7 Hausdorff Measure and Fractals

Carathéodory developed a remarkably simple generalization of Lebesgue's measure theory which in particular allowed him to define the p-dimensional measure of a set in q-dimensional space. In what follows, I present a small addition.... a clarification of p-dimensional measure that leads immediately to an extension to non-integral p, and thus gives rise to sets of fractional dimension.

F. Hausdorff, 1919

I coined *fractal* from the Latin adjective *fractus.* The corresponding Latin verb *frangere* means to "break": to create irregular fragments.

B. Mandelbrot, 1977

The deeper study of the geometric properties of sets often requires an analysis of their extent or "mass" that goes beyond what can be expressed in terms of Lebesgue measure. It is here that the notions of the dimension of a set (which can be fractional) and an associated measure play a crucial role.

Two initial ideas may help to provide an intuitive grasp of the concept of the dimension of a set. The first can be understood in terms of how the set replicates under scalings. Given the set E, let us suppose that for some positive number n we have that $nE = E_1 \cup \cdots \cup E_m$, where the sets E_j are m essentially disjoint congruent copies of E. Note that if E were a line segment this would hold with $m = n$; if E were a square, we would have $m = n^2$; if E were a cube, then $m = n^3$; etc. Thus, more generally, we might be tempted to say that E has dimension α if $m = n^\alpha$. Observe that if E is the Cantor set $\mathcal{C}$ in $[0, 1]$, then $3\mathcal{C}$ consists of 2 copies of $\mathcal{C}$, one in $[0, 1]$ and the other in $[2, 3]$. Here $n = 3$, $m = 2$, and we would be led to conclude that $\log 2/ \log 3$ is the dimension of the Cantor set.

Another approach is relevant for curves that are not necessarily rectifiable. Start with a curve $\Gamma = \{\gamma(t) : a \leq t \leq b\}$, and for each $\epsilon > 0$ consider polygonal lines joining $\gamma(a)$ to $\gamma(b)$, whose vertices lie on successive points of Γ, with each segment not exceeding ϵ in length. Denote by $\#(\epsilon)$ the least number of segments that arise for such polygonal lines. If $\#(\epsilon) \approx \epsilon^{-1}$ as $\epsilon \to 0$, then Γ is rectifiable. However, $\#(\epsilon)$ may well grow more rapidly than ϵ^{-1} as $\epsilon \to 0$. If we had $\#(\epsilon) \approx \epsilon^{-\alpha}$, $1 < \alpha$, then, in the spirit of the previous example, it would be natural to say that Γ has dimension α. These considerations have even an interest in other parts of science. For instance, in studying the question of determining the length of the border of a country or its coastline, L.F. Richardson found that the length of the west coast of Britain obeyed the empirical law $\#(\epsilon) \approx \epsilon^{-\alpha}$, with α approximately 1.5. Thus one might conclude that the coast has fractional dimension!

While there are a number of different ways to make some of these heuristic notions precise, the theory that has the widest scope and greatest flexibility is the one involving Hausdorff measure and Hausdorff dimension. Probably the most elegant and simplest illustration of this theory can be seen in terms of its application to a general class of self-similar sets, and this is what we consider first. Among these are the curves of von Koch type, and these can have any dimension between 1 and 2.

Next, we turn to an example of a space-filling curve, which, broadly speaking, falls under the scope of self-replicating constructions. Not only does this curve have an intrinsic interest, but its nature reveals the important fact that from the point of view of measure theory the unit interval and the unit square are the same.

Our final topic is of a somewhat different nature. It begins with the realization of an unexpected regularity that all subsets of $\mathbb{R}^d$ (of finite Lebesgue measure) enjoy, when $d \geq 3$. This property fails in two dimensions, and the key counter-example is the Besicovitch set. This set appears also in a number of other problems. While it has measure zero, this is barely so, since its Hausdorff dimension is necessarily 2.

1 Hausdorff measure

The theory begins with the introduction of a new notion of volume or mass. This "measure" is closely tied with the idea of dimension which prevails throughout the subject. More precisely, following Hausdorff, one considers for each appropriate set E and each $\alpha > 0$ the quantity $m_\alpha(E)$, which can be interpreted as the α-dimensional mass of E among sets of dimension α, where the word "dimension" carries (for now) only an intuitive meaning. Then, if α is larger than the dimension of the set E, the set has a negligible mass, and we have $m_\alpha(E) = 0$. If α is smaller than the dimension of E, then E is very large (comparatively), hence $m_\alpha(E) = \infty$. For the critical case when α is the dimension of E, the quantity $m_\alpha(E)$ describes the actual α-dimensional size of the set.

Two examples, to which we shall return in more detail later, illustrate this circle of ideas.

First, recall that the standard Cantor set $\mathcal{C}$ in $[0,1]$ has zero Lebesgue measure. This statement expresses the fact that $\mathcal{C}$ has one-dimensional mass or length equal to zero. However, we shall prove that $\mathcal{C}$ has a well-defined fractional Hausdorff dimension of $\log 2/\log 3$, and that the corresponding Hausdorff measure of the Cantor set is positive and finite.

Another illustration of the theory developed below consists of starting with Γ, a rectifiable curve in the plane. Then Γ has zero two-dimensional Lebesgue measure. This is intuitively clear, since Γ is a one-dimensional object in a two-dimensional space. This is where the Hausdorff measure comes into play: the quantity $m_1(\Gamma)$ is not only finite, but precisely equal to the length of Γ as we defined it in Section 3.1 of Chapter 3.

We first consider the relevant exterior measure, defined in terms of coverings, whose restriction to the Borel sets is the desired Hausdorff measure.

For any subset E of $\mathbb{R}^d$, we define the **exterior α-dimensional Hausdorff measure** of E by

$$m_\alpha^*(E) = \lim_{\delta \to 0} \inf \left\{ \sum_k (\text{diam } F_k)^\alpha : \ E \subset \bigcup_{k=1}^\infty F_k, \quad \text{diam } F_k \leq \delta \text{ all } k \right\},$$

where diam S denotes the diameter of the set S, that is, $\text{diam } S = \sup\{|x-y| : \ x, y \in S\}$. In other words, for each $\delta > 0$ we consider covers of E by countable families of (arbitrary) sets with diameter less than δ, and take the infimum of the sum $\sum_k (\text{diam } F_k)^\alpha$. We then define $m_\alpha^*(E)$ as the limit of these infimums as δ tends to 0. We note that the quantity

$$\mathcal{H}_\alpha^\delta(E) = \inf \left\{ \sum_k (\text{diam } F_k)^\alpha : \ E \subset \bigcup_{k=1}^\infty F_k, \quad \text{diam } F_k \leq \delta \text{ all } k \right\}$$

is increasing as δ decreases, so that the limit

$$m_\alpha^*(E) = \lim_{\delta \to 0} \mathcal{H}_\alpha^\delta(E)$$

exists, although $m_\alpha^*(E)$ could be infinite. We note that in particular, one has $\mathcal{H}_\alpha^\delta(E) \leq m_\alpha^*(E)$ for all $\delta > 0$. When defining the exterior measure $m_\alpha^*(E)$ it is important to require that the coverings be of sets of arbitrarily small diameters; this is the thrust of the definition $m_\alpha^*(E) = \lim_{\delta\to 0}\mathcal{H}_\alpha^\delta(E)$. This requirement, which is not relevant for Lebesgue measure, is needed to ensure the basic additive feature stated in Property 3 below. (See also Exercise 12.)

Scaling is the key notion that appears at the heart of the definition of the exterior Hausdorff measure. Loosely speaking, the measure of a set scales according to its dimension. For instance, if Γ is a one-dimensional subset of $\mathbb{R}^d$, say a smooth curve of length L, then $r\Gamma$ has total length rL. If Q is a cube in $\mathbb{R}^d$, the volume of rQ is $r^d|Q|$. This feature is captured in the definition of exterior Hausdorff measure by the fact that if the set F is scaled by r, then (diam $F)^\alpha$ scales by r^α. This key idea reappears in the study of self-similar sets in Section 2.2.

We begin with a list of properties satisfied by the Hausdorff exterior measure.

Property 1 (Monotonicity) *If $E_1 \subset E_2$, then $m_\alpha^*(E_1) \leq m_\alpha^*(E_2)$.*

This is straightforward, since any cover of E_2 is also a cover of E_1.

Property 2 (Sub-additivity) $m_\alpha^*(\bigcup_{j=1}^\infty E_j) \leq \sum_{j=1}^\infty m_\alpha^*(E_j)$ *for any countable family $\{E_j\}$ of sets in $\mathbb{R}^d$.*

For the proof, fix δ, and choose for each j a cover $\{F_{j,k}\}_{k=1}^\infty$ of E_j by sets of diameter less than δ such that $\sum_k(\text{diam } F_{j,k})^\alpha \leq \mathcal{H}_\alpha^\delta(E_j) + \epsilon/2^j$. Since $\bigcup_{j,k} F_{j,k}$ is a cover of E by sets of diameter less than δ, we must have

$$\begin{aligned}\mathcal{H}_\alpha^\delta(E) &\leq \sum_{j=1}^\infty \mathcal{H}_\alpha^\delta(E_j) + \epsilon \\ &\leq \sum_{j=1}^\infty m_\alpha^*(E_j) + \epsilon.\end{aligned}$$

Since ϵ is arbitrary, the inequality $\mathcal{H}_\alpha^\delta(E) \leq \sum m_\alpha^*(E_j)$ holds, and we let δ tend to 0 to prove the countable sub-additivity of m_α^*.

Property 3 *If $d(E_1, E_2) > 0$, then $m_\alpha^*(E_1 \cup E_2) = m_\alpha^*(E_1) + m_\alpha^*(E_2)$.*

It suffices to prove that $m_\alpha^*(E_1 \cup E_2) \geq m_\alpha^*(E_1) + m_\alpha^*(E_2)$ since the reverse inequality is guaranteed by sub-additivity. Fix $\epsilon > 0$ with $\epsilon < d(E_1, E_2)$. Given any cover of $E_1 \cup E_2$ with sets $F_1, F_2 \ldots$, of diameter less than δ, where $\delta < \epsilon$, we let

$$F_j' = E_1 \cap F_j \quad \text{and} \quad F_j'' = E_2 \cap F_j.$$

Then $\{F'_j\}$ and $\{F''_j\}$ are covers for E_1 and E_2, respectively, and are disjoint. Hence,

$$\sum_j (\text{diam } F'_j)^\alpha + \sum_i (\text{diam } F''_i)^\alpha \leq \sum_k (\text{diam } F_k)^\alpha.$$

Taking the infimum over the coverings, and then letting δ tend to zero yields the desired inequality.

At this point, we note that m^*_α satisfies all the properties of a metric Carathéodory exterior measure as discussed in Chapter 6. Thus m^*_α is a countably additive measure when restricted to the Borel sets. We shall therefore restrict ourselves to Borel sets and write $m_\alpha(E)$ instead of $m^*_\alpha(E)$. The measure m_α is called the **α-dimensional Hausdorff measure**.

Property 4 *If $\{E_j\}$ is a countable family of disjoint Borel sets, and $E = \bigcup_{j=1}^\infty E_j$, then*

$$m_\alpha(E) = \sum_{j=1}^\infty m_\alpha(E_j).$$

For what follows in this chapter, the full additivity in the above property is not needed, and we can manage with a weaker form whose proof is elementary and not dependent on the developments of Chapter 6. (See Exercise 2.)

Property 5 *Hausdorff measure is invariant under translations*

$$m_\alpha(E + h) = m_\alpha(E) \quad \textit{for all } h \in \mathbb{R}^d,$$

and rotations

$$m_\alpha(rE) = m_\alpha(E),$$

where r is a rotation in $\mathbb{R}^d$.

Moreover, it scales as follows:

$$m_\alpha(\lambda E) = \lambda^\alpha m_\alpha(E) \quad \textit{for all } \lambda > 0.$$

These conclusions follow once we observe that the diameter of a set S is invariant under translations and rotations, and satisfies $\text{diam}(\lambda S) = \lambda \,\text{diam}(S)$ for $\lambda > 0$.

We describe next a series of properties of Hausdorff measure, the first of which is immediate from the definitions.

Property 6 *The quantity $m_0(E)$ counts the number of points in E, while $m_1(E) = m(E)$ for all Borel sets $E \subset \mathbb{R}$. (Here m denotes the Lebesgue measure on $\mathbb{R}$.)*

In fact, note that in one dimension every set of diameter δ is contained in an interval of length δ (and for an interval its length equals its Lebesgue measure).

In general, d-dimensional Hausdorff measure in $\mathbb{R}^d$ is, up to a constant factor, equal to Lebesgue measure.

Property 7 *If E is a Borel subset of $\mathbb{R}^d$, then $c_d m_d(E) = m(E)$ for some constant c_d that depends only on the dimension d.*

The constant c_d equals $m(B)/(\text{diam } B)^d$, for the unit ball B; note that this ratio is the same for all balls B in $\mathbb{R}^d$, and so $c_d = v_d/2^d$ (where v_d denotes the volume of the unit ball). The proof of this property relies on the so-called iso-diametric inequality, which states that among all sets of a given diameter, the ball has largest volume. (See Problem 2.) Without using this geometric fact one can prove the following substitute.

Property 7′ *If E is a Borel subset of $\mathbb{R}^d$ and $m(E)$ is its Lebesgue measure, then $m_d(E) \approx m(E)$, in the sense that*

$$c_d m_d(E) \leq m(E) \leq 2^d c_d m_d(E).$$

Using Exercise 26 in Chapter 3 we can find for every $\epsilon, \delta > 0$, a covering of E by balls $\{B_j\}$, such that diam $B_j < \delta$, while $\sum_j m(B_j) \leq m(E) + \epsilon$. Now,

$$\mathcal{H}_d^\delta(E) \leq \sum_j (\text{diam } B_j)^d = c_d^{-1} \sum_j m(B_j) \leq c_d^{-1}(m(E) + \epsilon).$$

Letting δ and ϵ tend to 0, we get $m_d(E) \leq c_d^{-1} m(E)$. For the reverse direction, let $E \subset \bigcup_j F_j$ be a covering with $\sum_j (\text{diam } F_j)^d \leq m_d(E) + \epsilon$. We can always find closed balls B_j centered at a point of F_j so that $B_j \supset F_j$ and diam $B_j = 2\,\text{diam } F_j$. However, $m(E) \leq \sum_j m(B_j)$, since $E \subset \bigcup_j B_j$, and the last sum equals

$$\sum c_d (\text{diam } B_j)^d = 2^d c_d \sum (\text{diam } F_j)^d \leq 2^d c_d \left(m_d(E) + \epsilon\right).$$

Letting $\epsilon \to 0$ gives $m(E) \leq 2^d c_d m_d(E)$.

Property 8 *If $m_\alpha^*(E) < \infty$ and $\beta > \alpha$, then $m_\beta^*(E) = 0$. Also, if $m_\alpha^*(E) > 0$ and $\beta < \alpha$, then $m_\beta^*(E) = \infty$.*

Indeed, if diam $F \leq \delta$, and $\beta > \alpha$, then

$$(\text{diam } F)^\beta = (\text{diam } F)^{\beta-\alpha}(\text{diam } F)^\alpha \leq \delta^{\beta-\alpha}(\text{diam } F)^\alpha.$$

Consequently

$$\mathcal{H}_\beta^\delta(E) \leq \delta^{\beta-\alpha}\mathcal{H}_\alpha^\delta(E) \leq \delta^{\beta-\alpha} m_\alpha^*(E).$$

Since $m_\alpha^*(E) < \infty$ and $\beta - \alpha > 0$, we find in the limit as δ tends to 0, that $m_\beta^*(E) = 0$.

The contrapositive gives $m_\beta^*(E) = \infty$ whenever $m_\alpha^*(E) > 0$ and $\beta < \alpha$.

We now make some easy observations that are consequences of the above properties.

1. If I is a finite line segment in $\mathbb{R}^d$, then $0 < m_1(I) < \infty$.
2. More generally, if Q is a k-cube in $\mathbb{R}^d$ (that is, Q is the product of k non-trivial intervals and $d - k$ points), then $0 < m_k(Q) < \infty$.
3. If $\mathcal{O}$ is a non-empty open set in $\mathbb{R}^d$, then $m_\alpha(\mathcal{O}) = \infty$ whenever $\alpha < d$. Indeed, this follows because $m_d(\mathcal{O}) > 0$.
4. Note that we can always take $\alpha \leq d$. This is because when $\alpha > d$, m_α vanishes on every ball, and hence on all of $\mathbb{R}^d$.

2 Hausdorff dimension

Given a Borel subset E of $\mathbb{R}^d$, we deduce from Property 8 that there exists a unique α such that

$$m_\beta(E) = \begin{cases} \infty & \text{if } \beta < \alpha, \\ 0 & \text{if } \alpha < \beta. \end{cases}$$

In other words, α is given by

$$\alpha = \sup\{\beta : m_\beta(E) = \infty\} = \inf\{\beta : m_\beta(E) = 0\}.$$

We say that E has **Hausdorff dimension** α, or more succinctly, that E has dimension α. We shall write $\alpha = \dim E$. At the critical value α we can say no more than that in general the quantity $m_\alpha(E)$ satisfies $0 \leq m_\alpha(E) \leq \infty$. If E is bounded and the inequalities are strict, that is, $0 < m_\alpha(E) < \infty$, we say that E has **strict Hausdorff dimension** α. The term **fractal** is commonly applied to sets of fractional dimension.

In general, calculating the Hausdorff measure of a set is a difficult problem. However, it is possible in some cases to bound this measure from above and below, and hence determine the dimension of the set in question. A few examples will illustrate these new concepts.

2.1 Examples

The Cantor set

The first striking example consists of the Cantor set $\mathcal{C}$, which was constructed in Chapter 1 by successively removing the middle-third intervals in $[0,1]$.

Theorem 2.1 *The Cantor set $\mathcal{C}$ has strict Hausdorff dimension $\alpha = \log 2/\log 3$.*

The inequality

$$m_\alpha(\mathcal{C}) \leq 1$$

follows from the construction of $\mathcal{C}$ and the definitions. Indeed, recall from Chapter 1 that $\mathcal{C} = \bigcap C_k$, where each C_k is a finite union of 2^k intervals of length 3^{-k}. Given $\delta > 0$, we first choose K so large that $3^{-K} < \delta$. Since the set C_K covers $\mathcal{C}$ and consists of 2^K intervals of diameter $3^{-K} < \delta$, we must have

$$\mathcal{H}_\alpha^\delta(\mathcal{C}) \leq 2^K(3^{-K})^\alpha.$$

However, α satisfies precisely $3^\alpha = 2$, hence $2^K(3^{-K})^\alpha = 1$, and therefore $m_\alpha(\mathcal{C}) \leq 1$.

The reverse inequality, which consists of proving that $0 < m_\alpha(\mathcal{C})$, requires a further idea. Here we rely on the Cantor-Lebesgue function, which maps $\mathcal{C}$ *surjectively* onto $[0,1]$. The key fact we shall use about this function is that it satisfies a precise continuity condition that reflects the dimension of the Cantor set.

A function f defined on a subset E of $\mathbb{R}^d$ satisfies a **Lipschitz condition** on E if there exists $M > 0$ such that

$$|f(x) - f(y)| \leq M|x - y| \quad \text{for all } x, y \in E.$$

More generally, a function f satisfies a **Lipschitz condition with exponent** γ (or is **Hölder** γ) if

$$|f(x) - f(y)| \leq M|x - y|^\gamma \quad \text{for all } x, y \in E.$$

The only interesting case is when $0 < \gamma \leq 1$. (See Exercise 3.)

Lemma 2.2 *Suppose a function f defined on a compact set E satisfies a Lipschitz condition with exponent γ. Then*

(i) *$m_\beta(f(E)) \leq M^\beta m_\alpha(E)$ if $\beta = \alpha/\gamma$.*

(ii) $\dim f(E) \le \frac{1}{\gamma} \dim E$.

Proof. Suppose $\{F_k\}$ is a countable family of sets that covers E. Then $\{f(E \cap F_k)\}$ covers $f(E)$ and, moreover, $f(E \cap F_k)$ has diameter less than $M(\text{diam } F_k)^\gamma$. Hence

$$\sum_k (\text{diam } f(E \cap F_k))^{\alpha/\gamma} \le M^{\alpha/\gamma} \sum_k (\text{diam } F_k)^\alpha,$$

and part (i) follows. This result now immediately implies conclusion (ii).

Lemma 2.3 *The Cantor-Lebesgue function F on $\mathcal{C}$ satisfies a Lipschitz condition with exponent $\gamma = \log 2/\log 3$.*

Proof. The function F was constructed in Section 3.1 of Chapter 3 as the limit of a sequence $\{F_n\}$ of piecewise linear functions. The function F_n increases by at most 2^{-n} on each interval of length 3^{-n}. So the slope of F_n is always bounded by $(3/2)^n$, and hence

$$|F_n(x) - F_n(y)| \le \left(\frac{3}{2}\right)^n |x - y|.$$

Moreover, the approximating sequence also satisfies $|F(x) - F_n(x)| \le 1/2^n$. These two estimates together with an application of the triangle inequality give

$$\begin{aligned} |F(x) - F(y)| &\le |F_n(x) - F_n(y)| + |F(x) - F_n(x)| + |F(y) - F_n(y)| \\ &\le \left(\frac{3}{2}\right)^n |x - y| + \frac{2}{2^n}. \end{aligned}$$

Having fixed x and y, we then minimize the right hand side by choosing n so that both terms have the same order of magnitude. This is achieved by taking n so that $3^n|x - y|$ is between 1 and 3. Then, we see that

$$|F(x) - F(y)| \le c2^{-n} = c(3^{-n})^\gamma \le M|x - y|^\gamma,$$

since $3^\gamma = 2$ and 3^{-n} is not greater than $|x - y|$. This argument is repeated in Lemma 2.8 below.

