT}. We treat A, B' as union of dyadic intervals of length 2^{-sT}. Then we treat $A + B'$ as a union of dyadic intervals of length 2×2^{-sT}. Those intervals may not be disjoint. For each such dyadic interval, we consider the scale 2^{-sT-T}. To obtain such an interval, we choose one interval for A and one for B' and perform the sumset for the numbers inside A, B' and the chosen intervals. Multiple choices are possible. Trivially, the branching number for this interval of length 2×2^{-sT} from scale 2^{-sT} to 2^{-sT-T} is at least the branching number for A' in the corresponding layer. From here, we have the following trivial estimate that
$$T_{A+B',sT+T}/T_{A+B',sT} \geq 2^{r_s}.$$
Combining the above estimate for each layer we obtain that
$$|A + B'| \geq |A|,$$
which is trivial, however, it illustrates how one can accumulate gains of dimensions in layers.

Now, we want to bring in additive combinatorics. For a dyadic interval of length 2×2^{-sT} and some intervals for A, B' to obtain the sum, we write A_s, B'_s and $A'_s+B'_s$. We can zoom A_s, B'_s so that they are subsets of $[0,1]$ and they are 2^{-T}-separated so w.o.l.g. we assume that they are subsets of $\mathbb{Z}2^{-T}$. Let $c_s = |A_s + B'_s|/|A_s|$. If B'_s is not trivial (i.e. this layer is not NULL), we see that (by PR inequality for 2^{2T} iterated sum),

$$c_s \geq (\frac{2^{T-n_s+n_{s-1}}}{|A_s|})^{1/2^{2T}},$$

where it is understood that $n_s = T$ for NULL layers. From here, it is convenient to assume that for all T being large enough,

$$|A_s| \leq 2^{\alpha^+ T}$$

for some number $\alpha^+ > \alpha$. We do not require that $\alpha^+ < 1$ for now. Thus, the choice of T also depends on α^+. From here, we see that

$$c_s \geq 2^{((T-n_s)-\alpha^+ T)/2^{2T}}.$$

From here we see that branching number of $A'_s + B'_s$ in levels sT to $(s+1)T$ is

$$\geq 2^{((T-n_s+n_{s-1})-\alpha^+ T)/2^{2T}}.$$

From here, we see that branching number of $A' + B'$ in levels sT to $(s+1)T$ is

$$\geq 2^{((T-n_s+n_{s-1})-\alpha^+ T)/2^{2T}}.$$

If on the other hand, we have a NULL layer for $T_{B'}$, then we see that the branching number of $A' + B'$ in levels sT to $(s+1)T$ is

$$\geq 1.$$

We can multiply all the estimates from each layer to obtain

$$|A + B'|/|A| \geq 2^{\Delta}$$

where (some $-n_s$ may not be added back by $+n_s$ but it will be added back by T, except for the last n_s)

$$\Delta \geq \eta \frac{mT - \alpha^+ mT - T}{2^{2T}} = \eta \frac{1-\alpha^+}{2^{2T}} mT + o_m(1)$$

where ηmT is the number of non-NULL layers for $T_{B'}$. Notice that $\eta \geq \beta$. We then have the following extension of Theorem 6.2.2 where we no longer have a strong constraint for B. However, we added some conditions on A.

Theorem 6.3.1:Bourgain's sumset estimate: weak version

Let $T > 1$ be an integer. Let $N = mT$ with $m > 1$ be an integer. Let $A, B \subset \mathbb{Z}2^{-N} \cap [0,1]$ with $|A| = 2^{\alpha N}, |B| = 2^{\beta N}$ and

$$c = 2^{\delta N}$$

such that $0 < \alpha, \beta < 1, |A + B| = c|A|$. Suppose that A has the uniform branching property. Suppose that T-layer branching number for A is at most $2^{\alpha^+ T}$ for some $\alpha^+ < 1$. Then for all $\beta^- < \beta$, all large enough $m > 0$,

$$\delta > \beta^- \frac{1-\alpha^+}{2^{2T}}.$$

This is almost what we needed in (6.1). Unfortunately, the uniform branching property for A is still too strong to ask for even a self-similar set. Nonetheless, this idea of Bourgain is deep and far-reaching with many applications.

6.3.1. dropping the regularity requirement on A. In this section, we will weaken the uniform branching property requirement for A in Theorem 6.3.1.

To do this, it is helpful to have a look at the single-layer situation. Let A, B be finite sets that are 2^{-N}-sep and $N = mT$ for integers m, T. For $s \in \{0, \dots, m-1\}$, we look at T_A, T_B at layer s, i.e. levels from sT to $(s+1)T$. Take T_A, we consider its subtrees of level T in this layer. For each $x \in A$, we write $T_A(s, T, x)$ to the T level subtree whose root corresponds to a dyadic interval D_{sT+1} which contains x. Clearly, $T_A(s, T, x)$ and $T_A(s, T, y)$ are the same tree if x, y is contained in a same dyadic interval D_{sT+1}.

Given all such subtrees $T_A(s, T, x)$ we can compute the total branching number of A from level sT to level $(s+1)T$ as follows

$$N_{sT,(s+1)T}(A) := \frac{|T_{A,(s+1)T}|}{|T_{A,sT}|} = \frac{\sum_{T_A(s,T,x)} |T_A(s,T,x)|}{\sum_{T_A(s,T,x)} 1}$$

where the sum is taken over all possible subtrees $T_A(s, T, x)$. The reason to study this number is clear from the following simple relation

$$|A| = \prod_{s=1}^{m} N_{sT,(s+1)T}(A).$$

Now consider the tree T_B, we want to study T_{A+B} at layer s. We will always assume that T_B is regular in the sense that all subtrees $T_B(s, T, x)$ have roughly the same branching number. Let us assume that this branching number is $2^{1^- T}$ for some number $1^- < 1$ that is close to 1.

