

HOW TO DEFINE YOUR DIMENSION

A discourse on Hausdorff dimension and self-similarity

SATVIK SINGH

ABSTRACT. One often distinguishes between a line and a plane by saying that the former is one-dimensional while the latter is two. But, what does it mean for an object to have d -dimensions? Can we define a consistent notion of dimension rigorously for arbitrary objects, say a snowflake, perhaps? And must the dimension always be integer-valued? After highlighting some crucial problems that one encounters while defining a sensible notion of dimension for a certain class of objects, we attempt to answer the above questions by exploring the concept of Hausdorff dimension – a remarkable method of assigning dimension to subsets of arbitrary metric spaces. In order to properly formulate the definition and properties of the Hausdorff dimension, we review the critical measure-theoretic terminology beforehand. Finally, we discuss the notion of self-similarity and show how it often defies our quotidian intuition that dimension must always be integer-valued.

1. INTRODUCTION

Intuitively, the dimension of a set can be thought of as the minimum number of parameters needed to specify a unique element within the set. Following this reasoning, one can roughly think of points, curves, surfaces and solids to have, respectively, zero, one, two and three dimensions. However, shortcomings in this simple-minded logic surface when one considers, for instance, the fact that the unit line has the same cardinality as the unit plane [Can77, NS14], i.e. there exists a one-one mapping from $[0, 1]$ onto $[0, 1]^2$. Thus, any point in the plane can be specified by only one parameter by uniquely mapping it to its pre-image in $[0, 1]$. Does this mean that the above intuition is flawed? What if we consider the additional topological (or metric) structure on these spaces? In this scenario, the seminal works of Peano [Pea90], Hilbert [Hil91] and others [Bad12] show that the simplest *space-filling curves* (continuous mappings from the unit line *onto* the unit plane, see Figure 1) can be used to “fill up” two dimensions from a supposed one-dimensional set, thus apparently undermining the above notion of dimension again.

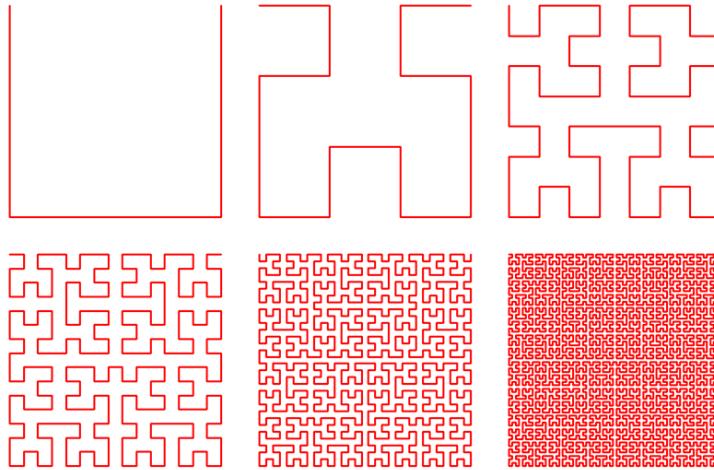


FIGURE 1. Ranges of the first six iterations of “Psuedo-Hilbert” curves, whose pointwise limit yields the continuous space-filling Hilbert curve from $[0, 1]$ onto $[0, 1] \times [0, 1]$

However, such space-filling curves are bound to be non-injective, meaning that although each point in the unit square is hit atleast once by the curve, its pre-image on the unit line is not unique. More generally, it can be shown that $[0, 1]$ and $[0, 1]^2$ are not topologically *homeomorphic*, i.e. there is no continuous bijection between the two which can also be continuously inverted. The concept of *topological dimension* [Edg08, Chapter 3] uses this fact to dimensionally distinguish between the two spaces. Roughly speaking, the topological dimension of a topological space is the minimum d such that, given any open cover of the space, there exists a refinement with the property that no point in the space lies in the intersection of more than $d + 1$ covering sets in the refinement. In this language, the unit line has topological dimension $d = 1$, since any open cover of it is bound to count some points twice. Similarly, it is not too hard to see that the unit plane has topological dimension $d = 2$. Clearly, *homeomorphic* spaces have identical topological dimensions. However, is invariance under homeomorphisms really a desirable property that every notion of dimension must posses?