With $E = \mathcal{C}$, f the Cantor-Lebesgue function, and $\alpha = \gamma = \log 2/\log 3$, the two lemmas give

$$m_1([0, 1]) \le M^\beta m_\alpha(\mathcal{C}).$$

Thus $m_\alpha(\mathcal{C}) > 0$, and we find that $\dim \mathcal{C} = \log 2/\log 3$.

The proof of this example is typical in the sense that the inequality $m_\alpha(\mathcal{C}) < \infty$ is usually easier to obtain than $0 < m_\alpha(\mathcal{C})$. Also, with some extra effort, it is possible to show that the $\log 2/\log 3$-dimensional Hausdorff measure of $\mathcal{C}$ is precisely 1. (See Exercise 7.)

Rectifiable curves

A further example of the role of dimension comes from looking at continuous curves in $\mathbb{R}^d$. Recall that a continuous curve $\gamma : [a, b] \to \mathbb{R}^d$ is said to be **simple** if $\gamma(t_1) \neq \gamma(t_2)$ whenever $t_1 \neq t_2$, and **quasi-simple** if the mapping $t \mapsto z(t)$ is injective for t in the complement of finitely many points.

Theorem 2.4 *Suppose the curve γ is continuous and quasi-simple. Then γ is rectifiable if and only if $\Gamma = \{\gamma(t) : a \leq t \leq b\}$ has strict Hausdorff dimension one. Moreover, in this case the length of the curve is precisely its one-dimensional measure $m_1(\Gamma)$.*

Proof. Suppose to begin with that Γ is a rectifiable curve of length L, and consider an arc-length parametrization $\tilde{\gamma}$ such that $\Gamma = \{\tilde{\gamma}(t) : 0 \leq t \leq L\}$. This parametrization satisfies the Lipschitz condition

$$|\tilde{\gamma}(t_1) - \tilde{\gamma}(t_2)| \leq |t_1 - t_2|.$$

This follows since $|t_1 - t_2|$ is the length of the curve between t_1 and t_2, which is greater than the distance from $\tilde{\gamma}(t_1)$ to $\tilde{\gamma}(t_2)$. Since $\tilde{\gamma}$ satisfies the conditions of Lemma 2.2 with exponent 1 and $M = 1$, we find that

$$m_1(\Gamma) \leq L.$$

To prove the reverse inequality, we let $a = t_0 < t_1 < \cdots < t_N = b$ denote a partition of $[a, b]$ and let

$$\Gamma_j = \{\gamma(t) : t_j \leq t \leq t_{j+1}\},$$

so that $\Gamma = \bigcup_{j=0}^{N-1} \Gamma_j$, and hence

$$m_1(\Gamma) = \sum_{j=0}^{N-1} m_1(\Gamma_j)$$

by an application of Property 4 of the Hausdorff measure and the fact that Γ is quasi-simple. Indeed, by removing finitely many points the union $\bigcup_{j=0}^{N-1} \Gamma_j$ becomes disjoint, while the points removed clearly have zero m_1-measure. We next claim that $m_1(\Gamma_j) \geq \ell_j$, where ℓ_j is the distance from $\gamma(t_j)$ to $\gamma(t_{j+1})$, that is, $\ell_j = |\gamma(t_{j+1}) - \gamma(t_j)|$. To see this, recall that Hausdorff measure is rotation-invariant, and introduce new orthogonal coordinates x and y such that $[\gamma(t_j), \gamma(t_j + 1)]$ is the segment $[0, \ell_j]$ on the x-axis. The projection $\pi(x, y) = x$ satisfies the Lipschitz condition

$$|\pi(P) - \pi(Q)| \leq |P - Q|,$$

and clearly the segment $[0, \ell_j]$ on the x-axis is contained in the image $\pi(\Gamma_j)$. Therefore, Lemma 2.2 guarantees

$$\ell_j \leq m_1(\Gamma_j),$$

and thus $m_1(\Gamma) \geq \sum \ell_j$. Since by definition the length L of Γ is the supremum of the sums $\sum \ell_j$ over all partitions of $[a, b]$, we find that $m_1(\Gamma) \geq L$, as desired.

Conversely, if Γ has strict Hausdorff dimension 1, then $m_1(\Gamma) < \infty$, and the above argument shows that Γ is rectifiable.

The reader may note the resemblance of this characterization of rectifiability and an earlier one in terms of Minkowski content, given in Chapter 3. In this connection we point out that there is a different notion of dimension that is sometimes used instead of Hausdorff dimension. For a compact set E, this dimension is given in terms of the size of $E^\delta = \{x \in \mathbb{R}^d : \ d(x, E) < \delta\}$ as $\delta \to 0$. One observes that if E is a k-dimensional cube in $\mathbb{R}^d$, then $m(E^\delta) \leq c\delta^{d-k}$ as $\delta \to 0$, with m the Lebesgue measure of $\mathbb{R}^d$. With this in mind, the **Minkowski dimension** of E is defined by

$$\inf \{\beta : \ m(E^\delta) = O(\delta^{d-\beta}) \text{ as } \delta \to 0\}.$$

One can show that the Hausdorff dimension of a set does not exceed its Minkowski dimension, but that equality does not hold in general. More details may be found in Exercises 17 and 18.

The Sierpinski triangle

A Cantor-like set can be constructed in the plane as follows. We begin with a (solid) closed equilateral triangle S_0, whose sides have unit length. Then, as a first step we remove the shaded open equilateral triangle pictured in Figure 1.

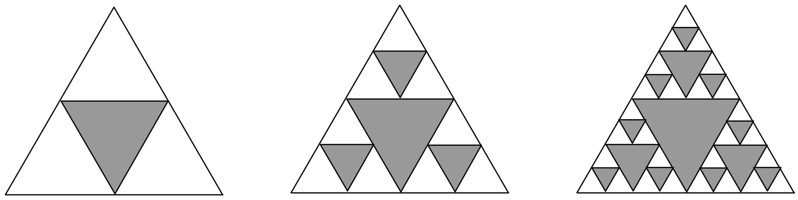

Figure 1. Construction of the Sierpinski triangle

This leaves three closed triangles whose union we denote by S_1. Each triangle is half the size of the original (or parent) triangle S_0, and these smaller closed triangles are said to be of the first **generation**: the triangles in S_1 are the children of the parent S_0. In the second step, we repeat the process in each triangle of the first generation. Each such triangle has three children of the second generation. We denote by S_2 the union of the three triangles in the second generation. We then repeat this process to find a sequence S_k of compact sets which satisfy the following properties:

(a) Each S_k is a union of 3^k closed equilateral triangles of side length 2^{-k}. (These are the triangles of the k^{th} generation.)

(b) $\{S_k\}$ is a decreasing sequence of compact sets; that is, $S_{k+1} \subset S_k$ for all $k \geq 0$.

The **Sierpinski triangle** is the compact set defined by

$$\mathcal{S} = \bigcap_{k=0}^{\infty} S_k.$$

Theorem 2.5 *The Sierpinski triangle $\mathcal{S}$ has strict Hausdorff dimension $\alpha = \log 3/\log 2$.*

The inequality $m_\alpha(\mathcal{S}) \leq 1$ follows immediately from the construction. Given $\delta > 0$, choose K so that $2^{-K} < \delta$. Since the set S_K covers $\mathcal{S}$ and consists of 3^K triangles each of diameter $2^{-K} < \delta$, we must have

$$\mathcal{H}_\alpha^\delta(\mathcal{S}) \leq 3^K(2^{-K})^\alpha.$$

But since $2^\alpha = 3$, we find $\mathcal{H}_\alpha^\delta(\mathcal{S}) \leq 1$, hence $m_\alpha(\mathcal{S}) \leq 1$.

The inequality $m_\alpha(\mathcal{S}) > 0$ is more subtle. For its proof we need to fix a special point in each triangle that appears in the construction of $\mathcal{S}$.

We choose to call the lower left vertex of a triangle *the* vertex of that triangle. With this choice there are 3^k vertices of the k^{th} generation. The argument that follows is based on the important fact that all these vertices belong to $\mathcal{S}$.

Suppose $\mathcal{S} \subset \bigcup_{j=1}^{\infty} F_j$, with diam $F_j < \delta$. We wish to prove that

$$\sum_j (\text{diam } F_j)^\alpha \geq c > 0$$

for some constant c. Clearly, each F_j is contained in a ball of twice the diameter of F_j, so upon replacing 2δ by δ and noting that $\mathcal{S}$ is compact, it suffices to show that if $\mathcal{S} \subset \bigcup_{j=1}^{N} B_j$, where $\mathcal{B} = \{B_j\}_{j=1}^{N}$ is a finite collection of balls whose diameters are less than δ, then

$$\sum_{j=1}^{N} (\text{diam } B_j)^\alpha \geq c > 0.$$

Suppose we have such a covering by balls. Consider the minimum diameter of the B_j, and choose k so that

$$2^{-k} \leq \min_{1 \leq j \leq N} \text{diam } B_j < 2^{-k+1}.$$

Lemma 2.6 *Suppose B is a ball in the covering $\mathcal{B}$ that satisfies*

$$2^{-\ell} \leq \text{diam } B < 2^{-\ell+1} \quad \textit{for some } \ell \leq k.$$

Then B contains at most $c3^{k-\ell}$ vertices of the k^{th} generation.

In this chapter, we shall continue use the common practice of denoting by $c, c', \ldots$ generic constants whose values are unimportant and may change from one usage to another. We also use $A \approx B$ to denote that the quantities A and B are **comparable**, that is, $cB \leq A \leq c'B$, for appropriate constants c and c'.

Proof of Lemma 2.6. Let B^* denote the ball with same center as B but three times its diameter, and let $\triangle_k$ be a triangle of the k^{th} generation whose vertex v lies in B. If $\triangle'_\ell$ denotes the triangle of the ℓ^{th} generation that contains $\triangle_k$, then since diam $B \geq 2^{-\ell}$,

$$v \in \triangle_k \subset \triangle'_\ell \subset B^*,$$

as shown in Figure 2.

Next, there is a positive constant c such that B^* can contain at most c distinct triangles of the ℓ^{th} generation. This is because triangles of the

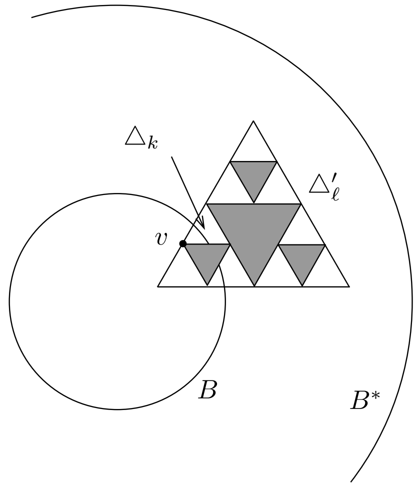

Figure 2. The setting in Lemma 2.6

ℓ^{th} generation have disjoint interiors and area equal to $c'4^{-\ell}$, while B^* has area at most equal to $c''4^{-\ell}$. Finally, each $\triangle'_\ell$ contains $3^{k-\ell}$ triangles of the k^{th} generation, hence B can contain at most $c3^{k-\ell}$ vertices of triangles of the k^{th} generation.

To complete the proof that $\sum_{j=1}^N(\text{diam } B_j)^\alpha \geq c > 0$, note that

$$\sum_{j=1}^N(\text{diam } B_j)^\alpha \geq \sum_\ell N_\ell 2^{-\ell\alpha},$$

where N_ℓ denotes the number of balls in $\mathcal{B}$ that satisfy $2^{-\ell} \leq \text{diam } B_j \leq 2^{-\ell+1}$. By the lemma, we see that the total number of vertices of triangles in the k^{th} generation that can be covered by the collection $\mathcal{B}$ can be no more than $c\sum_\ell N_\ell 3^{k-\ell}$. Since all 3^k vertices of triangles in the k^{th} generation belong to $\mathcal{S}$, and all vertices of the k^{th} generation must be covered, we must have $c\sum_\ell N_\ell 3^{k-\ell} \geq 3^k$. Hence

$$\sum_\ell N_\ell 3^{-\ell} \geq c.$$

It now suffices to recall the definition of α which guarantees $2^{-\ell\alpha} = 3^{-\ell}$, and therefore

$$\sum_{j=1}^N(\text{diam } B_j)^\alpha \geq c,$$

as desired.

We give a final example that exhibits properties similar to the Cantor set and Sierpinski triangle. It is the curve discovered by von Koch in 1904.

The von Koch curve

Consider the unit interval $K_0 = [0, 1]$, which we may think of as lying on the x-axis in the xy-plane. Then consider the polygonal path K_1 illustrated in Figure 3, which consists of four equal line segments of length 1/3.

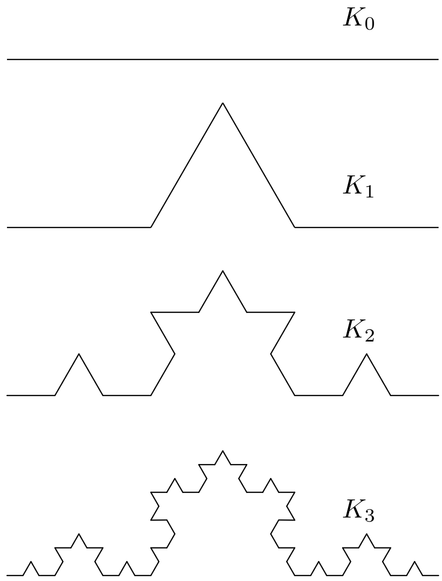

Figure 3. The first few stages in the construction of the von Koch curve

Let $K_1(t)$, for $0 \leq t \leq 1$, denote the parametrization of K_1 that has constant speed. In other words, as t travels from 0 to 1/4, the point $K_1(t)$ travels on the first line segment. As t travels from 1/4 to 1/2, the point $K_1(t)$ travels on the second line segment, and so on. In particular, we see that $K_1(\ell/4)$ for $0 \leq \ell \leq 4$ correspond to the five vertices of K_1.

At the second stage of the construction we repeat the process of replacing each line segment in stage one by the corresponding polygonal line. We then obtain the polygonal curve K_2 illustrated in Figure 3. It has $16 = 4^2$ segments of length $1/9 = 3^{-2}$. We choose a parametrization

$K_2(t)$ $(0 \leq t \leq 1)$ of K_2 that has constant speed. Observe that $K_2(\ell/4^2)$ for $0 \leq \ell \leq 4^2$ gives all vertices of K_2, and that the vertices of K_1 belong to K_2, with

$$K_2(\ell/4) = K_1(\ell/4) \quad \text{for } 0 \leq \ell \leq 4.$$

Repeating this process indefinitely, we obtain a sequence of continuous polygonal curves $\{K_j\}$, where K_j consists of 4^j segments of length 3^{-j} each. If $K_j(t)$ $(0 \leq t \leq 1)$ is the parametrization of K_j that has constant speed, then the vertices are precisely at the points $K_j(\ell/4^j)$, and

$$K_{j'}(\ell/4^j) = K_j(\ell/4^j) \quad \text{for } 0 \leq \ell \leq 4^j$$

whenever $j' \geq j$.

In the limit as j tends to infinity, the polygonal lines K_j tend to the **von Koch curve** $\mathcal{K}$. Indeed, we have

$$|K_{j+1}(t) - K_j(t)| \leq 3^{-j} \quad \text{for all } 0 \leq t \leq 1 \text{ and } j \geq 0.$$

This is clear when $j = 0$, and follows by induction in j when we consider the nature of the construction of the j^{th} stage. Since we may write

$$K_J(t) = K_1(t) + \sum_{j=1}^{J-1} (K_{j+1}(t) - K_j(t)),$$

the above estimate proves that the series

$$K_1(t) + \sum_{j=1}^{\infty} (K_{j+1}(t) - K_j(t))$$

converges absolutely and uniformly to a continuous function $\mathcal{K}(t)$ that is a parametrization of $\mathcal{K}$. Besides continuity, the function $\mathcal{K}(t)$ satisfies a regularity assumption that takes the form of a Lipschitz condition, as in the case of the Cantor-Lebesgue function.

Theorem 2.7 *The function $\mathcal{K}(t)$ satisfies a Lipschitz condition of exponent $\gamma = \log 3/\log 4$, that is:*

$$|\mathcal{K}(t) - \mathcal{K}(s)| \leq M|t-s|^\gamma \quad \textit{for all } t, s \in [0,1].$$

We have already observed that $|K_{j+1}(t) - K_j(t)| \leq 3^{-j}$. Since K_j travels a distance of 3^{-j} in 4^{-j} units of time, we see that

$$|K_j'(t)| \leq \left(\frac{4}{3}\right)^j \quad \text{except when } t = \ell/4^j.$$

Consequently we must have

$$|K_j(t) - K_j(s)| \leq \left(\frac{4}{3}\right)^j |t - s|.$$

Moreover, $\mathcal{K}(t) = K_1(t) + \sum_{j=1}^{\infty}(K_{j+1}(t) - K_j(t))$. We now find ourselves in precisely the same situation as in the proof that the Cantor-Lebesgue function satisfies a Lipschitz condition with exponent $\log 2/\log 3$. We generalize that argument in the following lemma.

Lemma 2.8 *Suppose* $\{f_j\}$ *is a sequence of continuous functions on the interval* $[0, 1]$ *that satisfy*

$$|f_j(t) - f_j(s)| \leq A^j|t - s| \quad \text{for some } A > 1,$$

and

$$|f_j(t) - f_{j+1}(t)| \leq B^{-j} \quad \text{for some } B > 1.$$

Then the limit $f(t) = \lim_{j\to\infty} f_j(t)$ *exists and satisfies*

$$|f(t) - f(s)| \leq M|t - s|^\gamma,$$

where $\gamma = \log B/\log(AB)$.

Proof. The continuous limit f is given by the uniformly convergent series

$$f(t) = f_1(t) + \sum_{k=1}^{\infty}(f_{k+1}(t) - f_k(t)),$$

and therefore

$$|f(t) - f_j(t)| \leq \sum_{k=j}^{\infty} |f_{k+1}(t) - f_k(t)| \leq \sum_{k=j}^{\infty} B^{-k} \leq cB^{-j}.$$

The triangle inequality, an application of the inequality just obtained, and the inequality in the statement of the lemma give

$$\begin{aligned} |f(t) - f(s)| &\leq |f_j(t) - f_j(s)| + |(f - f_j)(t)| + |(f - f_j)(s)| \\ &\leq c(A^j|t - s| + B^{-j}). \end{aligned}$$

For a fixed pair of numbers t and s with $t \neq s$, we choose j to minimize the sum $A^j|t - s| + B^{-j}$. This is essentially achieved by picking j so that two terms $A^j|t-s|$ and B^{-j} are comparable. More precisely, we choose a j that satisfies

$$(AB)^j|t-s| \leq 1 \quad \text{ and } \quad 1 \leq (AB)^{j+1}|t-s|.$$

Since $|t-s| \leq 2$ and $AB > 1$, such a j must exist. The first inequality then gives

$$A^j|t-s| \leq B^{-j},$$

while raising the second inequality to the power γ, and using the fact that $(AB)^\gamma = B$ gives

$$1 \leq B^j|t-s|^\gamma.$$

Thus $B^{-j} \leq |t-s|^\gamma$, and consequently

$$|f(t)-f(s)| \leq c(A^j|t-s| + B^{-j}) \leq M|t-s|^\gamma,$$

as was to be shown.

In particular, this result with Lemma 2.2 implies that

$$\dim \mathcal{K} \leq \frac{1}{\gamma} = \frac{\log 4}{\log 3}.$$

To prove that $m_\gamma(\mathcal{K}) > 0$ and hence $\dim \mathcal{K} = \log 4/\log 3$ requires an argument similar to the one given for the Sierpinski triangle. In fact, this argument generalizes to cover a general family of sets that have a self-similarity property. We therefore turn our attention to this general theory next.

Remarks. We mention some further facts about the von Koch curve. More details can be found in Exercises 13, 14, and 15 below.

1. The curve $\mathcal{K}$ is one in a family of similarly constructed curves. For each ℓ, $1/4 < \ell < 1/2$, consider at the first stage the curve $K_1^\ell(t)$ given by four line segments each of length ℓ, the first and last on the x-axis, and the second and third forming two sides of an isoceles triangle whose base lies on the x-axis. (See Figure 4.) The case $\ell = 1/3$ corresponds to the previously defined von Koch curve.