To study T_{A+B} in layer s, we look at the geometric information encoded in T_B, T_A. For B, we see that it intersects certain dyadic intervals in $\mathcal{D}_{sT}$. Let D be such an interval. We see that the dyadic decomposition of $D \cap B$ involves $2^{1^- T}$ intervals in $\mathcal{D}_{(s+1)T}$. For each $x \in A, y \in B$, we can locate $x+y \in A+B$. consider the trees $T_A(s, T, x), T_B(s, T, y)$. They represent dyadic intervals $I_A(s, T, x), I_B(s, T, y)$. Clearly $x + y \in I_A(s, T, x) \cap A + I_B(s, T, y) \cap B$. In some sense,

$$I_A(s, T, x) \cap A + I_B(s, T, y) \cap B$$

tells us some information of $T_{A+B}(s, T, x + y)$. To be more precise, the sum of two dyadic intervals $I_A(s, T, x) + I_B(s, T, y)$ is a union of two adjacent dyadic intervals of level sT. Therefore $(I_A(s, T, x) + I_B(s, T, y)) \cap (A + B)$ represents a tree of level $T + 1$ which encodes $(I_A(s, T, x) + I_B(s, T, y)) \cap (A + B)$ from level $sT - 1$ up to $(s + 1)T$. Clearly,

$$I_A(s, T, x) \cap A + I_B(s, T, y) \cap B \subset (I_A(s, T, x) + I_B(s, T, y)) \cap (A + B).$$

To study $A + B$ in the layer s, we need to consider all possible $x + y$ for $x \in A, y \in B$. Observe that

$$\begin{aligned} N(sT,(s+1)T)(A+B) &= \frac{\sum_{T_{A+B}(s,T,z)} |T_{A+B}(s,T,z)|}{\sum_{T_{A+B}(s,T,z)} 1} \\ &\geq \frac{1}{2} \frac{\sum_{T_{A+B}(s,T,z)} |(T_{A+B}(s,T,z)| + |T'_{A+B}(s,T,z)|)}{\sum_{T_{A+B}(s,T,z)} 1} \end{aligned}$$

where $T'_{A+B}(s,T,z)$ is the subtree that corresponds to the right adjacent dyadic interval of $I_{A+B}(s,T,z)$. Then we see that

$$\begin{aligned}&\sum_{T_{A+B}(s,T,z)} |(T_{A+B}(s,T,z)| + |T'_{A+B}(s,T,z)|)\\ \geq &\sum_{T_{A+B}(s,T,z)} \max_{x+y\in I_{A+B}(s,T,z)} N_{sT,(s+1)T}(I_A(s,T,x)\cap A + I_B(s,T,y)\cap B).\end{aligned}$$

This is because for each x,y with $x+y\in I_{A+B}(s,T,z)$,

$$I_A(s,T,x)\cap A + I_B(s,T,y)\cap B \subset I_{A+B}(s,T,z)\cup I'_{A+B}(s,T,z)$$

where $I'_{A+B}(s,T,z)$ is the right adjacent dyadic interval of $I_{A+B}(s,T,z)$. We can further obtain

$$\begin{aligned}&\sum_{T_{A+B}(s,T,z)} \max_{x+y\in I_{A+B}(s,T,z)} N_{sT,(s+1)T}(I_A(s,T,x)\cap A + I_B(s,T,y)\cap B)\\ \geq &\sum_{T_{A+B}(s,T,z)} \frac{\sum_{(x,y)\in K(z)} N_{sT,(s+1)T}(I_A(s,T,x)\cap A + I_B(s,T,y)\cap B)}{|K(z)|}\end{aligned}$$

where $K(z)\subset A\times B$ is a maximal subset with the property that for $(x,y)\in K(z)$,

$$x+y\in I_{A+B}(s,T,z)$$

and that for each $I_A(s,T,.)$ exactly one x is chosen, for each $I_B(s,T,.)$ exactly one y is chosen.

From here, we first make the trivial observation that

$$N_{sT,(s+1)T}(I_A(s,T,x)\cap A + I_B(s,T,y)\cap B) \geq \frac{1}{2} N_{sT,(s+1)T}(I_A(s,T,x)\cap A).$$

Then we see that

$$\begin{aligned}&\sum_{T_{A+B}(s,T,z)} |T_{A+B}(s,T,z)|\\ \geq &\frac{1}{4}\sum_{T_{A+B}(s,T,z)} \frac{\sum_{(x,y)\in K(z)} N_{sT,(s+1)T}(I_A(s,T,x)\cap A)}{|K(z)|}\\ = &4^{-1}\sum_{T_A(s,T,x)} N_{sT,(s+1)T}(I_A(s,T,x)\cap A) \sum_{T_{A+B}(s,T,z),T_B(s,T,y):(x,y)\in K(z)} \frac{1}{|K(z)|}\\ \geq &4^{-1}\sum_{T_A(s,T,x)} N_{sT,(s+1)T}(I_A(s,T,x)\cap A) \sum_{T_{A+B}(s,T,z)} \frac{1}{|T_{A,sT}|}\\ \geq &4^{-1}|T_{A+B}(s,T,z)| N_{sT,(s+1)T}(A).\end{aligned}$$

This implies that

$$N_{sT,(s+1)T}(A+B) \geq 4^{-1} N_{sT,(s+1)T}(A).$$

We can make a second trivial observation that

$$N_{sT,(s+1)T}(I_A(s,T,x)\cap A + I_B(s,T,y)\cap B) \geq \frac{1}{2} N_{sT,(s+1)T}(I_B(s,T,x)\cap A) \geq \frac{1}{2} 2^{1-T}.$$

From here we have

$$N_{sT,(s+1)T}(A+B) \geq 4^{-1}2^{1-T}.$$

Now, consider the layers $s \in \{0, \dots, m-1\}$. Let $\mathcal{S}$ be such that

$$S = \{s \in \{0, \dots, m-1\} : N_{sT,(s+1)T}(B) \geq 2^{1^-T}\}.$$

Remember that we assumed that B is regular in this sense that all $T_B(s, T, x)$ have roughly the same branching number. From here we see that

$$|A + B| = \prod_{s=1}^{m} N_{sT,(s+1)T}(A + B) \geq \frac{1}{4^m} 2^{1^-T|S|} \prod_{s \notin \mathcal{S}} N_{sT,(s+1)T}(A).$$

6.3.2. The wrong estimate. In this section, we adopt the following wrong version of PR-inequality:

Wrong version:Let A, B be finite subsets of an infinite abelian group G. Suppose that $|A + B| \leq c|A|$ for a number $c > 0$. Then for each n,

$$|A + nB| < c^n |A|.$$

Although this estimate is wrong, it will be useful in illustrating ideas. Later on, we will introduce the notion of entropy to replace our usage of the cardinality.

Let $T > 1$ be an integer. Let $N = mT$ with $m > 1$ be an integer. Let $A, B \subset \mathbb{Z}2^{-N} \cap [0, 1]$ with $|A| = 2^{\alpha N}, |B| = 2^{\beta N}$.