To answer the above question, let us consider the graph $\Gamma(f) := \{(x, f(x)) \in \mathbb{R}^3 : x \in [0, 1]\}$ of some space-filling curve $f : [0, 1] \rightarrow [0, 1]^2$. Intuitively, one would expect that the dimension of $\Gamma(f)$ must atleast be greater than one, because the range of f completely fills the unit plane (it is not hard to imagine that $\Gamma(f)$ is a ragged object in the unit cube, which looks like the unit square when seen head on from a direction parallel to the x -axis)! However, since the graph of any continuous function is homeomorphic to its domain, the topological dimension of $\Gamma(f)$ is again one. This is just one example of a large family of *self-similar* objects (or *fractals*), for which the topological dimension is smaller than what one might expect. The problem with the topological notion of dimension is that it is completely oblivious to any *local structure* (or roughness) that the concerned object might posses, since homeomorphisms can only discriminate between those spaces which *cannot* be transformed into one another by means of a *global* continuous deformation. Let us consider the self-similar *Koch snowflake* from Figure 2 to further elaborate on this point. While it is evident that the curve doesn't fill up the entire plane, it is definitely more complex than a line because of all the intricate spikes and crevices. Being impervious to this additional structure, a topologist would again call this a one-dimensional curve, since it can be shown to be homeomorphic to an interval in \mathbb{R} . It is thus clear that a more sophisticated notion of dimension needs to be developed in order to deal with such exotic objects.

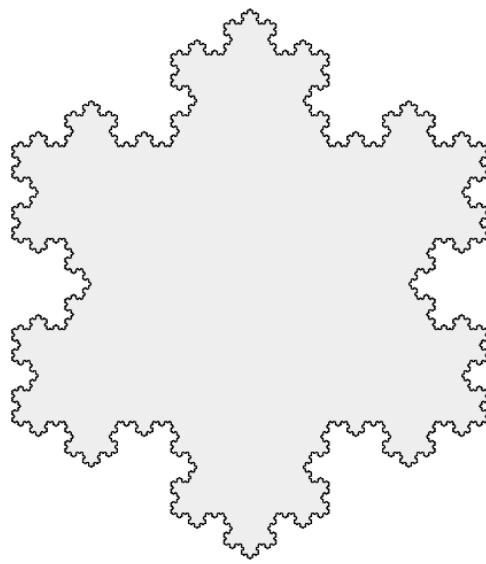


FIGURE 2. The 7th level in the construction of the Koch snowflake.

Remark 1.1. One can also exploit the linear structure of the real line \mathbb{R} and the real plane \mathbb{R}^2 to dimensionally distinguish between them. In this regime, the notion of vector space dimension becomes important, which, for any finite-dimensional vector space, is just the (unique) cardinality of its basis set. The structure preserving maps in the category of vector spaces are linear bijections, and consequently, the vector space dimension stays invariant under them. One can extend this idea of linear dimension to spaces which only locally look like \mathbb{R}^d to obtain the class of d -dimensional manifolds. However, this notion of dimension is too restrictive in its domain, since many objects occurring in nature are not manifolds! Or, as Mandelbrot would say [MFC83, p. 1], “clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line.” In this regard, it would be instructive for the reader to show that the closed unit interval $[0, 1]$, for instance, is neither a vector space nor a manifold.

During the early 20th century, a remarkable method was developed by Hausdorff (along with Lebesgue) for assigning dimension to subsets of an arbitrary metric space, which not only generalizes the topological notion but also takes the sets’ local structure into account. The idea is to explicitly employ the underlying metric (or distance function, which is sensitive to the object’s local structure at small scales) in order to associate a “proper” kind of measurement to the concerned object, which will be used to measure its size. For instance, if we consider an interval $(a, b) \subset \mathbb{R}$, we can say that its cardinality ($\dim = 0$) is infinite, its length ($\dim = 1$) is $b - a$ and its area ($\dim = 2$) is zero. The step d at which the size of an object changes from infinity to zero serves as an excellent candidate for its dimension. For the interval (a, b) , this step is $d = 1$. Moreover, if the step size for deciding this proper measurement is allowed to be arbitrarily small, then nothing stops the dimension of any given object from taking an arbitrary real value between zero and infinity. The only challenge is to come up with a meaningful notion of the δ -dimensional size of a given object for every $\delta > 0$.