   Proceeding as in the case $\ell = 1/3$, one obtains a curve $\mathcal{K}^\ell$, and it can be seen that

   $$\dim(\mathcal{K}^\ell) = \frac{\log 4}{\log 1/\ell}.$$

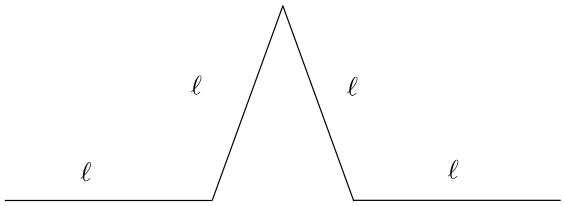

Figure 4. The curve $K_1^\ell(t)$

Thus for every α, $1 < \alpha < 2$, we have a curve of this kind of dimension α. Note that when $\ell \to 1/4$ the limiting curve is a straight line segment, which has dimension 1. When $\ell \to 1/2$, the limit can be seen to correspond to a "space-filling" curve.

2. The curves $t \mapsto \mathcal{K}^\ell(t)$, $1/4 < \ell \leq 1/2$, are each nowhere differentiable. One can also show that each curve is simple when $1/4 \leq \ell < 1/2$.

2.2 Self-similarity

The Cantor set $\mathcal{C}$, the Sierpinski triangle $\mathcal{S}$, and von Koch curve $\mathcal{K}$ all share an important property: each of these sets contains scaled copies of itself. Moreover, each of these examples was constructed by iterating a process closely tied to its scaling. For instance, the interval $[0, 1/3]$ contains a copy of the Cantor set scaled by a factor of 1/3. The same is true for the interval $[2/3, 1]$, and therefore

$$\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2,$$

where $\mathcal{C}_1$ and $\mathcal{C}_2$ are scaled versions of $\mathcal{C}$. Also, each interval $[0, 1/9]$, $[2/9, 3/9]$, $[6/9, 7/9]$ and $[8/9, 1]$ contains a copy of $\mathcal{C}$ scaled by a factor of 1/9, and so on.

In the case of the Sierpinski triangle, each of the three triangles in the first generation contains a copy of $\mathcal{S}$ scaled by the factor of 1/2. Hence

$$\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3,$$

where each $\mathcal{S}_j$, $j = 1, 2, 3$, is obtained by scaling and translating the original Sierpinski triangle. More generally, every triangle in the k^{th} generation is a copy of $\mathcal{S}$ scaled by the factor of $1/2^k$.

Finally, each line segment in the initial stage of the construction of the von Koch curve gives rise to a scaled and possibly rotated copy of the von Koch curve. In fact

$$\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2 \cup \mathcal{K}_3 \cup \mathcal{K}_4,$$

where $\mathcal{K}_j$, $j = 1, 2, 3, 4$, is obtained by scaling $\mathcal{K}$ by the factor of $1/3$ and translating and rotating it.

Thus these examples each contain replicas of themselves, but on a smaller scale. In this section, we give a precise definition of the resulting notion of self-similarity and prove a theorem determining the Hausdorff dimension of these sets.

A mapping $S : \mathbb{R}^d \to \mathbb{R}^d$ is said to be a **similarity** with **ratio** $r > 0$ if

$$|S(x) - S(y)| = r|x - y|.$$

It can be shown that every similarity of $\mathbb{R}^d$ is the composition of a translation, a rotation, and a dilation by r. (See Problem 3.)

Given finitely many similarities $S_1, \ldots, S_m$ with the same ratio r, we say that the set $F \subset \mathbb{R}^d$ is **self-similar** if

$$F = S_1(F) \cup \cdots \cup S_m(F).$$

We point out the relevance of the various examples we have already seen.

When $F = \mathcal{C}$ is the Cantor set, there are two similarities given by

$$S_1(x) = x/3 \quad \text{and} \quad S_2(x) = x/3 + 2/3$$

of ratio $1/3$. So $m = 2$ and $r = 1/3$.

In the case of $F = \mathcal{S}$, the Sierpinski triangle, the ratio is $r = 1/2$ and there are $m = 3$ similarities given by

$$S_1(x) = \frac{x}{2}, \quad S_2(x) = \frac{x}{2} + \alpha \quad \text{and} \quad S_3(x) = \frac{x}{2} + \beta.$$

Here, α and β are the points drawn in the first diagram in Figure 5.

If $F = \mathcal{K}$, the von Koch curve, we have

$$S_1(x) = \frac{x}{3}, \quad S_2(x) = \rho\frac{x}{3} + \alpha, \quad S_3(x) = \rho^{-1}\frac{x}{3} + \beta,$$

and

$$S_4(x) = \frac{x}{3} + \gamma,$$

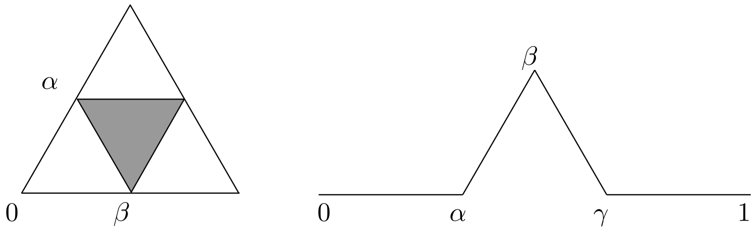

Figure 5. Similarities of the Sierpinski triangle and von Koch curve

where ρ is the rotation centered at the origin and of angle $\pi/3$. There are $m = 4$ similarities which have ratio $r = 1/3$. The points α, β, and γ are shown in the second diagram in Figure 5.

Another example, sometimes called the **Cantor dust** $\mathcal{D}$, is another two-dimensional version of the standard Cantor set. For each fixed $0 < \mu < 1/2$, the set $\mathcal{D}$ may be constructed by starting with the unit square $Q = [0,1] \times [0,1]$. At the first stage we remove everything but the four open squares in the corners of Q that have side length μ. This yields a union D_1 of four squares, as illustrated in Figure 6.

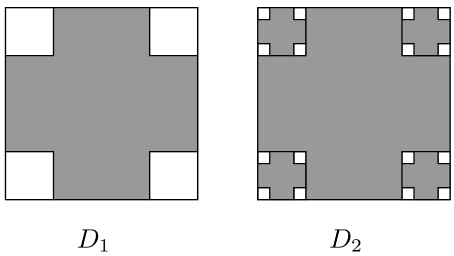

Figure 6. Construction of the Cantor dust

We repeat this process in each sub-square of D_1; that is, we remove everything but the four squares in the corner, each of side length μ^2. This gives a union D_2 of 16 squares. Repeating this process, we obtain a family $D_1 \supset D_2 \supset \cdots \supset D_k \supset \cdots$ of compact sets whose intersection defines the Cantor dust corresponding to the parameter μ.

There are here $m = 4$ similarities of ratio μ given by

$$\begin{aligned}
S_1(x) &= \mu x,\\
S_2(x) &= \mu x + (0, 1-\mu),\\
S_3(x) &= \mu x + (1-\mu, 1-\mu),\\
S_4(x) &= \mu x + (1-\mu, 0).
\end{aligned}$$

It is to be noted that $\mathcal{D}$ is the product $\mathcal{C}_\xi \times \mathcal{C}_\xi$, with $\mathcal{C}_\xi$ the Cantor set of constant dissection ξ, as defined in Exercise 3, of Chapter 1. Here $\xi = 1 - 2\mu$.

The first result we prove guarantees the existence of self-similar sets under the assumption that the similarities are contracting, that is, that their ratio satisfies $r < 1$.

Theorem 2.9 *Suppose $S_1, S_2, \ldots, S_m$ are m similartities, each with the same ratio r that satisfies $0 < r < 1$. Then there exists a unique non-empty compact set F such that*

$$F = S_1(F) \cup \cdots \cup S_m(F).$$

The proof of this theorem is in the nature of a fixed point argument. We shall begin with some large ball B and iteratively apply the mappings $S_1, \ldots, S_m$. The fact that the similarities have ratio $r < 1$ will suffice to imply that this process contracts to a unique set F with the desired property.

Lemma 2.10 *There exists a closed ball B so that $S_j(B) \subset B$ for all $j = 1, \ldots, m$.*

Proof. Indeed, we note that if S is a similarity with ratio r, then

$$\begin{aligned}
|S(x)| &\leq |S(x) - S(0)| + |S(0)|\\
&\leq r|x| + |S(0)|.
\end{aligned}$$

If we require that $|x| \leq R$ implies $|S(x)| \leq R$, it suffices to choose R so that $rR + |S(0)| \leq R$, that is, $R \geq |S(0)|/(1-r)$. In this fashion, we obtain for each S_j a ball B_j centered at the origin that satisfies $S_j(B_j) \subset B_j$. If B denotes the ball among the B_j with the largest radius, then the above shows that $S_j(B) \subset B$ for all j.

Now for any set A, let $\tilde{S}(A)$ denote the set given by

$$\tilde{S}(A) = S_1(A) \cup \cdots \cup S_m(A).$$

Note that if $A \subset A'$, then $\tilde{S}(A) \subset \tilde{S}(A')$.

Also observe that while each S_j is a mapping from $\mathbb{R}^d$ to $\mathbb{R}^d$, the mapping $\tilde{S}$ is not a point mapping, but takes subsets of $\mathbb{R}^d$ to subsets of $\mathbb{R}^d$.

To exploit the notion of contraction with a ratio less than 1, we introduce the distance between two compact sets as follows. For each $\delta > 0$ and set A, we let

$$A^\delta = \{x : d(x, A) < \delta\}.$$

Hence A^δ is a set that contains A but which is slightly larger in terms of δ. If A and B are two compact sets, we define the **Hausdorff distance** as

$$\mathrm{dist}(A, B) = \inf\{\delta :\ B \subset A^\delta \text{ and } A \subset B^\delta\}.$$

Lemma 2.11 *The distance function* dist *defined on compact subsets of* $\mathbb{R}^d$ *satisfies*

(i) $\mathrm{dist}(A, B) = 0$ *if and only if* $A = B$.

(ii) $\mathrm{dist}(A, B) = \mathrm{dist}(B, A)$.

(iii) $\mathrm{dist}(A, B) \leq \mathrm{dist}(A, C) + \mathrm{dist}(C, B)$.

If $S_1, \ldots, S_m$ *are similarities with ratio* r, *then*

(iv) $\mathrm{dist}(\tilde{S}(A), \tilde{S}(B)) \leq r\, \mathrm{dist}(A, B)$.

The proof of the lemma is simple and may be left to the reader.

Using both lemmas we may now prove Theorem 2.9. We first choose B as in Lemma 2.10, and let $F_k = \tilde{S}^k(B)$, where $\tilde{S}^k$ denotes the k^{th} composition of $\tilde{S}$, that is, $\tilde{S}^k = \tilde{S}^{k-1} \circ \tilde{S}$ with $\tilde{S}^1 = \tilde{S}$. Each F_k is compact, non-empty, and $F_k \subset F_{k-1}$, since $\tilde{S}(B) \subset B$. If we let

$$F = \bigcap_{k=1}^{\infty} F_k,$$

then F is compact, non-empty, and clearly $\tilde{S}(F) = F$, since applying $\tilde{S}$ to $\bigcap_{k=1}^{\infty} F_k$ yields $\bigcap_{k=2}^{\infty} F_k$, which also equals F.

Uniqueness of the set F is proved as follows. Suppose G is another compact set so that $\tilde{S}(G) = G$. Then, an application of part (iv) in Lemma 2.11 yields $\mathrm{dist}(F, G) \leq r\, \mathrm{dist}(F, G)$. Since $r < 1$, this forces $\mathrm{dist}(F, G) = 0$, so that $F = G$, and the proof of Theorem 2.9 is complete.

Under an additional technical condition, one can calculate the precise Hausdorff dimension of the self-similar set F. Loosely speaking, the restriction holds if the sets $S_1(F), \ldots, S_m(F)$ do not overlap too much. Indeed, if these sets were disjoint, then we could argue that

$$m_\alpha(F) = \sum_{j=1}^{m} m_\alpha(S_j(F)).$$

Since each S_j scales by r, we would then have $m_\alpha(S_j(F)) = r^\alpha m_\alpha(F)$. Hence

$$m_\alpha(F) = mr^\alpha m_\alpha(F).$$

If $m_\alpha(F)$ were finite, then we would have that $mr^\alpha = 1$; thus

$$\alpha = \frac{\log m}{\log 1/r}.$$

The restriction we impose is as follows. We say that the similarities $S_1, \ldots, S_m$ are **separated** if there is an bounded open set $\mathcal{O}$ so that

$$\mathcal{O} \supset S_1(\mathcal{O}) \cup \cdots \cup S_m(\mathcal{O}),$$

and the $S_j(\mathcal{O})$ are disjoint. It is not assumed that $\mathcal{O}$ contains F.

Theorem 2.12 *Suppose $S_1, S_2, \ldots, S_m$ are m separated similarities with the common ratio r that satisfies $0 < r < 1$. Then the set F has Hausdorff dimension equal to $\log m / \log(1/r)$.*

Observe first that when F is the Cantor set we may take $\mathcal{O}$ to be the open unit interval, and note that we have already proved that its dimension is $\log 2/\log 3$. For the Sierpinski triangle the open unit triangle will do, and $\dim \mathcal{S} = \log 3/\log 2$. In the example of the Cantor dust the open unit square works, and $\dim \mathcal{D} = \log m/\log \mu^{-1}$. Finally, for the von Koch curve we may take the interior of the triangle pictured in Figure 7, and we will have $\dim \mathcal{K} = \log 4/\log 3$.

We now turn to the proof of Theorem 2.12, which will follow the same approach used in the case of the Sierpinski triangle. If $\alpha = \log m/\log(1/r)$, we claim that $m_\alpha(F) < \infty$, hence $\dim F \leq \alpha$. Moreover, this inequality holds even without the separation assumption. Indeed, recall that

$$F_k = \tilde{S}^k(B),$$

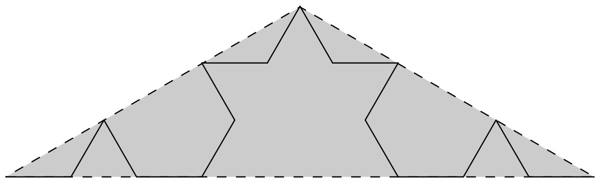

Figure 7. Open set in the separation of the von Koch similarities

and $\tilde{S}^k(B)$ is the union of m^k sets of diameter less than cr^k (with $c =$ diam B), each of the form

$$S_{n_1} \circ S_{n_2} \circ \cdots \circ S_{n_k}(B), \qquad \text{where } 1 \leq n_i \leq m \text{ and } 1 \leq i \leq k.$$

Consequently, if $cr^k \leq \delta$, then

$$\begin{aligned}\mathcal{H}_\alpha^\delta(F) &\leq \sum_{n_1,\ldots,n_k} (\text{diam } S_{n_1} \circ \cdots \circ S_{n_k}(B))^\alpha \\ &\leq c' m^k r^{\alpha k} \\ &\leq c',\end{aligned}$$

since $mr^\alpha = 1$, because $\alpha = \log m / \log(1/r)$. Since c' is independent of δ, we get $m_\alpha(F) \leq c'$.

To prove $m_\alpha(F) > 0$, we now use the separation condition. We argue in parallel with the earlier calculation of the Hausdorff dimension of the Sierpinski triangle.

Fix a point $\overline{x}$ in F. We define *the* "vertices" of the k^{th} generation as the m^k points that lie in F and are given by

$$S_{n_1} \circ \cdots \circ S_{n_k}(\overline{x}), \qquad \text{where } 1 \leq n_1 \leq m, \ldots, 1 \leq n_k \leq m.$$

Each vertex is labeled by $(n_1, \ldots, n_k)$. Vertices need not be distinct, so they are counted with their multiplicities.

Similarly, we define *the* "open sets" of the k^{th} generation to be the m^k sets given by

$$S_{n_1} \circ \cdots \circ S_{n_k}(\mathcal{O}), \qquad \text{where } 1 \leq n_1 \leq m, \ldots, 1 \leq n_k \leq m,$$

and where $\mathcal{O}$ is fixed and chosen to satisfy the separation condition. Such open sets are again labeled by multi-indices $(n_1, n_2, \ldots, n_k)$ with $1 \leq n_j \leq m$, $1 \leq j \leq k$.

Then the open sets of the k^{th} generation are disjoint, since those of the first generation are disjoint. Moreover if $k \geq \ell$, each open set of the ℓ^{th} generation contains $m^{k-\ell}$ open sets of the k^{th} generation.

Suppose v is a vertex of the k^{th} generation, and let $\mathcal{O}(v)$ denote the open set in the k^{th} generation which is associated to v, that is, v and $\mathcal{O}(v)$ carry the same label $(n_1, n_2, \ldots, n_k)$. Since $\overline{x}$ is at a fixed distance from the original open set $\mathcal{O}$, and $\mathcal{O}$ has a finite diameter, we find that

(a) $d(v, \mathcal{O}(v)) \leq cr^k$.

(b) $c'r^k \leq \text{diam } \mathcal{O}(v) \leq cr^k$.

As in the case of the Sierpinski triangle, it suffices to prove that if $\mathcal{B} = \{B_j\}_{j=1}^N$ is a finite collection of balls whose diameters are less than δ and whose union covers F, then

$$\sum_{j=1}^{N} (\text{diam } B_j)^\alpha \geq c > 0.$$

Suppose we have such a covering by balls, and choose k so that

$$r^k \leq \min_{1 \leq j \leq N} \text{diam } B_j < r^{k-1}.$$

Lemma 2.13 *Suppose B is a ball in the covering $\mathcal{B}$ that satisfies*

$$r^\ell \leq \text{diam } B < r^{\ell-1} \quad \textit{for some } \ell \leq k.$$

Then B contains at most $cm^{k-\ell}$ vertices of the k^{th} generation.

Proof. If v is a vertex of the k^{th} generation with $v \in B$, and $\mathcal{O}(v)$ denotes the corresponding open set of the k^{th} generation, then, for some fixed dilate B^* of B, properties (a) and (b) above guarantee that $\mathcal{O}(v) \subset B^*$, and B^* also contains the open set of generation ℓ that contains $\mathcal{O}(v)$.

Since B^* has volume $cr^{d\ell}$, and each open set in the ℓ^{th} generation has volume $\approx r^{d\ell}$ (by property (b) above), B^* can contain at most c open sets of generation ℓ. Hence B^* contains at most $cm^{k-\ell}$ open sets of the k^{th} generation. Consequently, B can contain at most $cm^{k-\ell}$ vertices of the k^{th} generation, and the lemma is proved.

For the final argument, let N_ℓ denote the number of balls in $\mathcal{B}$ so that

$$r^\ell \leq \text{diam } B_j \leq r^{\ell-1}.$$

By the lemma, we see that the total number of vertices of the k^{th} generation that can be covered by the collection $\mathcal{B}$ can be no more than

$c\sum_\ell N_\ell m^{k-\ell}$. Since all m^k vertices of the k^{th} generation belong to F, we must have $c\sum_\ell N_\ell m^{k-\ell} \geq m^k$, and hence

$$\sum_\ell N_\ell m^{-\ell} \geq c.$$

The definition of α gives $r^{\ell\alpha} = m^{-\ell}$, and therefore

$$\sum_{j=1}^{N} (\text{diam } B_j)^\alpha \geq \sum_\ell N_\ell r^{\ell\alpha} \geq c,$$

and the proof of Theorem 2.12 is complete.

3 Space-filling curves

The year 1890 heralded an important discovery: Peano constructed a continuous curve that filled an entire square in the plane. Since then, many variants of his construction have been given. We shall describe here a construction that has the feature of elucidating an additional significant fact. It is that from the point of measure theory, speaking broadly, the unit interval and unit square are "isomorphic."

Theorem 3.1 *There exists a curve $t \mapsto \mathcal{P}(t)$ from the unit interval to the unit square with the following properties:*

(i) *$\mathcal{P}$ maps $[0,1]$ to $[0,1] \times [0,1]$ continuously and surjectively.*

(ii) *$\mathcal{P}$ satisfies a Lipschitz condition of exponent $1/2$, that is,*

$$|\mathcal{P}(t) - \mathcal{P}(s)| \leq M|t-s|^{1/2}.$$

(iii) *The image under $\mathcal{P}$ of any sub-interval $[a,b]$ is a compact subset of the square of (two-dimensional) Lebesgue measure exactly $b-a$.*

The third conclusion can be elaborated further.

Corollary 3.2 *There are subsets $Z_1 \subset [0,1]$ and $Z_2 \subset [0,1] \times [0,1]$, each of measure zero, such that $\mathcal{P}$ is bijective from*

$$[0,1] - Z_1 \quad \textit{to} \quad [0,1] \times [0,1] - Z_2$$

and measure preserving. In other words, E is measurable if and only if $\mathcal{P}(E)$ is measurable, and

$$m_1(E) = m_2(\mathcal{P}(E)).$$

Here m_1 and m_2 denote the Lebesgue measures in $\mathbb{R}^1$ and $\mathbb{R}^2$, respectively.

We shall call the function $t \mapsto \mathcal{P}(t)$ the **Peano mapping**. Its image is called the **Peano curve**.

Several observations help clarify the nature of the conclusions of the theorem. Suppose that $F : [0,1] \to [0,1] \times [0,1]$ is continuous and surjective. Then:

(a) F cannot be Lipschitz of exponent $\gamma > 1/2$. This follows at once from Lemma 2.2, which states that

$$\dim\ F([0,1]) \leq \frac{1}{\gamma} \dim[0,1],$$

so that $2 \leq 1/\gamma$ as desired.