We first regulate B and obtain a subset B' so that $T_{B'}$ is the property that for all $s \in \{0, \dots, m-1\}$, all $T_{B'}(s, T, x)$ have roughly the same branching property. It can happen that $T_{B'}(s, T, x)$ are all singly branched (i.e. Null) for all x for some s. On the other hand, we have $|B'| = |T_{B',mT}| \geq 2^{mT(\beta - 2\log(T+1)/T)}$. This information tells that $T_{B'}$ cannot be Null for too many layers. Let

$$S = \{s : N_{sT,(s+1)T}(B') \geq 2\}$$

be the non-Null layers of B'. We have the following estimate

$$2^{mT(\beta - 2\log(T+1)/T)} \leq |B'| \leq 2^{|S|T}.$$

This tells us that

$$|S| \geq m(\beta - 2\log(T + 1)/T).$$

We now consider the sumset $2^{2T} B'$. It has the property that whenever $s \in \mathcal{S}$, we have

$$N_{(s-1)T,sT}(2^{2T} B') \gg 2^T.$$

Notice a layer shift in the above inequality. By knowing that B' is not null on some s-th layer, $2^{2T} B'$ is almost fully branching on the $(s-1)$-th layer. We can therefore obtain the following estimate

$$|A + 2^{2T} B'| \geq \frac{1}{4^m} 2^{1^-T|S|} \prod_{s+1 \notin \mathcal{S}} N_{sT,(s+1)T}(A).$$

Thus, if we write

$$|A + B'|/|A| = 2^{\delta N},$$

we know that

$$2^{\delta N} \geq \frac{1}{4^{m/2^{2T}}} \left(\prod_{s+1 \in \mathcal{S}} \frac{2^{1^-T}}{N_{sT,(s+1)T}(A)} \right)^{1/2^{2T}}.$$

Here we used the wrong version of the PR inequality.

Thus we have

$$\delta \geq O\left(\frac{1}{2^{2T}T}\right) + \frac{1}{2^{2T}}\frac{|S|1^-}{m} - \frac{1}{2^{2T}}\sum_{s+1\in\mathcal{S}}\frac{\log N_{sT,(s+1)T}(A)}{mT\log 2}.$$

We now have the following fake theorem.

THEOREM 6.3.1 (Bourgain's sumset estimate: based on the wrong version of PR inequality). Let $T > 1$ be an integer. Let $N = mT$ with $m > 1$ be an integer. Let $A, B \subset \mathbb{Z}2^{-N} \cap [0,1]$ with $|A| = 2^{\alpha N}, |B| = 2^{\beta N}$ and

$$c = 2^{\delta N}$$

such that $0 < \alpha, \beta < 1, |A+B| = c|A|$. Suppose that B has non-null layers $\mathcal{S}$. Then

$$\delta \geq O\left(\frac{1}{2^{2T}T}\right) + \frac{1}{2^{2T}}\frac{|S|1^-}{m} - \frac{1}{2^{2T}}\sum_{s+1\in\mathcal{S}}\frac{\log N_{sT,(s+1)T}(A)}{mT\log 2}.$$

This result is useful if the $O1/(T2^{2T})$ term is much smaller than the rest of the terms. This will require that $|S|$ grows at least linearly along with m. This is the case as

$$|S| \geq m(\beta - 2\log(T+1)/T).$$

The other requirement is that

$$\sum_{s+1\in\mathcal{S}}\frac{\log N_{sT,(s+1)T}(A)}{mT\log 2}$$

cannot be too large. This is a much weaker condition than the uniform branching property required in Theorem 6.3.1 in the sense that we do not need to require that all $T_A(s,T,x)$ are not too fully branching, instead, we only require that the averaged branching number

$$N_{sT,(s+1)T}(A) = \frac{\sum_{T_A(s,T,x)}|T_A(s,T,x)|}{\sum_{T_A(s,T,x)}1}$$

is not too large.

6.3.3. the super-exponential gap theorem. We make no delay in showing the following result with the fake Theorem 6.3.1.

Theorem 6.3.2:Hochman's super-exponential gap theorem

Let K_Λ be a self-similar system with a uniform scaling ratio. Then $\dim_{\mathrm{H}} K_\Lambda < \dim_s K_\Lambda$ only if there are super-exponential overlaps in Λ. Namely, for $N \to \infty$, $G_N = \min\{\Delta(\Lambda^N(\{0\}))\}$ where $\Delta(.)$ is the distance set satisfies

$$\frac{\log|G_N|}{N} \to \infty.$$

The first step is to understand the branching property of a self-similar set K. Assume that $K \subset [0,1]$. Let $\dim_{\mathrm{H}} K = \alpha$. For any $\alpha^- < \alpha$, for all large enough T, we have

$$|T_{K,T}| \geq 2^{\alpha^- T}.$$

This implies that for all $s \geq 1$,

$$N_{sT,(s+1)T}(K) \gg 2^{\alpha^- T}.$$

On the other hand, let $\alpha^+ > \alpha$. We consider the layers s so that

$$N_{sT,(s+1)T}(K) \geq 2^{\alpha^+ T}.$$

Denote this collection of layers to be $\mathcal{B}$ (bigger than expected layers). Then for all large enough m, we have

$$|T_{K,mT}| \geq 2^{\alpha^+ T|\mathcal{B}\cap[1,m]|} 2^{\alpha^- T|m-\mathcal{B}\cap[1,m]|}.$$

This implies that

$$\frac{\log |T_{K,mT}|}{mT \log 2} \geq \alpha^+ |\mathcal{B} \cap [1,m]| + \alpha^- |m - \mathcal{B} \cap [1,m]|.$$

Now that α^+, α^-n are fixed, there is some $\eta > 0$ such that

$$|\mathcal{B} \cap [1,m]| \leq \eta m$$

for all large enough m. Moreover, by choosing α^- to be sufficiently close to α and keeping α^+, we can achieve that η to be close to zero.

We can then use Theorem 6.3.1 to see that for each set B with the property that $|T_{B,mT}| \geq 2^{\beta mT}$ for some $\beta > 0$, we have

$$|T_{B+K,mT}|/|T_{K,mT}| \geq 2^{\delta mT}$$

with

$$\begin{aligned}
\delta &\geq O\left(\frac{1}{2^{2T}T}\right) + \frac{1}{2^{2T}} \frac{|S \cap [1,m]| 1^-}{m} - \frac{1}{2^{2T}} \sum_{s+1 \in \mathcal{S}\cap[1,m]} \frac{\log N_{sT,(s+1)T}(A)}{mT \log 2} \\
&\geq O\left(\frac{1}{2^{2T}T}\right) - \frac{1}{2^{2T}} \eta m \frac{\alpha^+ T \log 2}{mT \log 2} + \frac{1}{2^{2T}} \frac{|S \cap [1,m]| 1^-}{m} - \frac{1}{2^{2T}} \frac{\alpha^- T \log 2 |S \cap [1,m]|}{mT \log 2} \\
&= O\left(\frac{1}{2^{2T}T}\right) - \frac{1}{2^{2T}} \eta \alpha^+ + \frac{1}{2^{2T}} \frac{|S \cap [1,m]|(1^- - \alpha^-)}{m} \\
&\geq O\left(\frac{1}{2^{2T}T}\right) - \frac{1}{2^{2T}} \eta \alpha^+ + \frac{1}{2^{2T}} (1^- - \alpha^-)(\beta - 2\log(T+1)/T)
\end{aligned}$$

for all large enough m. Now, we make make the choice of α^+, α^- so that η is much smaller than β. Then we can choose a large enough T to achieve that

$$\delta > \rho > 0$$

for some number ρ. This number ρ depends on the initial data β. We now arrive at the following fake theorem.