In this article, we will formalize the above idea of *Hausdorff dimension* for subsets of arbitrary metric spaces. In order to do so, it is essential to first gather the appropriate measure-theoretic machinery, which will form the basis of our discussion in Section 2. Section 3 forms the core of this article, where we develop the promised definition of the Hausdorff dimension and elucidate several of its properties. Finally, we describe the notion of self-similarity for subsets of Euclidean spaces \mathbb{R}^k in Section 4 and compute the Hausdorff dimension of several famous self-similar objects like the *Cantor set*, the *Koch curve* and the *Sierpiński triangle*. It is probably worthwhile to emphasize here that this article is only meant to provide a light and accessible introduction to the basics of dimension theory and self-similarity. Hence, we do not claim to be exhaustive in our account of the aforementioned topics, neither do we claim to report on any modern advancements in the field. For a more sophisticated discussion on these topics, we refer the readers to the wonderful articles [Gra85, Man05, Sch05, BBE11, FM16] and books [MFC83, FF90, PJS⁺92, Edg08].

2. MEASURE THEORETIC PRELIMINARIES

In this section, we’ll briefly review some key notions from measure theory, which will be instrumental in defining the notion of Hausdorff dimension in the next section. For a much more thorough introduction to measure theory, the readers are referred to the classic texts [RRC87, Bil86, Hal76]. Before proceeding with our discussion, let us quickly review some basic notation. Given a non-empty set X , we denote its power set and the empty subset by $\mathcal{P}(X)$ and \emptyset , respectively. The complement of a subset $A \subseteq X$ is denoted by $X \setminus A$. The non-negative reals along with infinity are denoted by the interval $[0, \infty]$.

In a nutshell, measure theory deals with the task of sensibly assigning volumes to the subsets of a given set X . For $X = \mathbb{R}$, for instance, one would like to get hold of a function (called a *measure*) $\mu : \mathcal{P}(\mathbb{R}) \rightarrow [0, \infty]$ such that it is an extension of the usual notion of length for intervals, i.e. $\sigma((a - b]) = b - a \quad \forall a, b \in \mathbb{R}$. It is also rational to demand that this function be *countably additive* and *translationally invariant*, meaning that shifting a subset by a real number should not

change its volume and that the volume of countable disjoint unions of subsets should just be the sum of individual volumes of the subsets. However, by using the *axiom of choice*, one can show that such a function cannot exist, see [Hal76, Section 16]! It turns out that asking for a measure to be defined on all subsets of the real line is too strong a demand. Hence, one needs to restrict the domain of definition, which is where the notion of σ -algebras come into picture.

Definition 2.1. *A collection \mathcal{F} of subsets of a set X is called a σ -algebra if:*

- $\emptyset, X \in \mathcal{F}$.
- $A \in \mathcal{F} \implies X \setminus A \in \mathcal{F}$.
- $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{F} \implies \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F}$.

Using the last two properties, one can easily deduce that σ -algebras are also closed under countable intersections. The set X along with a σ -algebra \mathcal{F} is said to define a *measurable space* (X, \mathcal{F}) . The subsets in \mathcal{F} are then called *measurable*. Given a collection \mathcal{C} of subsets of a set X , it is natural to ask for a minimal σ -algebra on X which contains \mathcal{C} . For instance, suppose we have a topology τ (a collection of open sets) defined on X and we want to construct a σ -algebra on X such that it contains τ . The most economical way to do this is to obviously consider the minimal such σ -algebra. The following result assures us that such a minimal σ -algebra exists for every \mathcal{C} .

Lemma 2.2. *Let \mathcal{C} be a non-empty collection of subsets of a set X . Define*

$$\sigma(\mathcal{C}) := \bigcap \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-algebra which contains } \mathcal{C}\}.$$

Then, $\sigma(\mathcal{C})$ is the minimal σ -algebra which contains \mathcal{C} , i.e. if \mathcal{G} is a σ -algebra which contains \mathcal{C} , then $\sigma(\mathcal{C}) \subseteq \mathcal{G}$. We say that the collection \mathcal{C} generates $\sigma(\mathcal{C})$.

Given an arbitrary topological space (X, τ) , we can turn it into a measurable space by defining $\mathcal{F} = \sigma(\tau)$ to be the σ -algebra generated by the topology τ . The measurable sets are then called *Borel sets*. It is easy to see that all open, closed and compact sets (among others) in X are Borel.

With the concept of σ -algebra in place, we now arrive at the central definition of a measure.

Definition 2.3. *Consider a measurable space (X, \mathcal{F}) . Then, a set function $\mu : \mathcal{F} \mapsto [0, \infty]$ is said to be a measure if $\mu(\emptyset) = 0$ and if μ is countably additive, i.e. for all countable pairwise disjoint collections of measurable subsets $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{F}$, $\mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.*

Notice that measurable sets are allowed to have infinite measure. A measurable space (X, \mathcal{F}) equipped with a measure μ is called a *measure space* (X, \mathcal{F}, μ) . We list some important properties of measures below, which are straightforward consequences of the above definition.