(b) F cannot be injective. Indeed, if this were the case, then the inverse G of F would exist and would be continuous. Given any two points $a \neq b$ in $[0,1]$, we would get a contradiction by looking at two distinct curves in the square that join $F(a)$ and $F(b)$, since the image of these two curves under G would have to intersect at points between a and b. In fact, given any open disc D in the square, there always exists $x \in D$ so that $F(t) = F(s) = x$ yet $t \neq s$.

The proof of Theorem 3.1 will follow from a careful study of a natural class of mappings that associate sub-squares in $[0,1] \times [0,1]$ to subintervals in $[0,1]$. This implements the approach implicit in Hilbert's iterative procedure, which he set forth in the first three stages in Figure 8.

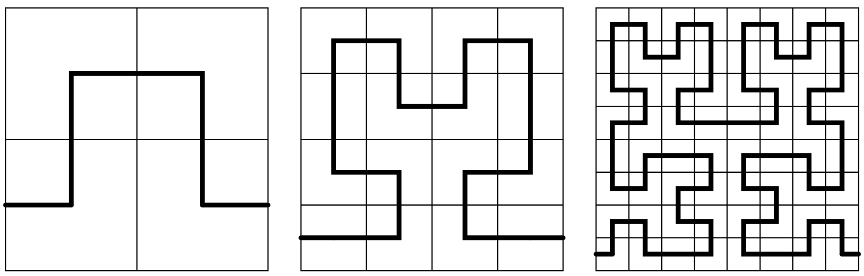

Figure 8. Construction of the Peano curve

We turn now to the study of the general class of mappings.

3.1 Quartic intervals and dyadic squares

The **quartic intervals** arise when $[0, 1]$ is successively sub-divided by powers of 4. For instance, the first generation quartic intervals are the closed intervals

$$I_1 = [0, 1/4], \quad I_2 = [1/4, 1/2], \quad I_3 = [1/2, 3/4], \quad I_4 = [3/4, 1].$$

The second generation quartic intervals are obtained by sub-dividing each interval of the first generation by 4. Hence there are $16 = 4^2$ quartic intervals of the second generation. In general, there are 4^k quartic intervals of the k^{th} generation, each of the form $[\frac{\ell}{4^k}, \frac{\ell+1}{4^k}]$, where ℓ is integral with $0 \leq \ell < 4^k$.

A **chain** of quartic intervals is a decreasing sequence of intervals

$$I^1 \supset I^2 \supset \cdots \supset I^k \supset \cdots,$$

where I^k is a quartic interval of the k^{th} generation (hence $|I^k| = 4^{-k}$).

Proposition 3.3 *Chains of quartic intervals satisfy the following properties:*

(i) *If $\{I^k\}$ is a chain of quartic intervals, then there exists a unique $t \in [0, 1]$ such that $t \in \bigcap_k I^k$.*

(ii) *Conversely, given $t \in [0, 1]$, there is a chain $\{I^k\}$ of quartic intervals such that $t \in \bigcap_k I^k$.*

(iii) *The set of t for which the chain in part* (ii) *is not unique is a set of measure zero (in fact, this set is countable).*

Proof. Part (i) follows from the fact that $\{I^k\}$ is a decreasing sequence of compact sets whose diameters go to 0.

For part (ii), we fix t and note that for each k there exists at least one quartic interval I^k with $t \in I^k$. If t is of the form $\ell/4^k$, where $0 < \ell < 4^k$, then there are exactly two quartic intervals of the k^{th} generation that contain t. Hence, the set of points for which the chain is not unique is precisely the set of **dyadic rationals**

$$\frac{\ell}{4^k}, \qquad \text{where } 1 \leq k, \text{ and } 0 < \ell < 4^k.$$

Note that of course, these fractions are the same as those of the form $\ell'/2^{k'}$ with $0 < \ell' < 2^{k'}$. This set is countable, hence has measure 0.

It is clear that each chain $\{I^k\}$ of quartic intervals can be represented naturally by a string $.a_1a_2\cdots a_k\cdots$, where each a_k is either $0, 1, 2$, or 3. Then the point t corresponding to this chain is given by

$$t = \sum_{k=1}^{\infty} \frac{a_k}{4^k}.$$

The points where ambiguity occurs are precisely those where $a_k = 3$ for all sufficiently large k, or equivalently where $a_k = 0$ for all sufficiently large k.

Part of our description of the Peano mapping will follow from associating to each quartic interval a dyadic square. These **dyadic squares** are obtained by sub-dividing the unit square $[0,1] \times [0,1]$ in the plane by successively bisecting the sides.

For instance, dyadic squares of the first generation arise from bisecting the sides of the unit square. This yields four closed squares S_1, S_2, S_3 and S_4, each of side length $1/2$ and area $|S_i| = 1/4$, for $i = 1, \ldots, 4$.

The dyadic squares of the second generation are obtained by bisecting each dyadic square of the first generation, and so on. In general, there are 4^k squares of the k^{th} generation, each of side length $1/2^k$ and area $1/4^k$.

A **chain** of dyadic squares is a decreasing sequence of squares

$$S^1 \supset S^2 \supset \cdots \supset S^k \supset \cdots,$$

where S^k is a dyadic square of the k^{th} generation.

Proposition 3.4 *Chains of dyadic squares have the following properties:*

(i) *If $\{S^k\}$ is a chain of dyadic squares, then there exists a unique $x \in [0,1] \times [0,1]$ such that $x \in \bigcap_k S^k$.*

(ii) *Conversely, given $x \in [0,1] \times [0,1]$, there is a chain $\{S^k\}$ of dyadic squares such that $x \in \bigcap_k S^k$.*

(iii) *The set of x for which the chain in part* (ii) *is not unique is a set of measure zero.*

In this case, the set of ambiguities consists of all points (x_1, x_2) where one of the coordinates is a dyadic rational. Geometrically, this set is the (countable) union of vertical and horizontal segments in $[0,1] \times [0,1]$ determined by the grid of dyadic rationals. This set has measure zero.

Moreover, each chain of dyadic squares can be represented by a string $.b_1 b_2 \cdots$, where each b_k is either $0, 1, 2$ or 3. Then

$$x = \sum_{k=1}^{\infty} \frac{\bar{b}_k}{2^k}, \tag{1}$$

where

$$\begin{aligned} \bar{b}_k &= (0,0) \quad \text{if } b_k = 0, \\ \bar{b}_k &= (0,1) \quad \text{if } b_k = 1, \\ \bar{b}_k &= (1,0) \quad \text{if } b_k = 2, \\ \bar{b}_k &= (1,1) \quad \text{if } b_k = 3. \end{aligned}$$

3.2 Dyadic correspondence

A **dyadic correspondence** is a mapping Φ from quartic intervals to dyadic squares that satisfies:

(1) Φ is bijective.

(2) Φ respects generations.

(3) Φ respects inclusion.

By (2), we mean that if I is a quartic interval of the k^{th} generation, then $\Phi(I)$ is a dyadic square of the k^{th} generation. By (3), we mean that if $I \subset J$, then $\Phi(I) \subset \Phi(J)$.

For example, the trivial, or standard correspondence assigns to the string $.a_1 a_2 \cdots$ the string $.b_1 b_2 \cdots$ with $b_k = a_k$.

Given a dyadic correspondence Φ, the **induced mapping** Φ^* maps $[0,1]$ to $[0,1] \times [0,1]$ and is given as follows. If $\{t\} = \bigcap I^k$ where $\{I^k\}$ is a chain of quartic intervals, then, since $\{\Phi(I^k)\}$ is a chain of dyadic squares, we may let

$$\Phi^*(t) = x = \bigcap \Phi(I^k).$$

We note that Φ^* is well-defined except on a (countable) set of measure zero, (those points t that are represented by more than one quartic chain.)

A moment's reflection will show that if I' is a quartic interval of the k^{th} generation, then the images $\Phi^*(I') = \{\Phi^*(t),\ t \in I'\}$, comprise the dyadic square of the k^{th} generation $\Phi(I')$. Thus $\Phi^*(I') = \Phi(I')$, and hence $m_1(I') = m_2(\Phi^*(I'))$.

Theorem 3.5 *Given a dyadic correspondence Φ, there exist sets $Z_1 \subset [0,1]$ and $Z_2 \subset [0,1] \times [0,1]$, each of measure zero, so that:*

(i) Φ^* *is a bijection on* $[0,1] - Z_1$ *to* $[0,1] \times [0,1] - Z_2$.

(ii) E *is measurable if and only if* $\Phi^*(E)$ *is measurable.*

(iii) $m_1(E) = m_2(\Phi^*(E))$.

Proof. First, let $\mathcal{N}_1$ denote the collection of chains of those quartic intervals arising in (iii) of Proposition 3.3, those for which the points in $I = [0,1]$ are not uniquely representable. Similarly, let $\mathcal{N}_2$ denote the collection of chains of those dyadic squares for which the corresponding points in the square $I \times I$ are not uniquely representable.

Since Φ is a bijection from chains of quartic intervals to chains of dyadic squares, it is also a bijection from $\mathcal{N}_1 \cup \Phi^{-1}(\mathcal{N}_2)$ to $\Phi(\mathcal{N}_1) \cup \mathcal{N}_2$, and hence also of their complements. Let Z_1 be the subset of I consisting of all points in I that can be represented (according to (i) of Proposition 3.3) by the chains in $\mathcal{N}_1 \cup \Phi^{-1}(\mathcal{N}_2)$, and let Z_2 be the set of points in the square that can be represented by dyadic squares in $\Phi(\mathcal{N}_1) \cup \mathcal{N}_2$. Then Φ^*, the induced mapping, is well-defined on $I - Z_1$, and gives a bijection of $I - Z_1$ to $(I \times I) - Z_2$. To prove that both Z_1 and Z_2 have measure zero, we invoke the following lemma. We suppose $\{f_k\}_{k=1}^{\infty}$ is a fixed given sequence, with each f_k either $0, 1, 2$, or 3.

Lemma 3.6 *Let*

$$E_0 = \{x = \sum_{k=1}^{\infty} a_k/4^k, \quad \text{where } a_k \neq f_k \text{ for all sufficiently large } k\}.$$

Then $m(E_0) = 0$.

Indeed, if we fix r, then $m(\{x : \ a_r \neq f_r\}) = 3/4$, and

$$m(\{x : \ a_r \neq f_r \text{ and } a_{r+1} \neq f_{r+1}\}) = (3/4)^2, \qquad \text{etc.}$$

Thus $m(\{x : \ a_k \neq f_k, \ \text{all } k \geq r\}) = 0$, and E_0 is a countable union of such sets, from which the lemma follows.

There is a similar statement for points in the square $S = I \times I$ in terms of the representation (1).

Note that as a result the set of points in I corresponding to chains in $\mathcal{N}_1$ form a set of measure zero. In fact, we may use the lemma for the sequence for which $f_k = 1$, for all k, since the elements of $\mathcal{N}_1$ correspond to sequences $\{a_k\}$ with $a_k = 0$ for all sufficiently large k, or $a_k = 3$ for all sufficiently large k.

Similarly, the points in the square S corresponding to $\mathcal{N}_2$ form a set of measure zero. To see this, take for example $f_k = 1$ for k odd, and $f_k = 2$

for k even, and note that $\mathcal{N}_2$ corresponds to all sequences $\{a_k\}$ where one of the following four exclusive alternatives holds for all sufficiently large k: either a_k is 0 or 1; or a_k is 2 or 3; or a_k is 0 or 2; or a_k is 1 or 3. By similar reasoning the points $\Phi^{-1}(\mathcal{N}_2)$ and $\Phi(\mathcal{N}_1)$ form sets of measure zero in I and $I \times I$ respectively.

We now turn to the proof that Φ^* (which is a bijection from $I - Z_1$ to $(I \times I) - Z_2$) is measure preserving. For this it is useful to recall Theorem 1.4 in Chapter 1, whereby any open set $\mathcal{O}$ in the unit interval I can be realized as a countable union $\bigcup_{j=1}^{\infty} I_j$, where each I_j is a closed interval and the I_j have disjoint interiors. Moreover, an examination of the proof shows that the intervals can be taken to be dyadic, that is, of the form $[\ell/2^j, (\ell+1)/2^j]$, for appropriate integers ℓ and j. Further, such an interval is itself a quartic interval if j is even, $j = 2k$, or the union of two quartic intervals $[(2\ell)/2^{2k}, (2\ell+1)/2^{2k}]$ and $[(2\ell+1)/2^{2k}, (2\ell+2)/2^{2k}]$, if j is odd, $j = 2k-1$. Thus any open set in I can be given as a union of quartic intervals whose interiors are disjoint. Similarly, any open set in the square $I \times I$ is a union of dyadic squares whose interiors are disjoint.

Now let E be any set of measure zero in $I - Z_1$ and $\epsilon > 0$. Then we can cover $E \subset \bigcup_j I_j$, where I_j are quartic intervals and $\sum_j m_1(I_j) < \epsilon$. Because $\Phi^*(E) \subset \bigcup_j \Phi^*(I_j)$, then

$$m_2(\Phi^*(E)) \leq \sum m_2(\Phi^*(I_j)) = \sum m_1(I_j) < \epsilon.$$

Thus $\Phi^*(E)$ is measurable and $m_2(\Phi^*(E)) = 0$. Similarly, $(\Phi^*)^{-1}$ maps sets of measure zero in $(I \times I) - Z_2$ to sets of measure zero in I.

Now the argument above also shows that if $\mathcal{O}$ is any open set in I, then $\Phi^*(\mathcal{O} - Z_1)$ is measurable, and $m_2(\Phi^*(\mathcal{O} - Z_1)) = m_1(\mathcal{O})$. Thus this identity goes over to G_δ sets in I. Since any measurable set differs from a G_δ set by a set of measure zero, we see that we have established that $m_2(\Phi^*(E)) = m_1(E)$ for any measurable subset of E of $I - Z_1$. The same argument can be applied to $(\Phi^*)^{-1}$, and this completes the proof of the theorem.

The Peano mapping will be obtained as Φ^* for a special correspondence Φ.

3.3 Construction of the Peano mapping

The particular dyadic correspondence we now present provides us with the steps to follow when tracing the approximations of the Peano curve. The main idea behind its construction is that as we go from one quartic interval in the k^{th} generation to the next quartic interval in the same generation, we move from a dyadic square of the k^{th} generation to another square of the k^{th} generation that shares a common side.

More precisely, we say that two quartic intervals in the same generation are **adjacent** if they share a point in common. Also, two squares in the same generation are **adjacent** if they share a side in common.

Lemma 3.7 *There is a unique dyadic correspondence Φ so that:*

(i) *If I and J are two adjacent intervals of the same generation, then $\Phi(I)$ and $\Phi(J)$ are two adjacent squares (of the same generation).*

(ii) *In generation k, if I_- is the left-most interval and I_+ the right-most interval, then $\Phi(I_-)$ is the left-lower square and $\Phi(I_+)$ is the right-lower square.*

Part (ii) of the lemma is illustrated in Figure 9.

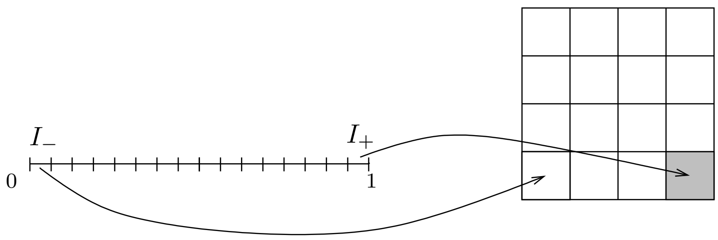

Figure 9. Special dyadic correspondence

Given a square S and its four immediate sub-squares, an acceptable **traverse** is an ordering of the sub-squares S_1, S_2, S_3, and S_4, so that S_j and S_{j+1} are adjacent for $j = 1, 2, 3$. With such an ordering, we note that if we color S_1 white, and then alternate black and white, the square S_3 is also white, while S_2 and S_4 are black. The important point to remember is that if the first square in a traverse is white, then the last square is black.

The key observation is the following. Suppose we are given a square S, and a side σ of S. If S_1 is any of the immediate four sub-squares in S, then there exists a unique traverse S_1, S_2, S_3, and S_4 so that the last square S_4 has a side in common with σ. With the initial square S_1 in the lower-left corner of S, the four possibilities which correspond to the four choices of σ, are illustrated in Figure 10.

We may now begin the inductive description of the dyadic correspondence satisfying the conditions in the lemma. On quartic intervals of the first generation we assign the square $S_j = \Phi(I_j)$, as pictured in Figure 11.

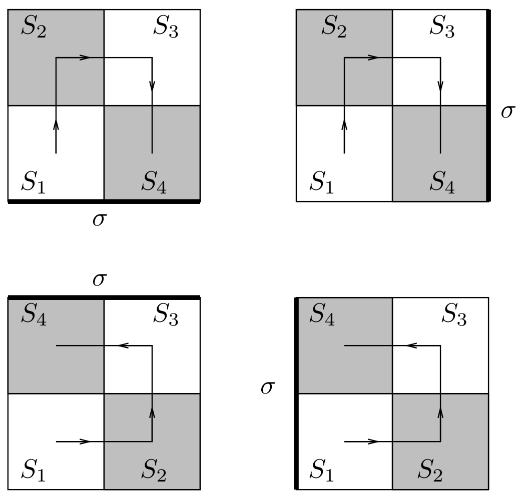

Figure 10. Traverses

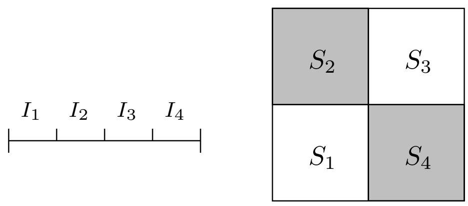

Figure 11. Initial step of the correspondence

Now suppose Φ has been defined for all quartic intervals of generation less than or equal to k. We now write the intervals in generation k in increasing order as $I_1, \ldots, I_{4^k}$, and let $S_j = \Phi(I_j)$. We then divide I_1 into four quartic intervals of generation $k+1$ and denote them by $I_{1,1}$, $I_{1,2}$, $I_{1,3}$, and $I_{1,4}$, where the intervals are chosen in increasing order.

Then, we assign to each interval $I_{1,j}$ a dyadic square $\Phi(I_{1,j}) = S_j$ of generation $k+1$ contained in S_1 so that:

(a) $S_{1,1}$ is the lower-left sub-square of S_1,

(b) $S_{1,4}$ touches the side that S_1 shares with S_2,

(c) $S_{1,1}$, $S_{1,2}$, $S_{1,3}$, and $S_{1,4}$ is a traverse.

This is possible, since the induction hypothesis guarantees that S_2 is adjacent to S_1.

This settles the assignments for the sub-squares of S_1, so we now turn our attention to S_2. Let $I_{2,1}$, $I_{2,2}$, $I_{2,3}$, and $I_{2,4}$ denote the quartic intervals of generation $k+1$ in I_2, written in increasing order. First, we take $S_{2,1} = \Phi(I_{2,1})$ to be the sub-square of S_2 which is adjacent to $S_{1,4}$. This can be done because $S_{1,4}$ touches S_2 by construction. Note that we leave S_1 from a black square $(S_{1,4})$, and enter S_2 in a white square $(S_{2,1})$. Since S_3 is adjacent to S_2, we may now find a traverse $S_{2,1}$, $S_{2,2}$, $S_{2,3}$ and $S_{2,4}$ so that $S_{2,4}$ touches S_3.

We may then repeat this process in each interval I_j and square S_j, $j = 3, \ldots, 4^k$. Note that at each stage the square $S_{j,1}$ (the "entering" square) is white, while $S_{j,4}$ (the "exiting" square) is black.

In the final step, the induction hypothesis guarantees that S_{4^k} is the lower-right corner square. Moreover, since S_{4^k-1} must be adjacent to S_{4^k} it must be either above it, or to the left of it, so we enter a square of the $(k+1)^{\text{st}}$ generation along an upper or left side. The entering square is a white square, and we traverse to the lower right corner sub-square of S_{4^k}, which is a black square.

This concludes the inductive step, hence the proof of Lemma 3.7.

We may now begin the actual description of the Peano curve. For each generation k we construct a polygonal line which consists of vertical and horizontal line segments connecting the centers of consecutive squares. More precisely, let Φ denote the dyadic correspondence in Lemma 3.7, and let $S_1, \ldots, S_{4^k}$ be the squares of the k^{th} generation ordered according to Φ, that is, $\Phi(I_j) = S_j$. Let t_j denote the middle point of I_j,

$$t_j = \frac{j - \frac{1}{2}}{4^k} \quad \text{for } j = 1, \ldots, 4^k.$$

Let x_j be the center of the square S_j, and define

$$\mathcal{P}_k(t_j) = x_j.$$

Also set

$$\mathcal{P}_k(0) = (0, 1/2^{k+1}) = x_0 \quad \text{where } t_0 = 0,$$

and

$$\mathcal{P}_k(1) = (1, 1/2^{k+1}) = x_{4^k+1} \quad \text{where } t_{4^k+1} = 1.$$

Then, we extend $\mathcal{P}_k(t)$ to the unit interval $0 \leq t \leq 1$ by linearity along the sub-intervals determined by the division points $t_0, \ldots, t_{4^k+1}$.