THEOREM 6.3.2. Let $K \subset [0,1]$ be a self-similar set with $\dim_{\mathrm{H}} K = \alpha$. Let $B \subset [0,1]$ be a finite set with $|B| = 2^{\beta N}$ where N is a large integer. Then there is a number $T > 0$ and a number $\rho > 0$ such that for all large enough $m > 0$ and $N = mT$,

$$N_{2^{-N}}(B+K)/N_{2^{-N}}(K) \geq 2^{\rho N}.$$

We can now prove Theorem 6.3.2. We have already almost done this. The problem is the wrong estimate (6.1). We now resume the context in that estimate ($K = K_\Lambda$).

From what we proved above, we see that for all large enough D, for all $\beta > 0$, and for all large enough N (after the choice of D, β), there is a number $\rho > 0$ such that

$$N_{r^{DN}}(A + r^N K) \gg r^{-DN\rho} N_{r^{DN}}(r^N K)$$

unless

$$N_{r^{DN}}(A) \leq r^{-DN\beta}.$$

This is not quite proving (6.1). However, this is already sufficient for Theorem 6.3.2.

Recall that the set A encodes the overlapping structure of K at scale r^N around somewhere. Assume that we only have exponential gaps. This implies that for some certain choice of D, all such sets A are r^{DN}-separated. This says that

$$N_{r^{DN}}(A) = |A|.$$

Recall that $\dim_{\mathrm{H}} K = \alpha, \dim_s K = s$. Then the sum the multiplicities of of all possible overlaps is

$$\sum_A |A| \gg r^{-sN}.$$

Finally, we need a result that confirms that $|A|$ must be in some sense large. This will be formulated later as Theorem 6.4.13. In particular, we have the result that for each α^-, $|A| \geq r^{-\alpha^- N}/r^{-sN}$ for at least half of the possible A's.

We can choose β to be so small that $\beta < s - \alpha^-$. Then we conclude that, as we only have exponential gaps,

$$N_{r^{DN}}(A + r^N K) \gg r^{DN\rho} N_{r^{DN}}(r^N K)$$

for those A with $|A| \geq r^{-\alpha^- N}/r^{-sN}$. However, this will imply that

$$N_{r^{(D+1)N}}(K) \geq \sum_{A:|A| \geq r^{-\alpha^- N}/r^{-sN}} r^{-\rho DN} N_{r^{(D+1)N}}(r^N K) \gg N_{r^N}(K) r^{-\rho DN} N_{r^{(D+1)N}}(r^N K).$$

Next, observe that

$$N_{r^N}(K) = r^{-N(\alpha+o(1))}, N_{r^{(D+1)N}}(r^N K) = r^{-DN(\alpha+o(1))}, N_{r^{(D+1)N}}(K) = r^{-(D+1)(\alpha+o(1))}.$$

Thus we see that

$$\begin{aligned}(D+1)N(\alpha + o(1)) &\geq \rho DN + DN(\alpha + o(1)) + N(\alpha + o(1)) \\ &= (D+1)N(\alpha + o(1)) + \rho DN\end{aligned}$$

which implies that for $N \to \infty$

$$o(1) = \rho \frac{D}{D+1} > 0$$

This is not possible. From here, we proved Theorem 6.3.2 under the wrong version of PR-inequality.

6.4. Hochman's inverse entropy theorem

We now provide a real proof of Theorem 6.3.2. This proof works for dealing with self-similar measures rather than self-similar sets. From Bourgain's sumset estimate, we see that $|A+B| < c|A|$ with some small number c implies that A, B must have some structural constrain in many scales. Namely, considering the binary coding trees of A, B we observe that there is a number $T > 0$, so that for most (proportion roughly 1) of the T layers, either B is singly branching (dimension zero) or else many subtrees of T_A in this layer is almost fully branching (dimension one). We already illustrated that if the following wrong version of PR inequality would hold:

$$|A+B| = c|A| \to |A+nB| \leq c^n|A|$$

then we would prove the super-exponential gap theorem. Unfortunately, it is a difficult task to keep track of the branching properties of a coding tree in various layers. For example, for some layers, the coding tree may not have a uniform branching property, i.e. some of the subtrees branch more heavily than others. We still want to capture the notion of the 'typical' branching property for those subtrees. It turns out that the notion of entropy encodes this type of information.

Definition 6.4.1:Entropy

Let μ be a probability measure on $\mathbb{R}$. Let $N > 0$, the N-level entropy $H_N(\mu)$ is defined to be

$$H_N(\mu) = \sum_{D \in \mathcal{D}_N} \mu(D)|\log_2 \mu(D)|.$$

If μ is countably supported then define its Shannon entropy to be

$$H(\mu) = \sum_{x \in \text{supp}(\mu)} \mu(x)|\log \mu(x)|.$$

For μ being a probability measure, we often want to study it with certain scales (much like the case for sets). For $N > 1$, we can define a probability measure μ_N on $\mathbb{Z}2^{-N}$ that approximate μ at 2^{-N}-scale. The entropy stores some branching information for the support of μ.

Theorem 6.4.2:

$H_N(\mu) \in [0, N]$. If $H_N(\mu) = 0$ then $\text{supp}(\mu)$ is singly branched upto level N and if $H_N(\mu) = N$ then $\text{supp}(\mu)$ is uniformly and fully branched up to level N. Write $H_N(\mu) = sN$, then $N_{2^{-N}}(\text{supp}(\mu)) \geq 2^{sN}$.

Proof. This is clear from the definition of entropy and the fact that for any finite sum $\sum_{i \in \{1,\dots,n\}} p_i|\log_2 p_i|$ takes the maximum value $\log n$ when all p_i's are equal to $1/n$. □

It is in general not true that if $H_N(\mu)$ is small then $N_{2^{-N}}(\text{supp}(\mu))$ is small. It can happen that a significant mass of concentrated in one 2^{-N} interval, and all other intervals may receive almost zero mass. However, this is essentially the only way to achieve small entropy.

Theorem 6.4.3:

$H_N(\mu) \in [0, N]$. Write $H_N(\mu) = sN$, then there is at least one $D \in \mathcal{N}$ so that $\mu(D) \geq 2^{-sN}$.