Lemma 2.4. *Let (X, \mathcal{F}, μ) be a measure space. Then,*

- $\mu(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mu(A_i)$, where $\{A_i\}_{i=1}^n \subseteq \mathcal{F}$ is a pairwise disjoint collection of subsets.
- $A, B \in \mathcal{F}$ with $A \subseteq B \implies \mu(A) \leq \mu(B)$.
- $A, B \in \mathcal{F}$ with $A \subseteq B$ and $\mu(A) < \infty \implies \mu(B \setminus A) = \mu(B) - \mu(A)$.

More often than not, the task of transforming sets into non-trivial measure spaces is quite complicated. One often tries to bypass the difficulty of choosing an appropriate σ -algebra by constructing an *outer measure* first, which is a special set function defined on *all* subsets of a given set. The price to pay is to relax the countable additivity condition to just countable *sub-additivity*.

Definition 2.5. *An outer measure on a set X is a set function $\bar{\mu} : \mathcal{P}(X) \mapsto [0, \infty]$ such that*

- $\bar{\mu}(\emptyset) = 0$.
- (Monotonicity) $A \subseteq B \implies \bar{\mu}(A) \leq \bar{\mu}(B)$.
- (Countable sub-additivity) $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{P}(X) \implies \bar{\mu}(\bigcup_{n \in \mathbb{N}} A_n) \leq \sum_{n=1}^{\infty} \mu(A_n)$.

Recall that a collection \mathcal{A} of subsets of X is said to cover a subset $B \subseteq X$ if $B \subseteq \bigcup_{A \in \mathcal{A}} A$. Now, if one already has a set function $f : \mathcal{A} \mapsto [0, \infty]$ defined on a family \mathcal{A} of subsets of a given set X which covers X , then the following theorem provides a powerful method to construct an outer measure on X . It should be noted that the assumptions here are quite easily fulfilled in practice. The real line \mathbb{R} with the family of all half-open intervals equipped with the length function $f((a, b]) = b - a$ readily provides an example.

Theorem 2.6. *Given an arbitrary set X along with a family of covering subsets \mathcal{A} and a set function $f : \mathcal{A} \rightarrow [0, \infty]$, there exists a unique outer measure $\bar{\mu}$ on X such that*

- $\bar{\mu}(A) \leq f(A) \quad \forall A \in \mathcal{A}$.
- if $\bar{\nu}$ is another outer measure on X with $\bar{\nu}(A) \leq f(A) \quad \forall A \in \mathcal{A}$, then $\bar{\nu}(B) \leq \bar{\mu}(B) \quad \forall B \subseteq X$.

Moreover, $\bar{\mu}$ is given by the following formula: $\bar{\mu}(B) = \inf \sum_{A \in \mathcal{C}} f(A)$, where the infimum is taken over all countable covers \mathcal{C} of B by sets of \mathcal{A} .

One can think of f in the above theorem as the candidate for defining the measure of certain subsets of X . Then, the theorem presents a way to assign an outer measure to these subsets which is as large as possible, with the constraint that it should not exceed the candidate values. Now, given an outer measure $\bar{\mu}$ on a set X , there is a canonical way to construct a σ -algebra \mathcal{F} on X and an associated measure μ . We present this method in the following theorem.

Theorem 2.7. *Let $\bar{\mu}$ be an outer measure on a set X . A subset $A \subseteq X$ is said to be $\bar{\mu}$ -measurable if $\bar{\mu}(E) = \bar{\mu}(E \cap A) + \bar{\mu}(E \setminus A) \quad \forall E \subseteq X$. The collection \mathcal{F} of all $\bar{\mu}$ -measurable subsets of X forms a σ -algebra on X and the restriction of $\bar{\mu}$ to \mathcal{F} is a measure.*

At this point, it is fair to ask about the structure of the σ -algebra that is generated by any given outer measure upon an application of Theorem 2.7. In particular, since metric spaces X will be our primary spaces of interest in the following section, we would like to get hold of outer measures $\bar{\mu}$ on X which are good enough to guarantee that the corresponding family of all $\bar{\mu}$ -measurable sets at least contains the usual Borel sets generated by the standard metric topology on X . This is not the case in general, for outer measures supplied by Theorem 2.6 above, see [Edg08, Section 5.4] for a counterexample. Fortunately, Theorem 2.6 can be slightly tweaked to remedy this flaw.