Note that the distance $|x_j - x_{j+1}| = 1/2^k$, while $|t_j - t_{j+1}| = 1/4^k$ for $0 \le j \le 4^k$. Also

$$|x_1 - x_0| = |x_{4^k} - x_{4^k+1}| = \frac{1}{2 \cdot 2^k},$$

while

$$|t_1 - t_0| = |t_{4^k} - t_{4^k+1}| = \frac{1}{2 \cdot 4^k}.$$

Therefore $\mathcal{P}_k'(t) = 4^k 2^{-k} = 2^k$ except when $t = t_j$.

As a result,

$$|\mathcal{P}_k(t) - \mathcal{P}_k(s)| \le 2^k |t - s|.$$

However,

$$|\mathcal{P}_{k+1}(t) - \mathcal{P}_k(t)| \le \sqrt{2}\, 2^{-k},$$

because when $\ell/4^k \le t \le (\ell+1)/4^k$, then $\mathcal{P}_{k+1}(t)$ and $\mathcal{P}_k(t)$ belong to the same dyadic square of generation k.

Therefore the limit

$$\mathcal{P}(t) = \lim_{k\to\infty} \mathcal{P}_k(t) = \mathcal{P}_1(t) + \sum_{j=1}^{\infty} \mathcal{P}_{j+1}(t) - \mathcal{P}_j(t)$$

exists, and defines a continuous function in view of the uniform convergence. By Lemma 2.8 we conclude that

$$|\mathcal{P}(t) - \mathcal{P}(s)| \le M|t - s|^{1/2},$$

and $\mathcal{P}$ satisfies a Lipschitz condition of exponent of $1/2$.

Moreover, each $\mathcal{P}_k(t)$ visits each dyadic square of generation k as t ranges in $[0, 1]$. Hence $\mathcal{P}$ is dense in the unit square, and by continuity we find that $t \mapsto \mathcal{P}(t)$ is a surjection.

Finally, to prove the measure preserving property of $\mathcal{P}$, it suffices to establish $\mathcal{P} = \Phi^*$.

Lemma 3.8 *If Φ is the dyadic correspondence in Lemma 3.7, then $\Phi^*(t) = \mathcal{P}(t)$ for every $0 \le t \le 1$.*

Proof. First, we observe that $\Phi^*(t)$ is unambiguously defined for every t. Indeed, suppose $t \in \bigcap_k I^k$ and $t \in \bigcap_k J^k$ are two chains of quartic intervals; then I^k and J^k must be adjacent for sufficiently large

k. Thus $\Phi(I^k)$ and $\Phi(J^k)$ must be adjacent squares for all sufficiently large k. Hence

$$\bigcap_k \Phi(I^k) = \bigcap_k \Phi(J^k).$$

Next, directly from our construction we have

$$\bigcap_k \Phi(I^k) = \lim \mathcal{P}_k(t) = \mathcal{P}(t).$$

This gives the desired conclusion.

The argument also shows that $\mathcal{P}(I) = \Phi(I)$ for any quartic interval I. Now recall that any interval (a, b) can be written as $\bigcup_j I_j$, where the I_j are quartic intervals with disjoint interiors. Because $\mathcal{P}(I_j) = \Phi(I_j)$, these are then dyadic squares with disjoint interiors. Since $\mathcal{P}(a, b) = \bigcup_k \mathcal{P}(I_j)$, we have

$$m_2(\mathcal{P}(a,b)) = \sum_{j=1}^{\infty} m_2(\mathcal{P}(I_j)) = \sum_{j=1}^{\infty} m_2(\Phi(I_j)) = \sum_{j=1}^{\infty} m_1(I_j) = m_1(a,b).$$

This proves conclusion (iii) of Theorem 3.1. The other conclusions having already been established, we need only note that the corollary is contained in Theorem 3.5.

As a result, we conclude that $t \mapsto \mathcal{P}(t)$ also induces a measure preserving mapping from $[0,1]$ to $[0,1] \times [0,1]$. This concludes the proof of Theorem 3.1.

4* Besicovitch sets and regularity

We begin by presenting a surprising regularity property enjoyed by all measurable subsets (of finite measure) of $\mathbb{R}^d$ when $d \geq 3$. As we shall see, the fact that the corresponding phenomenon does not hold for $d = 2$ is due to the existence of a remarkable set that was discovered by Besicovitch. A construction of a set of this kind will be detailed in Section 4.4.

We first fix some notation. For each unit vector γ on the sphere, $\gamma \in S^{d-1}$, and each $t \in \mathbb{R}$ we consider the **plane** $\mathcal{P}_{t,\gamma}$, which is defined as the $(d-1)$-dimensional affine hyperplane perpendicular to γ and of "signed distance" t from the origin.[1] The plane $\mathcal{P}_{t,\gamma}$ is given by

$$\mathcal{P}_{t,\gamma} = \{x \in \mathbb{R}^d : \ x \cdot \gamma = t\}.$$

[1]Note that there are two planes perpendicular to γ and of distance $|t|$ from the origin; this accounts for the fact that t may be either positive or negative.

We observe that each $\mathcal{P}_{t,\gamma}$ carries a natural $(d-1)$ Lebesgue measure, denoted by m_{d-1}. In fact, if we complete γ to an orthonormal basis $e_1, e_2, \ldots, e_{d-1}, \gamma$ of $\mathbb{R}^d$, then we can write any $x \in \mathbb{R}^d$ in terms of the corresponding coordinates as $x = x_1e_1 + x_2e_2 + \cdots + x_d\gamma$. When we set $x \in \mathbb{R}^d = \mathbb{R}^{d-1} \times \mathbb{R}$ with $(x_1, \ldots, x_{d-1}) \in \mathbb{R}^{d-1}$, $x_d \in \mathbb{R}$, then the measure m_{d-1} on $\mathcal{P}_{t,\gamma}$ is the Lebesgue measure on $\mathbb{R}^{d-1}$. This definition of m_{d-1} is independent of the choice of orthonormal vectors $e_1, e_2, \ldots, e_{d-1}$, since Lebesgue measure is invariant under rotations. (See Problem 4, Chapter 2, or Exercise 26, Chapter 3.)

With these preliminaries out of the way, we define for each subset $E \subset \mathbb{R}^d$ the **slice** of E cut out by the plane $\mathcal{P}_{t,\gamma}$ as

$$E_{t,\gamma} = E \cap \mathcal{P}_{t,\gamma}.$$

We now consider the slices $E_{t,\gamma}$ as t varies, where E is measurable and γ is fixed. (See Figure 12.)

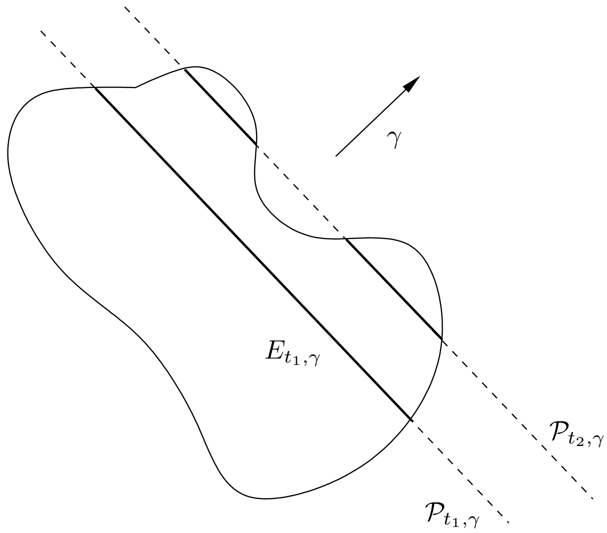

Figure 12. The slices $E \cap \mathcal{P}_{t,\gamma}$ as t varies

We observe that for almost every t the set $E_{t,\gamma}$ is m_{d-1} measurable and, moreover, $m_{d-1}(E_{t,\gamma})$ is a measurable function of t. This is a direct consequence of Fubini's theorem and the above decomposition, $\mathbb{R}^d = \mathbb{R}^{d-1} \times \mathbb{R}$. In fact, so long as the direction γ is pre-assigned, not much more can be said in general about the function $t \mapsto m_{d-1}(E_{t,\gamma})$.

However, when $d \geq 3$ the nature of the function is dramatically different for "most" γ. This is contained in the following theorem.

Theorem 4.1 *Suppose E is of finite measure in $\mathbb{R}^d$, with $d \geq 3$. Then for almost every $\gamma \in S^{d-1}$:*

(i) *$E_{t,\gamma}$ is measurable for all $t \in \mathbb{R}$.*

(ii) *$m_{d-1}(E_{t,\gamma})$ is continuous in $t \in \mathbb{R}$.*

Moreover, the function of t defined by $\mu(t, \gamma) = m_{d-1}(E_{t,\gamma})$ satisfies a Lipschitz condition with exponent α for any α with $0 < \alpha < 1/2$.

The almost everywhere assertion is with respect to the natural measure $d\sigma$ on S^{d-1} that arises in the polar coordinate formula in Section 3.2 of the previous chapter.

We recall that a function f is Lipschitz with exponent α if

$$|f(t_1) - f(t_2)| \leq A|t_1 - t_2|^\alpha \quad \text{for some } A.$$

A significant part of (i) is that for a.e. γ, the slice $E_{t,\gamma}$ is measurable for *all* values of the parameter t. In particular, one has the following.

Corollary 4.2 *Suppose E is a set of measure zero in $\mathbb{R}^d$ with $d \geq 3$. Then, for almost every $\gamma \in S^{d-1}$, the slice $E_{t,\gamma}$ has zero measure for all $t \in \mathbb{R}$.*

The fact that there is no analogue of this when $d = 2$ is a consequence of the existence of a **Besicovitch set**, (also called a "Kakeya set"), which is defined as a set that satisfies the three conditions in the theorem below.

Theorem 4.3 *There exists a set $\mathcal{B}$ in $\mathbb{R}^2$ that:*

(i) *is compact,*

(ii) *has Lebesgue measure zero,*

(iii) *contains a translate of every unit line segment.*

Note that with $F = \mathcal{B}$ and $\gamma \in S^1$ one has $m_1(F \cap \mathcal{P}_{t_0,\gamma}) \geq 1$ for some t_0. If $m_1(F \cap \mathcal{P}_{t,\gamma})$ were continuous in t, then this measure would be strictly positive for an interval in t containing t_0, and thus we would have $m_2(F) > 0$, by Fubini's theorem. This contradiction shows that the analogue of Theorem 4.1 cannot hold for $d = 2$.

While the set $\mathcal{B}$ has zero two-dimensional measure, this assertion cannot be improved by replacing this measure by α-dimensional Hausdorff measure, with $\alpha < 2$.

Theorem 4.4 *Suppose F is any set that satisfies the conclusions* (i) *and* (iii) *of Theorem* 4.3. *Then F has Hausdorff dimension* 2.

4.1 The Radon transform

Theorems 4.1 and 4.4 will be derived by an analysis of the regularity properties of the Radon transform $\mathcal{R}$. The operator $\mathcal{R}$ arises in a number of problems in analysis, and was already considered in Chapter 6 of Book I.

For an appropriate function f on $\mathbb{R}^d$, the **Radon transform** of f is defined by

$$\mathcal{R}(f)(t,\gamma)=\int_{\mathcal{P}_{t,\gamma}} f.$$

The integration is performed over the plane $\mathcal{P}_{t,\gamma}$ with respect to the measure m_{d-1} discussed above. We first make the following simple observation:

1. If f is continuous and has compact support, then f is of course integrable on *every* plane $\mathcal{P}_{t,\gamma}$, and so $\mathcal{R}(f)(t,\gamma)$ is defined for all $(t,\gamma)\in\mathbb{R}\times S^{d-1}$. Moreover it is a continuous function of the pair (t,γ) and has compact support in the t-variable.

2. If f is merely Lebesgue integrable, then f may fail to be measurable or integrable on $\mathcal{P}_{t,\gamma}$ for some (t,γ), and thus $\mathcal{R}(f)(t,\gamma)$ is not defined for those (t,γ).

3. Suppose f is the characteristic function of the set E, that is, $f=\chi_E$. Then $\mathcal{R}(f)(t,\gamma)=m_{d-1}(E_{t,\gamma})$ if $E_{t,\gamma}$ is measurable.

It is this last property that links the Radon transform to our problem. Key estimates in this conclusion involve a maximal "Radon transform" defined by

$$\mathcal{R}^*(f)(\gamma)=\sup_{t\in\mathbb{R}}|\mathcal{R}(f)(t,\gamma)|,$$

as well as corresponding expressions controlling the Lipschitz character of $\mathcal{R}(f)(t,\gamma)$ as a function of t. A basic fact inherent in our analysis is that the regularity of the Radon transform actually improves as the dimension of the underlying space increases.

Theorem 4.5 *Suppose f is continuous and has compact support in $\mathbb{R}^d$ with $d\geq 3$. Then*

$$\int_{S^{d-1}}\mathcal{R}^*(f)(\gamma)\,d\sigma(\gamma)\leq c\left[\|f\|_{L^1(\mathbb{R}^d)}+\|f\|_{L^2(\mathbb{R}^d)}\right] \tag{2}$$

for some constant $c>0$ that does not depend on f.

An inequality of this type is a typical "a priori" estimate. It is obtained first under some regularity assumption on the function f, and then a limiting argument allows one to pass to the more general case when f belongs to $L^1 \cap L^2$.

We make some comments about the appearance of both the L^1-norm and L^2-norm in (2). The L^2-norm imposes a crucial local control of the kind that is necessary for the desired regularity. (See Exercise 27.) However, without some restriction on f of a global nature, the function f might fail to be integrable on *any* plane $\mathcal{P}_{t,\gamma}$, as the example $f(x) = 1/(1+|x|^{d-1})$ shows. Note that this function belongs to $L^2(\mathbb{R}^d)$ if $d \geq 3$, but not to $L^1(\mathbb{R}^d)$.

The proof of Theorem 4.5 actually gives an essentially stronger result, which we state as a corollary.

Corollary 4.6 *Suppose f is continuous and has compact support in $\mathbb{R}^d$, $d \geq 3$. Then for any α, $0 < \alpha < 1/2$, the inequality* (2) *holds with $\mathcal{R}^*(f)(\gamma)$ replaced by*

$$\sup_{t_1 \neq t_2} \frac{|\mathcal{R}(f)(t_1,\gamma) - \mathcal{R}(f)(t_2,\gamma)|}{|t_1 - t_2|^\alpha}. \tag{3}$$

The proof of the theorem relies on the interplay between the Radon transform and the Fourier transform.

For fixed $\gamma \in S^{d-1}$, we let $\hat{\mathcal{R}}(f)(\lambda,\gamma)$ denote the Fourier transform of $\mathcal{R}(f)(t,\gamma)$ in the t-variable

$$\hat{\mathcal{R}}(f)(\lambda,\gamma) = \int_{-\infty}^{\infty} \mathcal{R}(f)(t,\gamma) e^{-2\pi i \lambda t}\, dt.$$

In particular, we use $\lambda \in \mathbb{R}$ to denote the dual variable of t.

We also write $\hat{f}$ for the Fourier transform of f as a function on $\mathbb{R}^d$, namely

$$\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x) e^{-2\pi i x\cdot\xi}\, dx.$$

Lemma 4.7 *If f is continuous with compact support, then for every $\gamma \in S^{d-1}$ we have*

$$\hat{\mathcal{R}}(f)(\lambda,\gamma) = \hat{f}(\lambda\gamma).$$

The right-hand side is just the Fourier transform of f evaluated at the point $\lambda\gamma$.

Proof. For each unit vector γ we use the adapted coordinate system described above: $x = (x_1, \ldots, x_d)$ where γ coincides with the x_d direction. We can then write each $x \in \mathbb{R}^d$ as $x = (u, t)$ with $u \in \mathbb{R}^{d-1}$, $t \in \mathbb{R}$, where $x \cdot \gamma = t = x_d$ and $u = (x_1, \ldots, x_{d-1})$. Moreover

$$\int_{\mathcal{P}_{t,\gamma}} f = \int_{\mathbb{R}^{d-1}} f(u,t)\, du,$$

and Fubini's theorem shows that $\int_{\mathbb{R}^d} f(x)\, dx = \int_{-\infty}^{\infty} \left(\int_{\mathcal{P}_{t,\gamma}} f \right) dt$. Applying this to $f(x)e^{-2\pi i x\cdot(\lambda\gamma)}$ in place of $f(x)$ gives

$$\begin{aligned}\hat{f}(\lambda\gamma) &= \int_{\mathbb{R}^d} f(x) e^{-2\pi i x\cdot(\lambda\gamma)}\, dx = \int_{-\infty}^{\infty} \left(\int_{\mathbb{R}^{d-1}} f(u,t)\, du \right) e^{-2\pi i \lambda t}\, dt \\ &= \int_{-\infty}^{\infty} \left(\int_{\mathcal{P}_{t,\gamma}} f \right) e^{-2\pi i \lambda t}\, dt.\end{aligned}$$

Therefore $\hat{f}(\lambda\gamma) = \hat{\mathcal{R}}(f)(\lambda, \gamma)$, and the lemma is proved.

Lemma 4.8 *If f is continuous with compact support, then*

$$\int_{S^{d-1}} \left(\int_{-\infty}^{\infty} |\hat{\mathcal{R}}(f)(\lambda,\gamma)|^2 |\lambda|^{d-1} d\lambda \right) d\sigma(\gamma) = 2 \int_{\mathbb{R}^d} |f(x)|^2 dx.$$

Let us observe the crucial point that the greater the dimension d, the larger the factor $|\lambda|^{d-1}$ as $|\lambda|$ tends to infinity. Hence the greater the dimension, the better the decay of the Fourier transform $\hat{\mathcal{R}}(f)(\lambda, \gamma)$, and so the better the regularity of the Radon transform $\mathcal{R}(f)(t, \gamma)$ as a function of t.

Proof. The Plancherel formula in Chapter 5 guarantees that

$$2 \int_{\mathbb{R}^d} |f(x)|^2\, dx = 2 \int_{\mathbb{R}^d} |\hat{f}(\xi)|^2\, d\xi.$$

Changing to polar coordinates $\xi = \lambda\gamma$ where $\lambda > 0$ and $\gamma \in S^{d-1}$, we obtain

$$2 \int_{\mathbb{R}^d} |\hat{f}(\xi)|^2\, d\xi = 2 \int_{S^{d-1}} \int_0^{\infty} |\hat{f}(\lambda\gamma)|^2 \lambda^{d-1}\, d\lambda\, d\sigma(\gamma).$$

We now observe that a simple change of variables provides

$$\int_{S^{d-1}} \int_0^{\infty} |\hat{f}(\lambda\gamma)|^2 \lambda^{d-1}\, d\lambda\, d\sigma(\gamma) = \int_{S^{d-1}} \int_{-\infty}^{0} |\hat{f}(\lambda\gamma)|^2 |\lambda|^{d-1}\, d\lambda\, d\sigma(\gamma),$$

and the proof is complete once we invoke the result of Lemma 4.7.

The final ingredient in the proof of Theorem 4.5 consists of the following:

Lemma 4.9 *Suppose*

$$F(t) = \int_{-\infty}^{\infty} \hat{F}(\lambda) e^{2\pi i \lambda t}\, d\lambda,$$

where

$$\sup_{\lambda \in \mathbb{R}} |\hat{F}(\lambda)| \leq A \quad \text{and} \quad \int_{-\infty}^{\infty} |\hat{F}(\lambda)|^2 |\lambda|^{d-1} d\lambda \leq B^2.$$

Then

$$\sup_{t \in \mathbb{R}} |F(t)| \leq c(A + B). \tag{4}$$

Moreover, if $0 < \alpha < 1/2$*, then*

$$|F(t_1) - F(t_2)| \leq c_\alpha |t_1 - t_2|^\alpha (A + B) \quad \text{for all } t_1,\ t_2. \tag{5}$$

Proof. The first inequality is obtained by considering separately the two cases $|\lambda| \leq 1$ and $|\lambda| > 1$. We write

$$F(t) = \int_{|\lambda| \leq 1} \hat{F}(\lambda) e^{2\pi i \lambda t}\, d\lambda + \int_{|\lambda| > 1} \hat{F}(\lambda) e^{2\pi i \lambda t}\, d\lambda.$$

Clearly, the first integral is bounded by cA. To estimate the second integral it suffices to bound $\int_{|\lambda|>1} |\hat{F}(\lambda)|\, d\lambda$. An application of the Cauchy-Schwarz inequality gives

$$\int_{|\lambda|>1} |\hat{F}(\lambda)| d\lambda \leq \left(\int_{|\lambda|>1} |\hat{F}(\lambda)|^2 |\lambda|^{d-1} d\lambda \right)^{1/2} \left(\int_{|\lambda|>1} |\lambda|^{-d+1} d\lambda \right)^{1/2}.$$

This last integral is convergent precisely when $-d + 1 < -1$, which is equivalent to $d > 2$, namely $d \geq 3$, which we assume. Hence $|F(t)| \leq c(A + B)$ as desired.