PROOF. Let $\delta > 0$. If all $\mu(D) \leq \delta$, then $H_N(\mu) \geq |\log \delta|$. We choose that $\delta = 2^{-sN}$ and obtain that there are D so that

$$\mu(D) > 2^{-sN}.$$

$\square$

So if s is close to zero, then there must be some D with a significant mass concentration.

6.4.1. An entropy version of PR inequality. Let μ, ν be two probability measure on $[0, 1]$. For a large number N, we want to consider

$$H_N(\mu * \nu)$$

as a analogy of the consideration of $\mathrm{supp}(\mu) + \mathrm{supp}(\nu)$ up to scale 2^{-N}. We want to understand the situation when

$$H_N(\mu * \nu) - H_N(\mu) \leq \delta N$$

for some small number $\delta > 0$. Bourgain's sumset estimate gives us the intuition that the above happens only when for most of T layers (T being a fixed integer), the 'branching property' for μ, ν in this layer should be either almost null (almost zero entropy) or almost full (almost T entropy).

Following Bourgain's idea, we can study the convolution $\mu * \nu$ are various scales and accumulate the increased entropy. For this reason, we first consider the convolution on a single scale. The following result serves as a replacement of the PR inequality in additive combinatorics.

Theorem 6.4.4:Kaimanovich-Vershik

Let μ, ν be finitely supported measure on $[0, 1]$. Let

$$c = H(\mu * \nu) - H(\mu).$$

Then for each $k \geq 1$,

$$H(\mu * \nu^k) - H(\mu) \leq kc.$$

Remark 6.4.5:

This result is much cleaner than the PR inequality in the sense that for PR inequality, with $|A + B| = c|A|$, we can only conclude that $|kB| \leq c^k|A|$ rather than

$$|A + kB| \leq c^k|A|.$$

The reason behind this cleanness is that the entropy only observes bits (small intervals) with significant μ or ν measures while in additive combinatorics, one would have to take care of all points. The situation is much like the case when $H_N(\mu)$ is small for which there can be so many small intervals with a very small μ measure and in this case, we'd only care about the bits with large μ measures.

PROOF. This result is nicely proved in the probability theory language. For a discrete random variable X, we write $H(X)$ for

$$\sum_x P(\{X=x\})|\log P(\{X=x\})|.$$

For two discrete random variables, we write

$$H(X,Y)=\sum_{x,y} P(\{X=x,Y=y\})|\log P(\{X=x,Y=y\})|$$

and

$$\begin{aligned}
H(X|Y) &= H(X,Y)-H(Y)\\
&= \sum_{x,y} P(\{X=x,Y=y\})|\log P(\{X=x,Y=y\})| - \sum_y P(\{Y=y\})|\log P(\{Y=y\})|\\
&= \sum_y P(\{Y=y\})\sum_x \frac{P(\{X=x,Y=y\})}{P(\{Y=y\})}\left|\log\frac{P(\{X=x,Y=y\})}{P(\{Y=y\})}\right|\\
&= \sum_y P(\{Y=y\})H(P(X|Y=y)).
\end{aligned}$$

In this language, we can find independent random variables $X_0\sim\mu, Y_1\sim\nu, Y_2\sim\nu\dots Y_i\sim\nu\dots$. We define

$$X_k = X_0+Y_1+\cdots+Y_k.$$

Then observe that

$$H(Y_1|X_k)=H(Y_1|X_k,X_{k+1}).$$

This is because

$$\begin{aligned}
H(Y_1|X_k,X_{k+1}) &= H(Y_1,X_k,X_{k+1})-H(X_k,X_{k+1})\\
&= H(X_k,X_{k+1}|Y_1)+H(Y_1)-H(X_k+Y_{k+1}|X_k)-H(X_k)\\
&= H(Y_{k+1}+X_k|Y_1,X_k)+H(X_k|Y_1)+H(Y_1)-H(X_k+Y_{k+1}|X_k)-H(X_k)\\
&= H(Y_{k+1})+H(X_k|Y_1)+H(Y_1)-H(Y_{k+1})-H(X_k)\\
&= H(Y_{k+1})+H(X_k,Y_1)-H(Y_1)-H(X_k)\\
&= H(Y_1|X_k).
\end{aligned}$$

Then we see that $(H(Y_1,X_k|X_{k+1})\le H(Y_1|X_{k+1})+H(X_k|X_{k+1}))$

$$H(Y_1|X_k)=H(Y_1|X_k,X_{k+1})\le H(Y_1|X_{k+1}).$$

Now we relate the entropy of random variables to the entropy of probability measures that we were interested in. First, observe that $(H(X_k|Y_1)=H(X_{k-1}))$

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

From here we obtain that

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

Thus
$$H(\mu * \nu^{k+1}) - H(\mu * \nu^k)$$
forms a decreasing sequence and from here we obtain the result. □

6.4.2. Iterated convolutions. Following Bourgain's idea, we want to obtain a non-trivial entropy increase of $H_N(\nu*\mu) - H_N(\mu)$ by simply knowing that $H_N(\nu^k)$ is large for certain large k. To achieve this, we need to know that finite scale entropy cannot drop by taking convolution (Shannon entropy has the property).

Theorem 6.4.6:

$H_N(\nu * \mu) \geq H_N(\mu)$ if $\mathrm{supp}(\nu) \subset \mathbb{Z}2^{-N}$. In general, $H_N(\nu * \mu) \geq H_N(\mu) - O(1)$.

Proof. $\mu * \nu$ is the following averaged version of μ,
$$\mu * \nu = \int \mu(. - x) d\nu(x).$$
Then we see that ($x \to -x \log x$ is convex)
$$\begin{aligned} H_N(\mu * \nu) &= H_N\Big(\int \mu(. - x) d\nu(x)\Big) \\ &= \sum_{D \in \mathcal{D}_\mathcal{N}} \int \mu(D_n - x) d\nu(x) \left|\log \int \mu(D_n - x) d\nu(x)\right| \\ &\geq \int \sum_{D \in \mathcal{D}_\mathcal{N}} \mu(D_n - x) |\log \mu(D_n - x)| d\nu(x). \end{aligned}$$
Notice that if $x \in \mathbb{Z}2^{-N}$, then
$$\sum_{D \in \mathcal{D}_\mathcal{N}} \mu(D_n - x) |\log \mu(D_n - x)| = H_N(\mu).$$
This proves the first part. Second part omit. □

Next, we want to study the iterated convolution of ν^k for large k. Clearly, if ν is a Dirac mass, then ν^k is always a Dirac mass there is no entropy growth. This is essentially the only case for ν^k not to have entropy growth. If we interpret ν^k as the law of the sum of independent random variables $Y_1 + \cdots + Y_k$, then it is very natural to consider the central limit theorem. However, we want to examine the technical details hidden behind the CLT.