Theorem 2.8. *Let X be a metric space and \mathcal{A} be a family of subsets of X such that for every $x \in X$ and $\varepsilon > 0$, there exists $A \in \mathcal{A}$ with $\text{diam } A \leq \varepsilon$ such that $x \in A$. Let $f : \mathcal{A} \rightarrow [0, \infty]$ be a candidate function for the measure. Now, define $\mathcal{A}_\varepsilon := \{A \in \mathcal{A} : \text{diam } A \leq \varepsilon\}$ and let $\bar{\mu}_\varepsilon$ be the outer measure defined using the family \mathcal{A}_ε and the function f , as in Theorem 2.6. Finally, define*

$$\bar{\mu}(B) = \lim_{\varepsilon \rightarrow 0} \bar{\mu}_\varepsilon(B) \quad \forall B \subseteq X$$

Then, $\bar{\mu}$ is an outer measure on X . Moreover, all Borel subsets of X are $\bar{\mu}$ -measurable.

Let us pause for a moment to appreciate the content of the above theorem. Starting from a simple initial candidate function f defined on some suitable family of subsets of X , we now have a systematic procedure to construct a rich σ -algebra \mathcal{F} containing all the Borel sets and a measure $\mu : \mathcal{F} \rightarrow [0, \infty]$ on X . This procedure will be at the heart of the definition of the Hausdorff measure in the next section.

Remark 2.9 (Lebesgue measure). *Before concluding this section, let us remark that the Lebesgue measure on \mathbb{R}^k is constructed by applying Theorems 2.6 and 2.7 on the family of all half-open rectangles of the form*

$$R = (a_1, b_1] \times (a_2, b_2] \times \dots \times (a_k, b_k] \subseteq \mathbb{R}^k$$

and the candidate function $f(R) = \prod_{i=1}^k (b_i - a_i)$, where $a_i < b_i$ for each i . It can be shown that all Borel sets in \mathbb{R}^k are also Lebesgue measurable (notice that in this special case, the construction elaborated in Theorem 2.8 is not required).

3. HAUSDORFF DIMENSION

Equipped with all the required measure-theoretic tools from the previous section, we can now begin to develop the notion of Hausdorff dimension for subsets of an arbitrary metric space X . The process unfolds in two steps. As was remarked in the introduction, the first (and the most challenging step) is to define a sound notion of δ -measurements which can be used to quantify the size of an object for each $\delta > 0$. The previous section informs us that measures are built to fulfil this precise function. However, as noted before, defining a measure from scratch can be an arduous task, so we start by formulating the definition of δ -Hausdorff outer measures. Note that unless stated otherwise, X will always denote a metric space in this section. Recall that a metric space X is a set equipped with a real valued function d on $X \times X$ such that $d(x, y) \geq 0$, $d(x, y) = d(y, x)$, $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$ and $d(x, y) = 0$ if and only if $x = y$. The diameter of a subset $A \subseteq X$ is defined as $\text{diam } A := \sup\{d(x, y) : x, y \in A\}$. For Euclidean spaces $X = \mathbb{R}^k$, we use the standard metric $d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$.

Definition 3.1. Let X be a metric space and $\mathcal{A} = \mathcal{P}(X)$ be the collection of all subsets of X . For each $\delta > 0$, let $f_\delta : \mathcal{A} \rightarrow [0, \infty]$ be the candidate function defined as $f(A) = (\text{diam } A)^\delta \forall A \subseteq X$. Then, the δ -Hausdorff outer measure \bar{H}^δ is constructed by applying Theorem 2.8 on X, \mathcal{A} and f_δ .

Using Theorems 2.6 and 2.8, we spell out the above definition in a bit more detail. For every $\varepsilon, \delta > 0$, we define the family of subsets $\mathcal{A}_\varepsilon = \{A \subseteq X : \text{diam } A \leq \varepsilon\}$ and set functions $f_\delta : \mathcal{A}_\varepsilon \rightarrow [0, \infty]$ such that $f_\delta(A) = (\text{diam } A)^\delta \forall A \in \mathcal{A}_\varepsilon$. Now, the outer measure $\bar{H}_\varepsilon^\delta$ on X is defined by applying Theorem 2.6 on the set family \mathcal{A}_ε and the set function $f_\delta : \mathcal{A}_\varepsilon \mapsto [0, \infty]$. More precisely,