To establish Lipschitz continuity, we first note that

$$F(t_1) - F(t_2) = \int_{-\infty}^{\infty} \hat{F}(\lambda) \left[e^{2\pi i \lambda t_1} - e^{2\pi i \lambda t_2} \right] d\lambda.$$

Since one has the inequality[2] $|e^{ix}-1| \le |x|$, we immediately see that

$$|e^{2\pi i\lambda t_1} - e^{2\pi i\lambda t_2}| \le c|t_1 - t_2|^\alpha \lambda^\alpha \quad \text{if } 0 \le \alpha < 1.$$

We may then write the difference $F(t_1) - F(t_2)$ as a sum of two integrals. The integral over $|\lambda| \le 1$ is clearly bounded by $cA|t_1 - t_2|^\alpha$. The second integral, the one over $|\lambda| > 1$, can be estimated from above by

$$|t_1 - t_2|^\alpha \int_{|\lambda|>1} |\hat{F}(\lambda)||\lambda|^\alpha \, d\lambda.$$

An application of the Cauchy-Schwarz inequality show that this last integral is majorized by

$$\left(\int_{|\lambda|>1} |\hat{F}(\lambda)|^2 |\lambda|^{d-1} \, d\lambda\right)^{1/2} \left(\int_{|\lambda|>1} |\lambda|^{-d+1+2\alpha} \, d\lambda\right)^{1/2} \le c_\alpha B,$$

since the second integral is finite if $-d+1+2\alpha < -1$, and in particular this holds if $\alpha < 1/2$ when $d \ge 3$. This concludes the proof of the lemma.

We now gather these results to prove the theorem. For each $\gamma \in S^{d-1}$ let

$$F(t) = \mathcal{R}(f)(t, \gamma).$$

Note that with this definition we have

$$\sup_{t\in\mathbb{R}} |F(t)| = \mathcal{R}^*(f)(\gamma).$$

Let

$$A(\gamma) = \sup_\lambda |\hat{F}(\lambda)| \quad \text{and} \quad B^2(\gamma) = \int_{-\infty}^{\infty} |\hat{F}(\lambda)|^2 |\lambda|^{d-1} \, d\lambda.$$

Then by (4)

$$\sup_{t\in\mathbb{R}} |F(t)| \le c(A(\gamma) + B(\gamma)).$$

However, we observed that $\hat{F}(\lambda) = \hat{f}(\lambda\gamma)$, and hence

$$A(\gamma) \le \|f\|_{L^1(\mathbb{R}^d)}.$$

[2]The distance in the plane from the point e^{ix} to the point 1 is shorter than the length of the arc on the unit circle joining them.

Therefore,

$$|\mathcal{R}^*(f)(\gamma)|^2 \leq c(A(\gamma)^2 + B(\gamma)^2),$$

and thus

$$\int_{S^{d-1}} |\mathcal{R}^*(f)(\gamma)|^2 \, d\sigma(\gamma) \leq c(\|f\|^2_{L^1(\mathbb{R}^d)} + \|f\|^2_{L^2(\mathbb{R}^d)}),$$

since $\int B^2(\gamma) \, d\sigma(\gamma) = 2\|f\|^2_{L^2}$ by Lemma 4.8. Consequently,

$$\int_{S^{d-1}} \mathcal{R}^*(f)(\gamma) \, d\sigma(\gamma) \leq c(\|f\|_{L^1(\mathbb{R}^d)} + \|f\|_{L^2(\mathbb{R}^d)}).$$

Note that the identity we have used,

$$\mathcal{R}(f)(t, \gamma) = \int_{-\infty}^{\infty} \hat{F}(\lambda) e^{2\pi i \lambda t} \, d\lambda,$$

with $F(t) = \mathcal{R}(f)(t, \gamma)$, is justified for almost every $\gamma \in S^{d-1}$ by the Fourier inversion result in Theorem 4.2 of Chapter 2. Indeed, we have seen that $A(\gamma)$ and $B(\gamma)$ are finite for almost every γ, and thus $\hat{F}$ is integrable for those γ. This completes the proof of the theorem. The corollary follows the same way if we use (5) instead of (4).

We now return to the situation in the plane to see what information we may deduce from the above analysis. The inequality (2) as it stands does not hold when $d = 2$. However, a modification of it does hold, and this will be used in the proof of Theorem 4.4.

If $f \in L^1(\mathbb{R}^d)$ we define

$$\begin{aligned} \mathcal{R}_\delta(f)(t, \gamma) &= \frac{1}{2\delta} \int_{t-\delta}^{t+\delta} \mathcal{R}(f)(s, \gamma) \, ds \\ &= \frac{1}{2\delta} \int_{t-\delta \leq x \cdot \gamma \leq t+\delta} f(x) \, dx. \end{aligned}$$

In this definition of $\mathcal{R}_\delta(f)(t, \gamma)$ we integrate the function f in a small "strip" of width 2δ around the plane $\mathcal{P}_{t,\gamma}$. Thus $\mathcal{R}_\delta$ is an average of Radon transforms.

We let

$$\mathcal{R}^*_\delta(f)(\gamma) = \sup_{t \in \mathbb{R}} |\mathcal{R}_\delta(f)(t, \gamma)|.$$

Theorem 4.10 *If f is continuous with compact support, then*

$$\int_{S^1} \mathcal{R}_\delta^*(f)(\gamma)\, d\sigma(\gamma) \leq c(\log 1/\delta)^{1/2} \left(\|f\|_{L^1(\mathbb{R}^2)} + \|f\|_{L^2(\mathbb{R}^2)}\right)$$

when $0 < \delta \leq 1/2$.

The same argument as in the proof of Theorem 4.5 applies here, except that we need a modified version of Lemma 4.9. More precisely, let us set

$$F_\delta(t) = \int_{-\infty}^{\infty} \hat{F}(\lambda) \left(\frac{e^{2\pi i(t+\delta)\lambda} - e^{2\pi i(t-\delta)\lambda}}{2\pi i\lambda(2\delta)} \right) d\lambda,$$

and suppose that

$$\sup_\lambda |\hat{F}(\lambda)| \leq A \quad \text{and} \quad \int_{-\infty}^{\infty} |\hat{F}(\lambda)|^2 |\lambda|\, d\lambda \leq B.$$

Then we claim that

$$\sup_t |F_\delta(t)| \leq c(\log 1/\delta)^{1/2}(A+B). \tag{6}$$

Indeed, we use the fact that $|(\sin x)/x| \leq 1$ to see that, in the definition of $F_\delta(t)$, the integral over $|\lambda| \leq 1$ gives the cA. Also, the integral over $|\lambda| > 1$ can be split and is bounded by the sum

$$\int_{1<|\lambda|\leq 1/\delta} |\hat{F}(\lambda)|\, d\lambda + \frac{c}{\delta} \int_{1/\delta \leq |\lambda|} |\hat{F}(\lambda)||\lambda|^{-1}\, d\lambda.$$

The first integral above can be estimated by the Cauchy-Schwarz inequality, as follows

$$\begin{aligned}\int_{1<|\lambda|\leq 1/\delta} |\hat{F}(\lambda)|\, d\lambda &\leq c\left(\int_{1<|\lambda|\leq 1/\delta} |\hat{F}(\lambda)|^2|\lambda|\, d\lambda\right)^{1/2}\left(\int_{1<|\lambda|\leq 1/\delta} |\lambda|^{-1}\, d\lambda\right)^{1/2} \\ &\leq cB(\log 1/\delta)^{1/2}.\end{aligned}$$

Finally, we also note that

$$\begin{aligned}\frac{c}{\delta} \int_{1/\delta \leq |\lambda|} |\hat{F}(\lambda)||\lambda|^{-1}\, d\lambda &\leq c\left(\int_{1/\delta \leq |\lambda|} |\hat{F}(\lambda)|^2|\lambda|\, d\lambda\right)^{1/2} \frac{1}{\delta}\left(\int_{1/\delta\leq|\lambda|} |\lambda|^{-3}\, d\lambda\right)^{1/2} \\ &\leq cB\end{aligned}$$

and this establishes (6), and hence the theorem.

4.2 Regularity of sets when $d \geq 3$

We now extend to the general context the basic estimates for the Radon transform, proved for continuous functions of compact support. This will yield the regularity result formulated in Theorem 4.1.

Proposition 4.11 *Suppose $d \geq 3$, and let f belong to $L^1(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$. Then for a.e. $\gamma \in S^{d-1}$ we can assert the following:*

(a) *f is measurable and integrable on the plane $\mathcal{P}_{t,\gamma}$, for every $t \in \mathbb{R}$.*

(b) *The function $\mathcal{R}(f)(t,\gamma)$ is continuous in t and satisfies a Lipschitz condition with exponent α for each $\alpha < 1/2$. Moreover, the inequality (2) of Theorem 4.5 and its variant with (3) hold for f.*

We prove this in a series of steps.

Step 1. We consider $f = \chi_{\mathcal{O}}$, the characteristic function of a bounded open set $\mathcal{O}$. Here the assertion (a) is evident since $\mathcal{O} \cap \mathcal{P}_{t,\gamma}$ is an open and bounded set in $\mathcal{P}_{t,\gamma}$ and is bounded. Thus $\mathcal{R}(f)(t,\gamma)$ is defined for all (t,γ).

Next we can find a sequence $\{f_n\}$ of non-negative continuous functions of compact support so that for every x, $f_n(x)$ increases to $f(x)$ as $n \to \infty$. Thus $\mathcal{R}(f_n)(t,\gamma) \to \mathcal{R}(f)(t,\gamma)$ for every (t,γ) by the monotone convergence theorem, and also $\mathcal{R}^*(f_n)(\gamma) \to \mathcal{R}^*(f)(\gamma)$ for each $\gamma \in S^{d-1}$. As a result we see that the inequality (2) is valid for $f = \chi_{\mathcal{O}}$, with $\mathcal{O}$ open and bounded.

Step 2. We now consider $f = \chi_E$, where E is a set of measure zero, and take first the case when the set E is bounded. Then we can find a decreasing sequence $\{\mathcal{O}_n\}$ of open and bounded sets, such that $E \subset \mathcal{O}_n$, while $m(\mathcal{O}_n) \to 0$ as $n \to \infty$.

Let $\tilde{E} = \bigcap \mathcal{O}_n$. Since $\tilde{E} \cap \mathcal{P}_{t,\gamma}$ is measurable for every (t,γ), the functions $\mathcal{R}(\chi_{\tilde{E}})(t,\gamma)$ and $\mathcal{R}^*(\chi_{\tilde{E}})(\gamma)$ are well-defined. However, $\mathcal{R}^*(\chi_{\tilde{E}})(\gamma) \leq \mathcal{R}^*(\chi_{\mathcal{O}_n})(\gamma)$, while the $\mathcal{R}^*(\chi_{\mathcal{O}_n})$ decrease. Thus the inequality (2) we have just proved for $f = \chi_{\mathcal{O}_n}$ shows that $\mathcal{R}^*(\chi_{\tilde{E}})(\gamma) = 0$ for a.e. γ. The fact that $E \subset \tilde{E}$ then implies that for a.e. γ, the set $E \cap \mathcal{P}_{t,\gamma}$ has $(d-1)$-dimensional measure zero for every $t \in \mathbb{R}$. This conclusion immediately extends to the case when E is not necessarily bounded, by writing E as a countable union of bounded sets of measure zero. Therefore Corollary 4.2 is proved.

Step 3. Here we assume that f is a bounded measurable function supported on a bounded set. Then by familiar arguments we can find a sequence $\{f_n\}$ of continuous functions that are uniformly bounded, supported in a fixed compact set, and so that $f_n(x) \to f(x)$ a.e. By the bounded convergence theorem, $\|f_n - f\|_{L^1}$ and $\|f_n - f\|_{L^2}$ both tend to zero as $n \to \infty$, and upon selecting a subsequence if necessary, we can suppose that $\|f_n - f\|_{L^1} + \|f_n - f\|_{L^2} \leq 2^{-n}$. By what we have just proved in Step 2 we have, for a.e. $\gamma \in S^{d-1}$, that $f_n(x) \to f(x)$ on $\mathcal{P}_{t,\gamma}$ a.e. with respect to the measure m_{d-1}, for each $t \in \mathbb{R}$. Thus again by the bounded convergence theorem for those (t, γ), we see that $\mathcal{R}(f_n)(t,\gamma) \to \mathcal{R}(f)(t,\gamma)$, and this limit defines $\mathcal{R}(f)$. Now applying Theorem 4.5 to $f_n - f_{n-1}$ gives

$$\sum_{n=1}^{\infty} \int_{S^{d-1}} \mathcal{R}^*(f_n - f_{n-1})(\gamma)\, d\sigma(\gamma) \leq c \sum_{n=1}^{\infty} 2^{-n} < \infty.$$

This means that

$$\sum_n \sup_t |\mathcal{R}(f_n)(t,\gamma) - \mathcal{R}(f_{n-1})(t,\gamma)| < \infty,$$

for a.e.$\gamma \in S^{d-1}$, and hence for those γ the sequence of functions $\mathcal{R}(f_n)(t,\gamma)$ converges uniformly. As a consequence, for those γ the function $\mathcal{R}(f)(t,\gamma)$ is continuous in t, and the inequality (2) is valid for this f. The inequality with (3) is deduced in the same way.

Finally, we deal with the general f in $L^1 \cap L^2$ by approximating it by a sequence of bounded functions each with bounded support. The details of the argument are similar to the case treated above and are left to the reader.

Observe that the special case $f = \chi_E$ of the proposition gives us Theorem 4.1.

4.3 Besicovitch sets have dimension 2

Here we prove Theorem 4.4, that any Besicovitch set necessarily has Hausdorff dimension 2. We use Theorem 4.10, namely, the inequality

$$\int_{S^1} \mathcal{R}^*_\delta(f)(\gamma)\, d\sigma(\gamma) \leq c(\log 1/\delta)^{1/2} \left(\|f\|_{L^1(\mathbb{R}^2)} + \|f\|_{L^2(\mathbb{R}^2)} \right).$$

This inequality was proved under the assumption that f was continuous and had compact support. In the present situation it goes over without difficulty to the general case where $f \in L^1 \cap L^2$, by an easy limiting argument, since it is clear that $\mathcal{R}^*_\delta(f_n)(\gamma)$ converges to $\mathcal{R}^*_\delta(f)(\gamma)$ for all γ if $f_n \to f$ in the L^1-norm.

Now suppose F is a Besicovitch set and α is fixed with $0 < \alpha < 2$. Assume that $F \subset \bigcup_{i=1}^{\infty} B_i$ is a covering, where B_i are balls with diameter less than a given number. We must show that

$$\sum_i (\text{diam } B_i)^\alpha \geq c_\alpha > 0.$$

We proceed in two steps, considering first a simple situation that will make clear the idea of the proof.

Case 1. We suppose first that all the balls B_i have the same diameter δ (with $\delta \leq 1/2$) and also that there are only a finite number, say N, of balls in the covering. We must prove that $N\delta^\alpha \geq c_\alpha$.

Let B_i^* denote the double of B_i and $F^* = \bigcup_i B_i^*$. Then, we clearly have

$$m(F^*) \leq cN\delta^2.$$

Since F is a Besicovitch set, for each $\gamma \in S^1$ there is a segment s_γ of unit length, perpendicular to γ, and which is contained in F. Also, by construction, any translate by less than δ of a point in s_γ must belong to F^*. Hence

$$\mathcal{R}_\delta^*(\chi_{F^*})(\gamma) \geq 1 \qquad \text{for every } \gamma.$$

If we take $f = \chi_{F^*}$ in the inequality (6), and note that the Cauchy-Schwarz inequality implies

$$\|\chi_{F^*}\|_{L^1(\mathbb{R}^2)} \leq c\|\chi_{F^*}\|_{L^2(\mathbb{R}^2)} \leq c(m(F^*))^{1/2},$$

then we obtain

$$c \leq N^{1/2}\delta(\log 1/\delta)^{1/2}.$$

This implies $N\delta^\alpha \geq c$ for $\alpha < 2$.

Case 2. We now treat the general case. Suppose $F \subset \bigcup_{i=1}^{\infty} B_i$, where the balls B_i each have diameter less than 1. For each integer k let N_k be the number of balls in the collection $\{B_i\}$ for which

$$2^{-k-1} \leq \text{diam } B_i \leq 2^{-k}.$$

We need to show that $\sum_{k=0}^{\infty} N_k 2^{-k\alpha} \geq c_\alpha$. In fact, we shall prove the stronger result that there exists a positive integer k' such that $N_{k'}2^{-k'\alpha} \geq c_\alpha$.

Let

$$F_k = F \bigcap \left(\bigcup_{2^{-k-1} \leq \text{diam } B_i \leq 2^{-k}} B_i \right),$$

and let

$$F_k^* = \bigcup_{2^{-k-1} \leq \text{diam } B_i \leq 2^{-k}} B_i^*,$$

where B_i^* denotes the double of B_i. Then we note that

$$m_1(F_k^*) \leq cN_k 2^{-2k} \quad \text{ for all } k.$$

Since F is a Besicovitch set, for each $\gamma \in S^1$ there is a segment s_γ of unit length entirely contained in F. We now make precise the fact that for some k, a large proportion of s_γ belongs to F_k.

We pick a sequence of real numbers $\{a_k\}_{k=0}^{\infty}$ such that $0 \leq a_k \leq 1$, $\sum a_k = 1$, but a_k does not tend to zero too quickly. For instance, we may choose $a_k = c_\epsilon 2^{-\epsilon k}$ with $c_\epsilon = 1 - 2^{-\epsilon}$, and $\epsilon > 0$ but ϵ sufficiently small.

Then, for some k we must have

$$m_1(s_\gamma \cap F_k) \geq a_k.$$

Otherwise, since $F = \bigcup F_k$, we would have

$$m_1(s_\gamma \cap F) < \sum a_k = 1,$$

and this contradicts the fact that $m_1(s_\gamma \cap F) = 1$, since s_γ is entirely contained in F.

Therefore, with this k, we must have

$$\mathcal{R}^*_{2^{-k}}(\chi_{F_k^*})(\gamma) \geq a_k,$$

because any point of distance less than 2^{-k} from F_k must belong to F_k^*. Since the choice of k may depend on γ, we let

$$E_k = \{\gamma : \ \mathcal{R}^*_{2^{-k}}(\chi_{F_k^*})(\gamma) \geq a_k\}.$$

By our previous observations, we have

$$S^1 = \bigcup_{k=1}^{\infty} E_k,$$

and so for at least one k, which we denote by k', we have

$$m(E_{k'}) \geq 2\pi a_{k'},$$

for otherwise $m(S_1) < 2\pi \sum a_k = 2\pi$. As a result

$$\begin{aligned} 2\pi a_{k'}^2 &= 2\pi a_{k'} a_{k'} \\ &\leq \int_{E_{k'}} a_{k'}\, d\sigma(\gamma) \\ &\leq \int_{S_1} \mathcal{R}^*_{2^{-k'}}(\chi_{F^*_{k'}})(\gamma)\, d\sigma(\gamma). \end{aligned}$$

By the fundamental inequality (6) we get

$$a_{k'}^2 \leq c(\log 2^{k'})^{1/2} \|\chi_{F^*_{k'}}\|_{L^2(\mathbb{R}^2)}.$$

Recalling that by our choice $a_k \approx 2^{-\epsilon k}$, and noting that $\|\chi_{F^*_{k'}}\|_{L^2} \leq cN_{k'}^{1/2} 2^{-k'}$, we obtain

$$2^{(1-2\epsilon)k'} \leq c(\log 2^{k'})^{1/2} N_{k'}^{1/2}.$$

Finally, this last inequality guarantees that $N_{k'} 2^{-\alpha k'} \geq c_\alpha$ as long as $4\epsilon < 2 - \alpha$.

This concludes the proof of the theorem.

4.4 Construction of a Besicovitch set

There are a number of different constructions of Besicovitch sets. The one we have chosen to describe here involves the concept of self-replicating sets, an idea that permeates much of the discussion of this chapter.

We consider the Cantor set of constant dissection $\mathcal{C}_{1/2}$, which for simplicity we shall write as $\mathcal{C}$, and which is defined in Exercise 3, Chapter 1. Note that $\mathcal{C} = \bigcap_{k=0}^{\infty} C_k$, where $C_0 = [0, 1]$, and C_k is the union of 2^k closed intervals of length 4^{-k} obtained by removing from C_{k-1} the 2^{k-1} centrally situated open intervals of length $\frac{1}{2} \cdot 4^{-k+1}$. The set $\mathcal{C}$ can also be represented as the set of points $x \in [0, 1]$ of the form $x = \sum_{k=1}^{\infty} \epsilon_k / 4^k$, with ϵ_k either 0 or 3.

We now place a copy of $\mathcal{C}$ on the x-axis of the plane $\mathbb{R}^2 = \{(x, y)\}$, and a copy of $\frac{1}{2}\mathcal{C}$ on the line $y = 1$. That is, we put $E_0 = \{(x, y) : x \in \mathcal{C},\ y = 0\}$ and $E_1 = \{(x, y) : 2x \in \mathcal{C},\ y = 1\}$. The set F that will play the central role is defined as the union of all line segments that join a point of E_0 with a point of E_1. (See Figure 13.)