Consider the Fourier transform $\hat{\nu}$. It is a continuous function with $\hat{\nu}(0) = 1$. In our case, ν is always compactly supported, thus $\hat{\nu}$ is smooth. Suppose that ν is centred, i.e. its mean is zero. We see that
$$\frac{d^k}{d^k \xi} \hat{\nu}(0) = \int x^k d\nu(x).$$
In particular, $\sigma(\nu) = \frac{d^2}{d^2\xi}\hat{\nu}(0)$. Assuming that $\mathbb{E}_\nu(|x|^3) < \infty$, we see that
$$\hat{\nu}(\xi) = 1 - \frac{1}{2}\sigma\xi^2 + \frac{1}{6}\lambda\xi^3 + O(\xi^4).$$

This tells us that near 0, the Fourier transform of $\hat{\nu}$ is smaller than one in a quantitative way. Thus $\hat{\nu}^k$ decays to zero near $\xi = 0$. Consider the following scaled function

$$\hat{\nu}^k(\xi) = \left(1 - \frac{1}{2}\sigma\xi^2 + \frac{1}{3}\lambda\xi^3 + O(\xi^4)\right)^k.$$

Then we see that

$$\begin{aligned}\log \hat{\nu}^k(\xi) &= k \log\left(1 - \frac{1}{2}\sigma\xi^2 + \frac{1}{3}\lambda\xi^3 + O(\xi^4)\right)\\ &= k\left(-\frac{1}{2}\sigma\xi^2 + \frac{1}{3}\lambda\xi^3 + O(\xi^4)\right)\\ &= -\frac{1}{2}\sigma\xi^2 k + \frac{1}{3}\lambda\xi^3 k + O(\xi^4 k).\end{aligned}$$

Thus we see that

$$\hat{\nu}^k(\xi) = e^{-k\sigma\xi^2/2} e^{3^{-1}\lambda k\xi^3 + O(k\xi^4)}.$$

Given $|\xi| \leq k^{-1/2}$, we have

$$e^{3^{-1}\lambda k\xi^3 + O(k\xi^4)} = 1 + O(k^{-0.5}).$$

Let $J \subset \mathbb{R}$ be any interval of length $\gg \sqrt{k}$. We approximate J with a smooth bump function on it, still written as J. Then $\hat{J}$ is essentially supported on $[-|J|^{-1}, |J|^{-1}]$. (This is one version of the uncertainty principle in harmonic analysis.) Then we see that

$$\nu^k(J) = \int \hat{J}(-\xi)e^{-k\sigma\xi^2/2}d\xi + O(k^{-0.5}) = \gamma(J) + O(k^{-0.5})$$

where γ is a gaussian distribution with variance $k\sigma$. The above argument holds for a series of probability measures $\nu_1, \ldots, \nu_k$ with variance and third momentum σ_k, λ_k. Let $\nu_k = \nu_1 * \cdots * \nu_k$ and we see that

$$\log \hat{\nu}_k(\xi) = -\frac{\sum_i \sigma_i}{2}\xi^2 + O(\frac{\sum_i \lambda_i}{3}\xi^3).$$

Thus if λ_i are uniformly bounded and $\sum_i \sigma_i \gg \sigma k$ for some $\sigma > 0$, we conclude that ν_k agrees with a Gaussian distribution of variance $\gg k$ on intervals of length $\gg k^{1/2}$.

Theorem 6.4.7:Iterated convolution

Let $\nu_1, \ldots, \nu_k$ be probability measures with bounded three moments. Suppose that the variance $\sum_i \sigma_i \geq \sigma k$ for some $\sigma > 0$. Then for each interval of length $\gg k^{1/2}$,

$$|\nu_k(J) - \gamma(J)| \ll k^{-1/2}$$

for some Gaussian distribution of variance σk.

6.4.3. From branching number to conditional entropy. In Bourgain's sumset estimate, we used a multi-scale analysis and then accumulated all local dimension gains to obtain an overall dimension gain. In our entropy consideration, this idea is encoded in the notion of conditional entropy.

Definition 6.4.8:

Let μ be a probability measure on $\mathbb{R}$. For integers $t < s$, define

$$H_{t|s}(\mu) = \sum_{D \in \mathcal{D}_t} \mu(D) H_{s-t}(\mu^D)$$

where μ^D is $T_D(\frac{1}{\mu_{|D}}\mu_{|D} = \mu_D)$ where T_D is the linear map that sends D to $[0,1]$.

Thus if all μ^D distributed uniformly at level $s-t$, then $H_{t|s}(\mu) = t-s$. Observe that

$$H_s(\mu) - H_t(\mu) = H_{t|s}(\mu).$$

Thus $H_{t|s}(\mu)$ plays the role of the notion of branching number in our coding tree argument in proving Bourgain's sumset estimate.

Let us now examine the 'branching number' of iterated convolutions measures $\nu_1, \ldots, \nu_k$ so that $\sum_i \sigma(\nu_i) \gg \sigma k$ for some $\sigma > 0$. Given $T, \varepsilon > 0$, there is some $n > 0$, and for all large enough k,

$$H_T((\nu_1 * \cdots * \nu_k)^D) \geq (1-\varepsilon)T.$$

holds for all $D \in \mathcal{D}_{n - \log k^{1/2}}$ except for some D's whose total $\nu_1 * \cdots * \nu_k$ mass is at most ε. To see this, observe that we can approximate $\nu_1 * \cdots * \nu_k$ on intervals $\geq k^{1/2}/2^{T+n}$ by the measure given by a Gaussian distribution of variance $\gg \sigma k$.

6.4.4. An inverse entropy theorem. We now want to obtain an entropy version of Theorem 6.3.1. Before that, we will introduce the notion of 'singly' and 'fully' branching property in entropy.

Let μ be a probability measure on $\mathbb{R}$.

Definition 6.4.9:

Let $N > 1$. We say that μ is (N, ε)-uniform if

$$H_N(\mu) > (1-\varepsilon)N.$$

We say that μ is (N, ε)-atomic if

$$H_N(\mu) < \varepsilon N.$$

We also need to relate the variance and the entropy.

Theorem 6.4.10:

Let $N > 0$. Let $\sigma > 0$. There is an $\varepsilon_1 > 0$ so that as long as μ is (N, ε_1)-atomic, $\sigma(\mu) < \sigma$. On the other hand, let $\varepsilon_2 > 0$ there is a σ so that if $\sigma(\mu) < \sigma$, then μ is (N, ε_2)-atomic.

Remark 6.4.11:

Thus small entropy measures also have small variance and vice versa. The relations between $\sigma, \varepsilon_1, \varepsilon_2$ depend on the choice of N.

PROOF. Omit for now. □

Let μ, ν be probability measures on $[0,1]$. Consider the measure $\mu * \nu^k$. For integers N, m, T with $N = mT$, we have

$$H_N(\mu * \nu^k) = \sum_{s=1}^{m-1} H_{(s-1)T|sT}(\mu * \nu^k).$$

For each s, consider for $n, k > 1$,

$$\begin{aligned} &H_{(s-1)T|sT}(\mu * \nu^k) \\ &= H_{(s-1)T|sT}(\mathbb{E}_{x\sim\mu, y_1,\dots,y_k\sim\nu}\mu_{x,(s-1)T} * \nu_{y_1,(s-1)T} * \cdots * \nu_{y_k,(s-1)T}) \\ &\geq \mathbb{E}_{x\sim\mu, y_1,\dots,y_k\sim\nu} H_{(s-1)T|sT}(\mu_{x,(s-1)T} * \nu_{y_1,(s-1)T} * \cdots * \nu_{y_k,(s-1)T}) \\ &= \mathbb{E}_{x\sim\mu, y_1,\dots,y_k\sim\nu} H_T(\mu^{x,(s-1)T} * \nu^{y_1,(s-1)T} * \cdots * \nu^{y_k,(s-1)T}) \end{aligned}$$

where we used $\mu^{x,i}, \mu_{x,i}$ to denote μ^D, μ_D for $D \in \mathcal{D}_i$ which contains x.

We now make sure of the CLT discussion. Suppose that for some $k > 1$,

$$H_{(s-1)T+\log k^{1/2}|sT+\log k^{1/2}}(\nu) > \sigma T$$

for some number $\sigma > 0$. Then for $\gg 1$ portion of y measured by ν,

$$H_T(\nu^{y,(s-1)T+\log k^{1/2}}) > \sigma T/100$$

Then for all large enough k, the law of large numbers tells us that with a probability close to one,

$$\sum_{i=1}^{k} \sigma(\nu^{y_i,(s-1)T+\log k^{1/2}}) > \frac{\sigma' T k}{100}$$

for some $\sigma' > 0$ depending on σ, T. Then we see that $\nu^{y_1,(s-1)T+\log k^{1/2}} * \cdots * \nu^{y_k,(s-1)T+\log k^{1/2}}$ is basically a Gaussian distribution with variance $\gg \sigma' T k$. The approximate holds at scale $\geq k^{1/2}/2^T$. We can then see that

$$\nu^{y_1,(s-1)T} * \cdots * \nu^{y_k,(s-1)T}$$

can be approximated by a (roughly normalised) Gaussian distribution on scale $\gg 1/2^T$. For each $\varepsilon_0, \sigma_0 > 0$, we can choose T so that

$$H_T(\gamma) > (1-\varepsilon_0)T$$

for all Gaussian distribution with $\sigma(\gamma) > \sigma_0'$, where σ_0, σ_0' are related as indicated in Theorem 6.4.10. Suppose that $\sigma' > \sigma_0'$, then we can choose a large enough $k > 0$[2] and achieve

$$H_T(\nu^{y_1,(s-1)T} * \cdots * \nu^{y_k,(s-1)T}) > (1-\varepsilon_0)T.$$

If this is not the case, then

$$H_{(s-1)T+\log k^{1/2}|sT+\log k^{1/2}}(\nu) < \sigma T$$

and for most portions of y measured by ν,

$$H_T(\nu^{y,(s-1)T+\log k^{1/2}}) \leq \sigma T.$$

Now our aim is to study $\mu * \nu$. If s is such that

$$H_{(s-1)T+\log k^{1/2}|sT+\log k^{1/2}}(\nu) > \sigma T,$$

[2] It is useful to record the dependencies of the parameters. We first choose $\varepsilon_0 > 0, \sigma_0 > 0$ be small numbers. Then we find an integer T and an integer k.

then $H_{(s-1)T|sT}(\mu * \nu^k) > (1-\varepsilon_0)T + O(1)$. On the other hand, we get

$$H_{(s-1)T|sT}(\mu * \nu^k) \geq H_{(s-1)T|sT}(\mu) + O(1).$$

Then we see that

$$H_N(\mu * \nu^k) = \sum_{s=1}^{m-1} H_{(s-1)T|sT}(\mu * \nu^k) \geq \sum_{s \in \mathcal{S}} (1-\varepsilon_0)T + \sum_{s \notin \mathcal{S}} H_{(s-1)T|sT}(\mu)$$

where $\mathcal{S}$ is the collection of s with

$$H_{(s-1)T+\log k^{1/2}|sT+\log k^{1/2}}(\nu) > \sigma T.$$

Thus $\mathcal{S}$ is the '(shifted) non-singly branching layers' for ν. To obtain a result in the form of Theorem 6.3.1, we can refine the collection $\mathcal{S}$

$$S' = \{s \in \mathcal{S} : H_{(s-1)T|sT}(\mu) < (1-2\varepsilon_0)T\}.$$

Then we have

$$H_N(\mu * \nu^k) \geq \#\mathcal{S}'\varepsilon_0 T + H_N(\mu) + O(m).$$

Use Kaimanovich-Vershik we see that

$$H_N(\mu * \nu) - H_N(\mu) \geq \frac{\#\mathcal{S}'\varepsilon_0 T}{k} + O(m/k).$$

Thus we see that

$$\frac{1}{N}(H_N(\mu * \nu) - H_N(\mu)) \geq \frac{\#\mathcal{S}'\varepsilon_0}{mk} + O(\frac{1}{kT}).$$

Theorem 6.4.12:Hochman's inverse entropy theorem

Let μ, ν be probability measures on $[0,1]$. Let ε_0, σ_0 be given. Then for a large enough T, for a large enough k depending on k, for all large enough $N = mT$,

$$\frac{1}{N}(H_N(\mu * \nu) - H_N(\mu)) \geq \frac{\#\mathcal{S}'\varepsilon_0}{km} + O(1/(kT)).$$

To obtain reasonable applications, we need $\#\mathcal{S}'$ to scale with $\gg m$ for otherwise, this theorem tells nothing new. This is the case if μ is 'mostly not fully branching' in the sense that for all $\varepsilon > 0$, for all large enough $N = mT$,

$$\mathcal{F} = \mathcal{F}_N(\mu) = \{H_{(s-1)T|sT}(\mu) > (1-\varepsilon)T\} \leq \varepsilon m.$$

This condition is much weaker than the corresponding uniform branching property in Bourgain's theorem in this sense we are now allowed to drop some bits that are not significant w.r.t the measure μ.