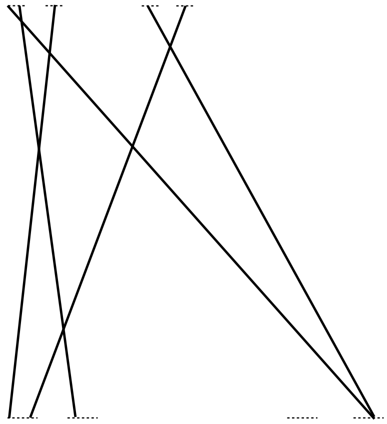

Figure 13. Several line segments joining E_0 with E_1

Theorem 4.12 *The set F is compact and of two-dimensional measure zero. It contains a translate of any unit line segment whose slope is a number s that lies outside the intervals* $(-1, 2)$.

Once the theorem is proved, our job is done. Indeed, a finite union of rotations of the set F contains unit segments of any slope, and that set is therefore a Besicovitch set.

The proof of the required properties of the set F amounts to showing the following paradoxical facts about the set $\mathcal{C} + \lambda\mathcal{C}$, for $\lambda > 0$. Here $\mathcal{C} + \lambda\mathcal{C} = \{x_1 + \lambda x_2 : \ x_1 \in \mathcal{C}, x_2 \in \mathcal{C}\}$:

- $\mathcal{C} + \lambda\mathcal{C}$ has one-dimensional measure zero, for a.e. λ.
- $\mathcal{C} + \frac{1}{2}\mathcal{C}$ is the interval $[0, 3/2]$.

Let us see how these two assertions imply the theorem. First, we note that the set F is closed (and hence compact), because both E_0 and E_1 are closed. Next observe that with $0 < y < 1$, the slice F^y of the set F is exactly $(1-y)\mathcal{C} + \frac{y}{2}\mathcal{C}$. This set is obtained from the set $\mathcal{C} + \lambda\mathcal{C}$, where $\lambda = y/(2(1-y))$, by scaling with the factor $1-y$. Hence F^y is of measure zero whenever $\mathcal{C} + \lambda\mathcal{C}$ is also of measure zero. Moreover, under the mapping $y \mapsto \lambda$, sets of measure zero in $(0, \infty)$ correspond to sets of measure zero in $(0, 1)$. (For this see, for example, Exercise 8 in Chapter 1, or Problem 1 in Chapter 6.) Therefore, the first assertion and Fubini's theorem prove that the (two-dimensional) measure of F is zero.

Finally the slope s of the segment joining the point $(x_0, 0)$, with the point $(x_1, 1)$ is $s = 1/(x_1 - x_0)$. Thus the quantity s can be realized if

$x_1 \in \mathcal{C}/2$ and $x_0 \in \mathcal{C}$, that is, if $1/s \in \mathcal{C}/2 - \mathcal{C}$. However, by an obvious symmetry $\mathcal{C} = 1 - \mathcal{C}$, and so the condition becomes $1/s \in \mathcal{C}/2 + \mathcal{C} - 1$, which by the second assertion is $1/s \in [-1, 1/2]$. This last is equivalent with $s \notin (-1, 2)$.

Our task therefore remains the proof of the two assertions above. The proof of the second is nearly trivial. In fact,

$$\frac{2}{3}\left(\mathcal{C} + \frac{1}{2}\mathcal{C}\right) = \frac{2}{3}\mathcal{C} + \frac{1}{3}\mathcal{C},$$

and this set consists of all x of the form $x = \sum_{k=1}^{\infty} \left(\frac{2\epsilon_k}{3} + \frac{\epsilon_k'}{3}\right) 4^{-k}$, where ϵ_k and ϵ_k' are independently 0 or 3. Since then $\frac{2\epsilon_k}{3} + \frac{\epsilon_k'}{3}$ can take any of the values 0, 1, 2, or 3, we have that $\frac{2}{3}\left(\mathcal{C} + \frac{1}{2}\mathcal{C}\right) = [0, 1]$, and hence $\mathcal{C} + \frac{1}{2}\mathcal{C} = [0, 3/2]$.

The proof that $m(\mathcal{C} + \lambda\mathcal{C}) = 0$ for a.e. λ

We come to the main point: that $\mathcal{C} + \lambda\mathcal{C}$ has measure zero for almost all λ. We show this by examining the self-replicating properties of the sets $\mathcal{C}$ and $\mathcal{C} + \lambda\mathcal{C}$.

We know that $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$, where $\mathcal{C}_1$ and $\mathcal{C}_2$ are two similar copies of $\mathcal{C}$, obtained with similarity ratio 1/4, and given by $\mathcal{C}_1 = \frac{1}{4}\mathcal{C}$ and $\mathcal{C}_2 = \frac{1}{4}\mathcal{C} + \frac{3}{4}$. Thus $\mathcal{C}_1 \subset [0, 1/4]$ and $\mathcal{C}_2 \subset [3/4, 1]$. Iterating ℓ times this decomposition of $\mathcal{C}$, that is, reaching the ℓ^{th} "generation," we can write

$$\mathcal{C} = \bigcup_{1 \leq j \leq 2^{\ell}} \mathcal{C}_j^{\ell}, \tag{7}$$

with $\mathcal{C}_1^{\ell} = (1/4)^{\ell}\mathcal{C}$ and each $\mathcal{C}_j^{\ell}$ a translate of $\mathcal{C}_1^{\ell}$.

We consider in the same way the set

$$\mathcal{K}(\lambda) = \mathcal{C} + \lambda\mathcal{C},$$

and we shall sometimes omit the λ and write $\mathcal{K}(\lambda) = \mathcal{K}$, when this causes no confusion. By its definition we have

$$\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2 \cup \mathcal{K}_3 \cup \mathcal{K}_4,$$

where $\mathcal{K}_1 = \mathcal{C}_1 + \lambda\mathcal{C}_1$, $\mathcal{K}_2 = \mathcal{C}_1 + \lambda\mathcal{C}_2$, $\mathcal{K}_3 = \mathcal{C}_2 + \lambda\mathcal{C}_1$, and $\mathcal{K}_4 = \mathcal{C}_2 + \lambda\mathcal{C}_2$. An iteration of this decomposition using (7) gives

$$\mathcal{K} = \bigcup_{1 \leq i \leq 4^{\ell}} \mathcal{K}_i^{\ell}, \tag{8}$$

where each $\mathcal{K}_i^\ell$ equals $\mathcal{C}_{j_1}^\ell + \lambda \mathcal{C}_{j_2}^\ell$ for a pair of indices j_1, j_2. In fact, this relation among the indices sets up a bijection between the i with $1 \leq i \leq 4^\ell$, and the pair j_1, j_2 with $1 \leq j_1 \leq 2^\ell$ and $1 \leq j_2 \leq 2^\ell$. Note that each $\mathcal{K}_i^\ell$ is a translate of $\mathcal{K}_1^\ell$, and each $\mathcal{K}_i^\ell$ is also obtained from $\mathcal{K}$ by a similarity of ratio $4^{-\ell}$. Now note that $\mathcal{C} = \mathcal{C}/4 \bigcup (\mathcal{C}/4 + 3/4)$ implies that

$$\begin{aligned}\mathcal{K}(\lambda) = \mathcal{C} + \lambda\mathcal{C} &= (\mathcal{C} + \frac{\lambda}{4}\mathcal{C}) \cup (\mathcal{C} + \frac{\lambda}{4}\mathcal{C} + \frac{3\lambda}{4}) \\ &= \mathcal{K}(\lambda/4) \cup (\mathcal{K}(\lambda/4) + \frac{3\lambda}{4}).\end{aligned}$$

Thus $\mathcal{K}(\lambda)$ has measure zero if and only if $\mathcal{K}(\lambda/4)$ has measure zero. Hence it suffices to prove that $\mathcal{K}(\lambda)$ has measure zero for a.e. $\lambda \in [1,4]$.

After these preliminaries let us observe that we immediately obtain that $m(\mathcal{K}(\lambda)) = 0$ for some special λ's, those for which the following **coincidence** takes place: for some ℓ and a pair i and i' with $i \neq i'$,

$$\mathcal{K}_i^\ell(\lambda) = \mathcal{K}_{i'}^\ell(\lambda).$$

Indeed, if we have this coincidence, then (8) gives

$$m(\mathcal{K}(\lambda)) \leq \sum_{i=1,\ i\neq i'}^{4^\ell} m(\mathcal{K}_i^\ell(\lambda)) = (4^\ell - 1)4^{-\ell} m(\mathcal{K}(\lambda)),$$

and this implies $m(\mathcal{K}(\lambda)) = 0$.

The key insight below is that, in a quantitative sense, the λ's for which this coincidence takes place are "dense" relative to the size of ℓ. More precisely, we have the following.

Proposition 4.13 *Suppose λ_0 and ℓ are given, with $1 \leq \lambda_0 \leq 4$ and ℓ a positive integer. Then, there exist a $\overline{\lambda}$ and a pair i, i' with $i \neq i'$ such that*

$$\mathcal{K}_i^\ell(\overline{\lambda}) = \mathcal{K}_{i'}^\ell(\overline{\lambda}) \quad \textit{and} \quad |\overline{\lambda} - \lambda_0| \leq c4^{-\ell}. \tag{9}$$

Here c is a constant independent of λ_0 and ℓ.

This is proved on the basis of the following observation.

Lemma 4.14 *For every λ_0 there is a pair $1 \leq i_1, i_2 \leq 4$, with $i_1 \neq i_2$ such that $\mathcal{K}_{i_1}(\lambda_0)$ and $\mathcal{K}_{i_2}(\lambda_0)$ intersect.*

Proof. Indeed, if the $\mathcal{K}_i$ are disjoint for $1 \leq i \leq 4$ then for sufficiently small δ the $\mathcal{K}_i^\delta$ are also disjoint. Here we have used the notation that F^δ denotes the set of points of distance less than δ from F. (See Lemma 3.1 in Chapter 1.) However, $\mathcal{K}^\delta = \bigcup_{i=1}^4 \mathcal{K}_i^\delta$, and by similarity $m(\mathcal{K}^{4\delta}) = 4m(\mathcal{K}_i^\delta)$. Thus by the disjointness of the $\mathcal{K}_i^\delta$ we have $m(\mathcal{K}^\delta) = m(\mathcal{K}^{4\delta})$, which is a contradiction, since $\mathcal{K}^{4\delta} - \mathcal{K}^\delta$ contains an open ball (of radius $3\delta/2$). The lemma is therefore proved.

Now apply the lemma for our given λ_0 and write $\mathcal{K}_{i_1} = \mathcal{C}_{\mu_1} + \lambda_0 \mathcal{C}_{\nu_1}$, $\mathcal{K}_{i_2} = \mathcal{C}_{\mu_2} + \lambda_0 \mathcal{C}_{\nu_2}$, where the μ's and ν's are either 1 or 2. However, since $i_1 \neq i_2$ we have $\mu_1 \neq \mu_2$ or $\nu_1 \neq \nu_2$ (or both). Assume for the moment that $\nu_1 \neq \nu_2$.

The fact that $\mathcal{K}_{i_1}(\lambda_0)$ and $\mathcal{K}_{i_2}(\lambda_0)$ intersect means that there are pairs of numbers (a, b) and (a', b'), with $a \in \mathcal{C}_{\mu_1}$, $b \in \mathcal{C}_{\nu_1}$, $a' \in \mathcal{C}_{\mu_2}$, and $b' \in \mathcal{C}_{\nu_2}$ such that

$$a + \lambda_0 b = a' + \lambda_0 b'. \tag{10}$$

Note that the fact that $\nu_1 \neq \nu_2$ means that $|b - b'| \geq 1/2$. Next, looking at the ℓ^{th} generation we find via (7) that there are indices $1 \leq j_1, j_2, j_1', j_2' \leq 2^\ell$, so that $a \in \mathcal{C}_{j_1}^\ell \subset \mathcal{C}_{\mu_1}$, $b \in \mathcal{C}_{j_2}^\ell \subset \mathcal{C}_{\nu_1}$, $a' \in \mathcal{C}_{j_1'}^\ell \subset \mathcal{C}_{\mu_2}$, $b' \in \mathcal{C}_{j_2'}^\ell \subset \mathcal{C}_{\nu_2}$. We also observe that the above sets are translates of each other, that is, $\mathcal{C}_{j_1}^\ell = \mathcal{C}_{j_1'}^\ell + \tau_1$ and $\mathcal{C}_{j_2}^\ell = \mathcal{C}_{j_2'}^\ell + \tau_2$, with $|\tau_k| \leq 1$. Hence if i and i' correspond to the pairs (j_1, j_2) and (j_1', j_2'), respectively, we have

$$\mathcal{K}_i^\ell(\lambda) = \mathcal{K}_{i'}^\ell(\lambda) + \tau(\lambda) \qquad \text{with } \tau(\lambda) = \tau_1 + \lambda\tau_2. \tag{11}$$

Now let (A, B) be the pair that corresponds to (a', b') under the above translations, namely

$$A = a' + \tau_1, \qquad B = b' + \tau_2. \tag{12}$$

We claim there is a $\overline{\lambda}$ such that

$$A + \overline{\lambda} B = a' + \overline{\lambda} b'. \tag{13}$$

In fact, by (12) we have put B in $\mathcal{C}_{j_2}^\ell \subset \mathcal{C}_{\nu_1}$, while b' is in $\mathcal{C}_{j_2'}^\ell \subset \mathcal{C}_{\nu_2}$. Thus $|B - b'| \geq 1/2$, since $\nu_1 \neq \nu_2$. We can therefore solve (13) by taking $\overline{\lambda} = (A - a')/(b' - B)$. Now we compare this with (10), and get $\lambda_0 = (a - a')/(b' - b)$. Moreover, $|A - a| \leq 4^{-\ell}$ and $|B - b| \leq 4^{-\ell}$, since A and a both lie in $\mathcal{C}_{j_1}^\ell$, and B and b lie in $\mathcal{C}_{j_2}^\ell$. This yields the inequality

$$|\overline{\lambda} - \lambda_0| \leq c4^{-\ell}. \tag{14}$$

Also, (12) and (13) clearly imply $\tau(\overline{\lambda}) = \tau_1 + \overline{\lambda}\tau_2 = 0$, and this together with (11) proves the coincidence.

Therefore our proposition is proved under the restriction we made earlier that $\nu_1 \neq \nu_2$. The situation where instead $\mu_1 \neq \mu_2$ is obtained from the case $\nu_1 \neq \nu_2$ if we replace λ_0 by λ_0^{-1}. Note that $\mathcal{K}_i^\ell(\lambda_0) = \mathcal{K}_{i'}^\ell(\lambda_0)$ if and only if $\mathcal{C}_{j_1}^\ell + \lambda_0 \mathcal{C}_{j_2}^\ell = \mathcal{C}_{j_1'}^\ell + \lambda_0 \mathcal{C}_{j_2'}^\ell$ and this is the same as $\mathcal{C}_{j_2}^\ell + \lambda_0^{-1}\mathcal{C}_{j_1}^\ell = \mathcal{C}_{j_2'}^\ell + \lambda_0^{-1}\mathcal{C}_{j_1'}^\ell$. This allows us to reduce to the case $\mu_1 \neq \mu_2$, since $\mathcal{C}_{j_1}^\ell \subset \mathcal{C}_{\mu_1}$ and $\mathcal{C}_{j_1'}^\ell \subset \mathcal{C}_{\mu_2}$. Here the fact that $1 \leq \lambda_0 \leq 4$ gives $\lambda_0^{-1} \leq 1$ and guarantees that the constant c in (9) can be taken to be independent of λ_0. The proposition is therefore established.

Note that as a consequence, the following holds near the points $\overline{\lambda}$ where the coincidence (9) takes place: If $|\lambda - \overline{\lambda}| \leq \epsilon 4^{-\ell}$, then

$$\mathcal{K}_i^\ell(\lambda) = \mathcal{K}_{i'}^\ell(\lambda) + \tau(\lambda) \quad \text{with } |\tau(\lambda)| \leq \epsilon 4^{-\ell}. \tag{15}$$

In fact, this is (11) together with the observation that

$$|\tau(\lambda)| = |\tau(\lambda) - \tau(\overline{\lambda})| \leq |\lambda - \overline{\lambda}|,$$

since $|\tau(\lambda)| = \tau_1 + \lambda\tau_2$ and $|\tau_2| \leq 1$.

The assertion (15) leads to the following more elaborate version of itself:

> *There is a set Λ of full measure such that whenever $\lambda \in \Lambda$ and $\epsilon > 0$ are given, there are ℓ and a pair i, i' so that* (15) *holds.*[3]

Indeed, for fixed $\epsilon > 0$, let Λ_ϵ denote the set of λ that satisfies (15) for some ℓ, i and i'. For any interval I of length not exceeding 1, we have

$$m(\Lambda_\epsilon \cap I) \geq \epsilon 4^{-\ell} \geq c^{-1}\epsilon m(I),$$

because of (9) and (15). Thus Λ_ϵ^c has no points of Lebesgue density, hence Λ_ϵ^c has measure zero, and thus Λ_ϵ is a set of full measure. (See Corollary 1.5 in Chapter 3.) Since $\Lambda = \bigcap_\epsilon \Lambda_\epsilon$, and Λ_ϵ decreases with ϵ, we see that Λ also has full measure and our assertion is proved.

Finally, our theorem will be established once we show that $m(\mathcal{K}(\lambda)) = 0$ whenever $\lambda \in \Lambda$. To prove this, we assume contrariwise that $m(\mathcal{K}(\lambda)) > 0$. Using again the point of density argument, there must be for any

[3]The terminology that Λ has "full measure" means that its complement has measure zero.

$0 < \delta < 1$, a non-empty open interval I with $m(\mathcal{K}(\lambda) \cap I) \geq \delta m(I)$. We then fix δ with $1/2 < \delta < 1$ and proceed. With this fixed δ, we select ϵ used below as $\epsilon = m(I)(1-\delta)$. Next, find ℓ, i, and i' for which (15) holds. The existence of such indices is guaranteed by the hypothesis that $\lambda \in \Lambda$.

We then consider the two similarities (of ratio $4^{-\ell}$) that map $\mathcal{K}(\lambda)$ to $\mathcal{K}_i^\ell(\lambda)$ and $\mathcal{K}_{i'}^\ell(\lambda)$, respectively. These take the interval I to corresponding intervals I_i and $I_{i'}$, respectively, with $m(I_i) = m(I_{i'}) = 4^{-\ell}m(I)$. Moreover,

$$m(\mathcal{K}_i^\ell \cap I_i) \geq \delta m(I_i) \quad \text{and} \quad m(\mathcal{K}_{i'}^\ell \cap I_{i'}) \geq \delta m(I_{i'}).$$

Also, as in (15), $I_{i'} = I_i + \tau(\lambda)$, with $|\tau(\lambda)| \leq \epsilon 4^{-\ell}$. This shows that

$$m(I_i \cap I_{i'}) \geq m(I_i) - \tau(\lambda) \geq 4^{-\ell}m(I) - \epsilon 4^{-\ell} \geq \delta m(I_i),$$

since $\epsilon 4^{-\ell} = (1-\delta)m(I_i)$. Thus $m(I_i - I_i \cap I_{i'}) \leq (1-\delta)m(I_i)$, and

$$\begin{aligned} m(\mathcal{K}_i^\ell \cap I_i \cap I_{i'}) &\geq m(\mathcal{K}_i^\ell \cap I_i) - m(I_i - I_i \cap I_{i'}) \\ &\geq (2\delta - 1)m(I_i) \\ &> \frac{1}{2}m(I_i) \geq \frac{1}{2}m(I_i \cap I_{i'}). \end{aligned}$$

So $m(\mathcal{K}_i^\ell \cap I_i \cap I_{i'}) > \frac{1}{2}m(I_i \cap I_{i'})$ and the same holds for i' in place of i. Hence $m(\mathcal{K}_i^\ell \cap \mathcal{K}_{i'}^\ell) > 0$, and this contradicts the decomposition (8) and the fact that $m(\mathcal{K}_i^\ell) = 4^{-\ell}m(\mathcal{K})$ for every i. Therefore we obtain that $m(\mathcal{K}(\lambda)) = 0$ for every $\lambda \in \Lambda$, and the proof of Theorem 4.12 is now complete.

5 Exercises

1. Show that the measure m_α is not σ-finite on $\mathbb{R}^d$ if $\alpha < d$.

2. Suppose E_1 and E_2 are two compact subsets of $\mathbb{R}^d$ such that $E_1 \cap E_2$ contains at most one point. Show directly from the definition of the exterior measure that if $0 < \alpha \leq d$, and $E = E_1 \cup E_2$, then

$$m_\alpha^*(E) = m_\alpha^*(E_1) + m_\alpha^*(E_2).$$

[Hint: Suppose $E_1 \cap E_2 = \{x\}$, let B_ϵ denote the open ball centered at x and of diameter ϵ, and let $E^\epsilon = E \cap B_\epsilon^c$. Show that

$$m_\alpha^*(E^\epsilon) \geq \mathcal{H}_\alpha^\epsilon(E) \geq m_\alpha^*(E) - \mu(\epsilon) - \epsilon^\alpha,$$

where $\mu(\epsilon) \to 0$. Hence $m_\alpha^*(E^\epsilon) \to m_\alpha^*(E)$.]