If μ is mostly not fully branching, then

$$\mathcal{S}' = \mathcal{F}^c \cap \mathcal{S}.$$

Now $\mathcal{F}$ occupies a proportion $\leq \varepsilon$ many layers. So that if $\mathcal{S}'$ is smaller than εm, it implies that

$$\#\mathcal{S} \leq 2\varepsilon m.$$

Thus ν does not have many layers with large branching numbers (i.e. large conditional entropy). This implies that

$$H_N(\nu) \leq \#\mathcal{S}T + \sigma_0 N \leq (\sigma_0 + 2\varepsilon)N.$$

Thus, by choosing σ_0 small at the beginning and choosing ε small, this will imply that ν is a very singular measure (with small entropy).

6.4.5. The entropy structure of self-similar measures. Let μ be a self-similar measure. Then there is more structural information. We first have the following fact,

Theorem 6.4.13:Exact dimensionality

Let μ be a non-trivial self-similar measure on $\mathbb{R}$. Then it is exact dimensional. This means that for μ.a.e x,

$$\lim_{r\to 0}\frac{\log\mu(B_r(x))}{\log r}=s$$

for some $s\in(0,1]$. Moreover, for the same s,

$$\lim_{N\to\infty}\frac{1}{N}H_N(\mu)=s.$$

This s is equal to the Hausdorff dimension of the support of μ if it is the canonical self-similar measure (i.e. with probability weight chosen to be according to $\sum_i r_i^s=1$).

PROOF. omit for now □

Given this result, consider

$$H_{(s-1)T|sT}(\mu)\geq\mathbb{E}_{x\sim\mu}H_T(\mu^{x,(s-1)T}).$$

Observe that $H_T(\mu^{x,(s-1)T})$ is basically $H_T(\mu*\nu)$ for some discrete measure ν because of the self-similarity of μ. Here ν encodes the overlaps of μ around a chosen x. Then we see that

$$H_{(s-1)T|sT}(\mu)\geq H_T(\mu)+O(1).$$

If T is large enough, we can get

$$H_T(\mu)>s^-.$$

Let $s^+>s$ be given. Suppose for large $N=mT$, in at least ε-portion of the layers,

$$H_{(s-1)T|sT}(\mu)\geq s^+T,$$

then we see that

$$H_N(\mu)\geq s^-(1-\varepsilon)N+s^+\varepsilon N.$$

Thus, for any given s^+,ε there is a large T (to make s^- close to s) so that for all large $N=mT$,

$$\#\{H_{(s-1)T|sT}(\mu)>s^+T\}\leq\varepsilon m.$$

This implies that μ mostly not fully branching provided that $s<1$. We thus obtained the following result which was 'proved' under a wrong assumption. With the help of the notion of entropy, we can now rescue the proof, again.

Assuming that there are no super exponentially decaying gaps. Assume that $\dim_{\mathrm{H}}K=\alpha<s$. We choose μ to be the canonical self-similar measure. Then we see that for $N>1,D>1$,

$$H_{(D+1)N}(\mu)\geq H_N(\mu)+\mathbb{E}_{x\sim\mu}H_{DN}(\mu^{x,N}).$$

As was discussed above,

$$\mu^{x,N} \approx \mu * \nu_x$$

for some discrete measure ν that encodes the overlaps. We use $\approx$ to denote that this is not equality in general. omit details As μ is mostly not fully branching, we see that $H_{DN}(\mu * \nu_x) > H_{DN}(\mu) + DN\delta$ for some $\delta > 0$ as long as $H_{DN}(\nu_x) > DN\varepsilon$ for some $\varepsilon > 0$. Because of the exact dimensionality, we see that for most of $\mu^{x,N}$, we have

$$|\mu^{x,N}| \gg 2^{-\alpha^+ N}.$$

Since $\alpha < s$, we choose $\alpha^+ < s$. Then we see that the number of overlaps at x must be

$$\gg 2^{-\alpha^+ N}/2^{-sN} = 2^{(s-\alpha^+)N}.$$

However, if on the other hand for some $> \varepsilon$-portion of $x \sim \mu$, $H_{DN}(\nu_x) > \varepsilon$, then

$$H_{(D+1)N}(\mu) - H_N(\mu) \geq \mathbb{E}_{x\sim\mu} H_{DN}(\mu^{x,N}) \geq H_{DN}(\mu) + DN\varepsilon\delta + O(1).$$

Then we see that

$$\frac{1}{(D+1)N} H_{(D+1)N}(\mu) - \frac{1}{(D+1)N} H_N(\mu) \geq \frac{1}{(D+1)N} \frac{DN}{DN} H_{DN}(\mu) + \frac{D}{D+1}\varepsilon\delta + o(1).$$

Taking the limit for $N \to \infty$, we see that

$$\alpha - \frac{\alpha}{D+1} \geq \frac{D}{D+1}\alpha + \frac{D}{D+1}\varepsilon\delta.$$

This is not possible. Therefore, there for most of $x \sim \mu$, we have

$$H_{DN}(\nu_x) \leq DN\varepsilon.$$

Therefore, not all overlaps around x can be separated by intervals of length $2^{-(D+1)N}$ for otherwise $H_{DN}(\nu_x)$ would be $\geq (s - \alpha^+)N$.

From here, we conclude that for each D, there are N-level overlaps with distance $\leq 2^{-DN}$. Since D can be arbitrary, we see that there are super-exponential gaps.

6.5. Shmerkin's inverse entropy theorem

Bibliography

[1] C. Bishop and Y. Peres, *Fractals in probability and analysis*, Cambridge University Press, Cambridge, 2017. vii

[2] T. Bodineau, I. Gallagher and L. Saint-Raymond, *The Brownian motion as the limit of a deterministic system of hard-spheres*, Invent. math. 203, 493,Äì553 (2016). 6

[3] J. Bourgain,*The discretized sum-product and projection theorems.*, JAMA 112, 193-236 (2010). 81

[4] K. Falconer, *Fractal geometry: mathematical foundations and applications*, John Wiley & Sons Inc., Hoboken, 2014. vii

[5] M. Hochman, *On self-similar sets with overlaps and inverse theorems for entropy*, Annals of Mathematics 180 (2014), 773-822. 81

[6] R. Kaufman, *On the theorem of Jarník and Besicovitch*, Acta Arithmetica, 39.3, 265-267 (1981). 18

[7] P. Mattila, *Geometry of sets and measures in Euclidean spaces: fractals and rectifiability*, Cambridge University Press, Cambridge, 1995. vii

[8] I. Ruzsa, *Sumsets and structure*, https://www.math.cmu.edu/users/af1p/Teaching/AdditiveCombinatorics/Additive-Combinatorics.pdf 84

[9] P. Shmerkin, *On Furstenberg's intersection conjecture, self-similar measures, and the Lq norms of convolutions.*, Ann. of Math.(2), 189(2), 319-391. 81

[10] P. Shmerkin and H. Yu, *On Sets Containing a Unit Distance in Every Direction*, Discrete Analysis,https://doi.org/10.19086/da.22058. 79