3. Prove that if $f : [0,1] \to \mathbb{R}$ satisfies a Lipschitz condition of exponent $\gamma > 1$, then f is a constant.

4. Suppose $f : [0,1] \to [0,1] \times [0,1]$ is surjective and satisfies a Lipschitz condition

$$|f(x) - f(y)| \leq C|x-y|^\gamma.$$

Prove that $\gamma \leq 1/2$ directly, without using Theorem 2.2.

[Hint: Divide $[0,1]$ into N intervals of equal length. The image of each sub-interval is contained in a ball of volume $O(N^{-2\gamma})$, and the union of all these balls must cover the square.]

5. Let $f(x) = x^k$ be defined on $\mathbb{R}$, where k is a positive integer and let E be a Borel subset of $\mathbb{R}$.

(a) Show that if $m_\alpha(E) = 0$ for some α, then $m_\alpha(f(E)) = 0$.

(b) Prove that $\dim(E) = \dim f(E)$.

6. Let $\{E_k\}$ be a sequence of Borel sets in $\mathbb{R}^d$. Show that if $\dim E_k \leq \alpha$ for some α and all k, then

$$\dim \bigcup_k E_k \leq \alpha.$$

7. Prove that the $(\log 2/\log 3)$-Hausdorff measure of the Cantor set is precisely equal to 1.

[Hint: Suppose we have a covering of $\mathcal{C}$ by finitely many closed intervals $\{I_j\}$. Then there exists another covering of $\mathcal{C}$ by intervals $\{I'_\ell\}$ each of length 3^{-k} for some k, such that $\sum_j |I_j|^\alpha \geq \sum_\ell |I'_\ell|^\alpha \geq 1$, where $\alpha = \log 2/\log 3$.]

8. Show that the Cantor set of constant dissection, $\mathcal{C}_\xi$, in Exercise 3 of Chapter 1 has strict Hausdorff dimension $\log 2/\log(2/(1-\xi))$.

9. Consider the set $\mathcal{C}_{\xi_1} \times \mathcal{C}_{\xi_2}$ in $\mathbb{R}^2$, with $\mathcal{C}_\xi$ as in the previous exercise. Show that $\mathcal{C}_{\xi_1} \times \mathcal{C}_{\xi_2}$ has strict Hausdorff dimension $\dim(\mathcal{C}_{\xi_1}) + \dim(\mathcal{C}_{\xi_2})$.

10. Construct a Cantor-like set (as in Exercise 4, Chapter 1) that has Lebesgue measure zero, yet Hausdorff dimension 1.

[Hint: Choose $\ell_1, \ell_2, \ldots, \ell_k, \ldots$ so that $1 - \sum_{j=1}^k 2^{j-1}\ell_j$ tends to zero sufficiently slowly as $k \to \infty$.]

11. Let $\mathcal{D} = \mathcal{D}_\mu$ be the Cantor dust in $\mathbb{R}^2$ given as the product $\mathcal{C}_\xi \times \mathcal{C}_\xi$, with $\mu = (1-\xi)/2$.

(a) Show that for any real number λ, the set $\mathcal{C}_\xi + \lambda\mathcal{C}_\xi$ is similar to the projection of $\mathcal{D}$ on the line in $\mathbb{R}^2$ with slope $\lambda = \tan\theta$.

(b) Note that among the Cantor sets $\mathcal{C}_\xi$, the value $\xi = 1/2$ is critical in the construction of the Besicovitch set in Section 4.4. In fact, prove that with $\xi > 1/2$, then $\mathcal{C}_\xi + \lambda\mathcal{C}_\xi$ has Lebesgue measure zero for *every* λ. See also Problem 10 below.

[Hint: $m_\alpha(\mathcal{C}_\xi + \lambda\mathcal{C}_\xi) < \infty$ for $\alpha = \dim \mathcal{D}_\mu$.]

12. Define a primitive one-dimensional "measure" $\tilde{m}_1$ as

$$\tilde{m}_1 = \inf \sum_{k=1}^{\infty} \text{diam } F_k, \qquad E \subset \bigcup_{k=1}^{\infty} F_k.$$

This is akin to the one-dimensional exterior measure m_α^*, $\alpha = 1$, except that no restriction is placed on the size of the diameters F_k.

Suppose I_1 and I_2 are two *disjoint* unit segments in $\mathbb{R}^d$, $d \geq 2$, with $I_1 = I_2 + h$, and $|h| < \epsilon$. Then observe that $\tilde{m}_1(I_1) = \tilde{m}_1(I_2) = 1$, while $\tilde{m}_1(I_1 \cup I_2) \leq 1 + \epsilon$. Thus

$$\tilde{m}_1(I_1 \cup I_2) < \tilde{m}_1(I_1) + \tilde{m}_1(I_2) \qquad \text{when } \epsilon < 1;$$

hence $\tilde{m}_1$ fails to be additive.

13. Consider the von Koch curve $\mathcal{K}^\ell$, $1/4 < \ell < 1/2$, as defined in Section 2.1. Prove for it the analogue of Theorem 2.7: the function $t \mapsto \mathcal{K}^\ell(t)$ satisfies a Lipschitz condition of exponent $\gamma = \log(1/\ell)/\log 4$. Moreover, show that the set $\mathcal{K}^\ell$ has strict Hausdorff dimension $\alpha = 1/\gamma$.

[Hint: Show that if $\mathcal{O}$ is the shaded open triangle indicated in Figure 14, then $\mathcal{O} \supset S_0(\mathcal{O}) \cup S_1(\mathcal{O}) \cup S_2(\mathcal{O}) \cup S_3(\mathcal{O})$, where $S_0(x) = \ell x$, $S_1(x) = \rho_\theta(\ell x) + a$, $S_2(x) = \rho_\theta^{-1}(\ell x) + c$, and $S_3(x) = \ell x + b$, with ρ_θ the rotation of angle θ. Note that the sets $S_j(\mathcal{O})$ are disjoint.]

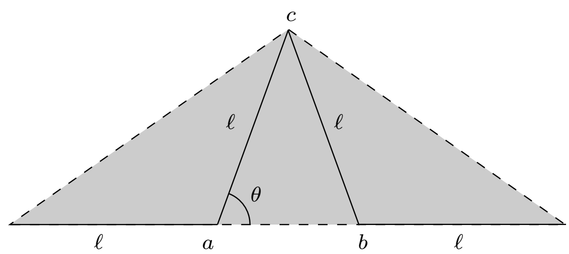

Figure 14. The open set $\mathcal{O}$ in Exercise 13

14. Show that if $\ell < 1/2$, the von Koch curve $t \mapsto \mathcal{K}^\ell(t)$ in Exercise 13 is a simple curve.

[Hint: Observe that if $t = \sum_{j=1}^{\infty} a_j/4^j$, with $a_j = 0, 1, 2,$ or 3, then

$$\{\mathcal{K}(t)\} = \bigcap_{j=1}^{\infty} S_{a_j}\left(\cdots S_{a_2}\left(S_{a_1}(\overline{\mathcal{O}})\right)\right).]$$

15. Note that if we take $\ell = 1/2$ in the definition of the von Koch curve in Exercise 13 we get a "space-filling" curve, one that fills the right triangle whose vertices are $(0,0)$, $(1,0)$, and $(1/2,1/2)$. The first three steps of the construction are as in Figure 15, with the intervals traced out in the indicated order.

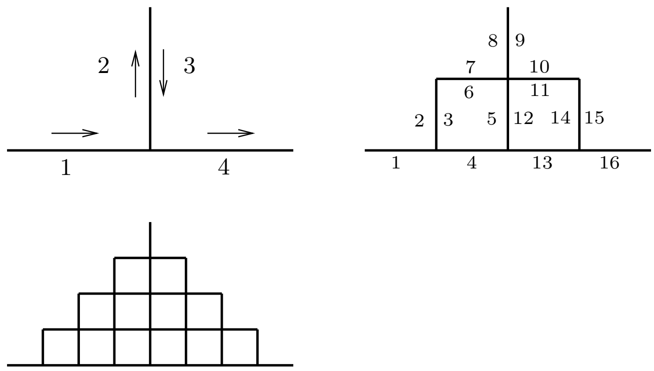

Figure 15. The first three steps of the von Koch curve when $\ell = 1/2$

16. Prove that the von Koch curve $t \mapsto \mathcal{K}^{\ell}(t)$, $1/4 < \ell \leq 1/2$ is continuous but nowhere differentiable.

[Hint: If $\mathcal{K}'(t)$ exists for some t, then

$$\lim_{n\to\infty} \frac{\mathcal{K}(u_n) - \mathcal{K}(v_n)}{u_n - v_n}$$

must exist, where $u_n \leq t \leq v_n$, and $u_n - v_n \to 0$. Choose $u_n = k/4^n$ and $v_n = (k+1)/4^n$.]

17. For a compact set E in $\mathbb{R}^d$, define $\#(\epsilon)$ to be the least number of balls of radius ϵ that cover E. Note that we always have $\#(\epsilon) = O(\epsilon^{-d})$ as $\epsilon \to 0$, and $\#(\epsilon) = O(1)$ if E is finite.

One defines the **covering dimension** of E, denoted by $\dim_C(E)$, as $\inf \beta$ such that $\#(\epsilon) = O(\epsilon^{-\beta})$, as $\epsilon \to 0$. Show that $\dim_C(E) = \dim_M(E)$, where $\dim_M$ is the Minkowski dimension discussed in Section 2.1, by proving the following inequalities for all $\delta > 0$:

(i) $m(E^\delta) \leq c\#(\delta)\delta^d$.

(ii) $\#(\delta) \leq c' m(E^\delta)\delta^{-d}$.

[Hint: To prove (ii), use Lemma 1.2 in Chapter 3 to find a collection of disjoint balls $B_1, B_2, \ldots, B_N$ of radius $\delta/3$, each centered at E, such that their "triples" $\tilde{B}_1, \tilde{B}_2, \ldots, \tilde{B}_N$ (of radius δ) cover E. Then $\#(\delta) \leq N$, while $Nm(B_j) = cN\delta^d \leq m(E^\delta)$, since the balls B_j are disjoint and are contained in E^δ.]

18. Let E be a compact set in $\mathbb{R}^d$.

(a) Prove that $\dim(E) \leq \dim_M(E)$, where dim and $\dim_M$ are the Hausdorff and Minkowski dimensions, respectively.

(b) However, prove that if $E = \{0, 1/\log 2, 1/\log 3, \ldots, 1/\log n, \ldots\}$, then $\dim_M E = 1$, yet $\dim E = 0$.

19. Show that there is a constant c_d, dependent only on the dimension d, such that whenever E is a compact set,

$$m(E^{2\delta}) \leq c_d m(E^\delta).$$

[Hint: Consider the maximal function f^*, with $f = \chi_{E^\delta}$, and take $c_d = 6^d$.]

20. Show that if F is the self-similar set considered in Theorem 2.12, then it has the same Minkowski dimension as Hausdorff dimension.

[Hint: Each F_k is the union of m^k balls of radius cr^k. In the converse direction one sees by Lemma 2.13 that if $\epsilon = r^k$, then each ball of radius ϵ can contain at most c' vertices of the k^{th} generation. So it takes at least m^k/c' such balls to cover F.]

21. From the unit interval, remove the second and fourth quarters (open intervals). Repeat this process in the remaining two closed intervals, and so on. Let F be the limiting set, so that

$$F = \{x : \; x = \sum_{k=1}^{\infty} a_k/4^k \quad a_k = 0 \text{ or } 2\}.$$

Prove that $0 < m_{1/2}(F) < \infty$.

22. Suppose F is the self-similar set arising in Theorem 2.9.

(a) Show that if $m \leq 1/r^d$, then $m_d(F_i \cap F_j) = 0$ if $i \neq j$.

(b) However, if $m \geq 1/r^d$, prove that $F_i \cap F_j$ is not empty for some $i \neq j$.

(c) Prove that under the hypothesis of Theorem 2.12

$$m_\alpha(F_i \cap F_j) = 0, \qquad \text{with } \alpha = \log m/\log(1/r), \text{ whenever } i \neq j.$$

23. Suppose $S_1, \dots, S_m$ are similarities with ratio r, $0 < r < 1$. For each set E, let

$$\tilde{S}(E) = S_1(E) \cup \cdots \cup S_m(E),$$

and suppose F denotes the unique non-empty compact set with $\tilde{S}(F) = F$.

(a) If $\overline{x} \in F$, show that the set of points $\{\tilde{S}^n(\overline{x})\}_{n=1}^{\infty}$ is dense in F.

(b) Show that F is **homogeneous** in the following sense: if $x_0 \in F$ and B is any open ball centered at x_0, then $F \cap B$ contains a set similar to F.

24. Suppose E is a Borel subset of $\mathbb{R}^d$ with $\dim E < 1$. Prove that E is totally disconnected, that is, any two distinct points in E belong to different connected components.

[Hint: Fix $x, y \in E$, and show that $f(t) = |t - x|$ is Lipschitz of order 1, and hence $\dim f(E) < 1$. Conclude that $f(E)$ has a dense complement in $\mathbb{R}$. Pick r in the complement of $f(E)$ so that $0 < r < f(y)$, and use the fact that $E = \{t \in E : |t - x| < r\} \cup \{t \in E : |t - x| > r\}$.]

25. Let $F(t)$ be an arbitrary non-negative measurable function on $\mathbb{R}$, and $\gamma \in S^{d-1}$. Then there exists a measurable set E in $\mathbb{R}^d$, such that $F(t) = m_{d-1}(E \cap \mathcal{P}_{t,\gamma})$.

26. Theorem 4.1 can be refined for $d \geq 4$ as follows.

Define $C^{k,\alpha}$ to be the class of functions $F(t)$ on $\mathbb{R}$ that are C^k and for which $F^{(k)}(t)$ satisfies a Lipschitz condition of exponent α.

If E has finite measure, then for a.e. $\gamma \in S^{d-1}$ the function $m(E \cap \mathcal{P}_{t,\gamma})$ is in $C^{k,\alpha}$ for $k = (d-3)/2$, $\alpha < 1/2$, if d is odd, $d \geq 3$; and for, $k = (d-4)/2$, $\alpha < 1$, if d is even, $d \geq 4$.

27. Show that the modification of the inequality (2) of Theorem 4.5 fails if we drop $\|f\|_{L^2(\mathbb{R}^d)}$ from the right-hand side.

[Hint: Consider $\mathcal{R}^*(f_\epsilon)$, with f_ϵ defined by $f_\epsilon(x) = (|x| + \epsilon)^{-d+\delta}$, for $|x| \leq 1$, with δ fixed, $0 < \delta < 1$, and $\epsilon \to 0$.]

28. Construct a compact set $E \subset \mathbb{R}^d$, $d \geq 3$, such that $m_d(E) = 0$, yet E contains translates of any segment of unit length in $\mathbb{R}^d$. (While particular examples of such sets can be easily obtained from the case $d = 2$, the determination of the least Hausdorff dimension among all such sets is an open problem.)

6 Problems

1. Carry out the construction below of two sets U and V so that

$$\dim U = \dim V = 0 \quad \text{but} \quad \dim(U \times V) \geq 1.$$

Let $I_1, \dots, I_n, \dots$ be given as follows:

- Each I_j is a finite sequence of consecutive positive integers; that is, for all j

$$I_j = \{n \in \mathbb{N} : A_j \le n \le B_j\} \qquad \text{for some given } A_j \text{ and } B_j.$$

- For each j, I_{j+1} is to the right of I_j; that is, $A_{j+1} > B_j$.

Let $U \subset [0,1]$ consist of all x which when written dyadically $x = .a_1 a_2 \cdots a_n \cdots$ have the property that $a_n = 0$ whenever $n \in \bigcup_j I_j$. Assume also that A_j and B_j tend to infinity (as $j \to \infty$) rapidly enough, say $B_j/A_j \to \infty$ and $A_{j+1}/B_j \to \infty$.

Also, let J_j be the complementary blocks of integers, that is,

$$J_j = \{n \in \mathbb{N} : B_j < n < A_{j+1}\}.$$

Let $V \subset [0,1]$ consist of those $x = .a_1 a_2 \cdots a_n \cdots$ with $a_n = 0$ if $n \in \bigcup_j J_j$.

Prove that U and V have the desired property.

2.* The iso-diametric inequality states the following: If E is a bounded subset of $\mathbb{R}^d$ and diam $E = \sup\{|x-y| : x, y \in E\}$, then

$$m(E) \le v_d \left(\frac{\text{diam } E}{2} \right)^d,$$

where v_d denotes the volume of the unit ball in $\mathbb{R}^d$. In other words, among sets of a given diameter, the ball has maximum volume. Clearly, it suffices to prove the inequality for $\overline{E}$ instead of E, so we can assume that E is compact.

(a) Prove the inequality in the special case when E is symmetric, that is, $-x \in E$ whenever $x \in E$.

In general, one reduces to the symmetric case by using a technique called Steiner symmetrization. If e is a unit vector in $\mathbb{R}^d$, and $\mathcal{P}$ is a plane perpendicular to e, the Steiner symmetrization of E with respect to E is defined by

$$S(E,e) = \{x + te : x \in \mathcal{P},\ |t| \le \frac{1}{2} L(E;e;x)\},$$

where $L(E;e;x) = m(\{t \in \mathbb{R} :\ x + t \cdot e \in E\})$, and m denotes the Lebesgue measure. Note that $x + te \in S(E,e)$ if and only if $x - te \in S(E,e)$.

(b) Prove that $S(E,e)$ is a bounded measurable subset of $\mathbb{R}^d$ that satisfies $m(S(E,e)) = m(E)$.

[Hint: Use Fubini's theorem.]

(c) Show that diam $S(E,e) \le$ diam E.

(d) If ρ is a rotation that leaves E and $\mathcal{P}$ invariant, show that $\rho S(E,e) = S(E,e)$.

(e) Finally, consider the standard basis $\{e_1, \ldots, e_d\}$ of $\mathbb{R}^d$. Let $E_0 = E$, $E_1 = S(E_0, e_1)$, $E_2 = S(E_1, e_2)$, and so on. Use the fact that E_d is symmetric to prove the iso-diametric inequality.

(f) Use the iso-diametric inequality to show that $m(E) = \frac{v_d}{2^d} m_d(E)$ for any Borel set E in $\mathbb{R}^d$.

3. Suppose S is a similarity.

(a) Show that S maps a line segment to a line segment.

(b) Show that if L_1 and L_2 are two segments that make an angle α, then $S(L_1)$ and $S(L_2)$ make an angle α or $-\alpha$.

(c) Show that every similarity is a composition of a translation, a rotation (possibly improper), and a dilation.

4.* The following gives a generalization of the construction of the Cantor-Lebesgue function.

Let F be the compact set in Theorem 2.9 defined in terms of m similarities $S_1, S_2, \ldots, S_m$ with ratio $0 < r < 1$. There exists a unique Borel measure μ supported on F such that $\mu(F) = 1$ and

$$\mu(E) = \frac{1}{m} \sum_{j=1}^{m} \mu(S_j^{-1}(E)) \qquad \text{for any Borel set } E.$$

In the case when F is the Cantor set, the Cantor-Lebesgue function is $\mu([0, x])$.

5. Prove a theorem of Hausdorff: Any compact subset K of $\mathbb{R}^d$ is a continuous image of the Cantor set $\mathcal{C}$.

[Hint: Cover K by 2^{n_1} (some n_1) open balls of radius 1, say $B_1, \ldots, B_\ell$ (with possible repetitions). Let $K_{j_1} = K \cap \overline{B_{j_1}}$ and cover each K_{j_1} with 2^{n_2} balls of radius 1/2 to obtain compact sets K_{j_1,j_2}, and so on. Express $t \in \mathcal{C}$ as a ternary expansion, and assign to t a unique point in K defined by the intersection $K_{j_1} \cap K_{j_1,j_2} \cap \cdots$ for appropriate $j_1, j_2, \ldots$. To prove continuity, observe that if two points in the Cantor set are close, then their ternary expansions agree to high order.]

6. A compact subset K of $\mathbb{R}^d$ is **uniformly locally connected** if given $\epsilon > 0$ there exists $\delta > 0$ so that whenever $x, y \in K$ and $|x - y| < \delta$, there is a continuous curve γ in K joining x to y, such that $\gamma \subset B_\epsilon(x)$ and $\gamma \subset B_\epsilon(y)$.

Using the previous problem, one can show that a compact subset K of $\mathbb{R}^d$ is the continuous image of the unit interval $[0, 1]$ if and only if K is uniformly locally connected.

7. Formulate and prove a generalization of Theorem 3.5 to the effect that once appropriate sets of measure zero are removed, there is a measure-preserving isomorphism of the unit interval in $\mathbb{R}$ and the unit cube in $\mathbb{R}^d$.

8.* There exists a *simple* continuous curve in the plane of positive two-dimensional measure.

9. Let E be a compact set in $\mathbb{R}^{d-1}$. Show that $\dim(E \times I) = \dim(E) + 1$, where I is the unit interval in $\mathbb{R}$.

10.* Let $\mathcal{C}_\xi$ be the Cantor set considered in Exercises 8 and 11. If $\xi < 1/2$, then $\mathcal{C}_\xi + \lambda \mathcal{C}_\xi$ has positive Lebesgue measure for almost every λ.