$$\forall B \subseteq X, \quad \bar{H}_\varepsilon^\delta(B) = \inf_{\mathcal{C} \subseteq \mathcal{A}_\varepsilon} \sum_{A \in \mathcal{C}} (\text{diam } A)^\delta \quad (1)$$

where the infimum is taken over all countable ε -covers \mathcal{C} of B , i.e. over all countable covers $\mathcal{C} \subseteq \mathcal{A}_\varepsilon$ of B . The δ -Hausdorff outer measure on X is then defined for all $B \subseteq X$ as the following limit:

$$\bar{H}^\delta(B) = \lim_{\varepsilon \rightarrow 0} \bar{H}_\varepsilon^\delta(B) = \sup_{\varepsilon > 0} \bar{H}_\varepsilon^\delta(B) \in [0, \infty] \quad (2)$$

Although Theorem 2.8 guarantees that \bar{H}^δ is an outer measure for each $\delta > 0$, it is instructive to look at the proof once. We begin by certifying the well-definition of \bar{H}^δ . If $0 < \varepsilon_1 \leq \varepsilon_2$, it is easy to see that $\mathcal{A}_{\varepsilon_1} \subseteq \mathcal{A}_{\varepsilon_2}$. Hence, there are more permissible covers in the definition of $\bar{H}_{\varepsilon_2}^\delta$ when compared to $\bar{H}_{\varepsilon_1}^\delta$, which allows the infimum to decrease. Thus, for arbitrary $B \subseteq X$, we have $\bar{H}_{\varepsilon_2}^\delta(B) \leq \bar{H}_{\varepsilon_1}^\delta(B)$, which implies that $\bar{H}_\varepsilon^\delta(B)$ is a non-increasing function of $\varepsilon \in (0, \infty)$. Now, if $\{\varepsilon_n\}_{n \in \mathbb{N}} \subseteq \mathbb{R}$ converges to zero, then $\{\bar{H}_{\varepsilon_n}^\delta(B)\}_{n \in \mathbb{N}}$ is monotonically increasing and hence either converges to some value in $[0, \infty)$ or diverges to ∞ . It is thus clear that \bar{H}^δ is well defined. We now prove that \bar{H}^δ is indeed an outer measure.

Theorem 3.2. For every $\delta > 0$, \bar{H}^δ is an outer measure on X .

Proof. We begin by showing that for each fixed $\varepsilon > 0$, $\bar{H}_\varepsilon^\delta$ is an outer measure. It is clear from the definition that $\bar{H}_\varepsilon^\delta(\emptyset) = 0$ and that $\bar{H}_\varepsilon^\delta$ is monotonous (if $A \subseteq B$, $\bar{H}_\varepsilon^\delta(A) \leq \bar{H}_\varepsilon^\delta(B)$). To show countable sub-additivity, we consider an arbitrary sequence $\{B_n\}_{n \in \mathbb{N}}$ of subsets of X and let $\gamma > 0$. For each $n \in \mathbb{N}$, the definition of infimum allows us to choose a countable ε -cover \mathcal{C}_n of B_n such that

$$\sum_{A \in \mathcal{C}_n} (\text{diam } A)^\delta \leq \bar{H}_\varepsilon^\delta(B_n) + \frac{\gamma}{2^n}. \quad (3)$$

Now, notice that

$$B_n \subseteq \bigcup_{A \in \mathcal{C}_n} A \implies \bigcup_{n \in \mathbb{N}} B_n \subseteq \bigcup_{n \in \mathbb{N}} \bigcup_{A \in \mathcal{C}_n} A.$$

Hence, we obtain

$$\begin{aligned}
\bar{H}_\varepsilon^\delta(\bigcup_{n \in \mathbb{N}} B_n) &\leq \bar{H}_\varepsilon^\delta(\bigcup_{n \in \mathbb{N}} \bigcup_{A \in \mathcal{C}_n} A) \leq \sum_{n=1}^{\infty} \sum_{A \in \mathcal{C}_n} (\text{diam } A)^\delta \\
&\leq \sum_{n=1}^{\infty} (\bar{H}_\varepsilon^\delta(B_n) + \frac{\gamma}{2^n}) \\
&= \sum_{n=1}^{\infty} \bar{H}_\varepsilon^\delta(B_n) + \gamma
\end{aligned}$$

where, the first inequality follows from the monotonicity of $\bar{H}_\varepsilon^\delta$, the second follows from the definition of $\bar{H}_\varepsilon^\delta$ and the rest are consequences of the above inequality Eq. (3). Since the above inequality holds for all $\gamma > 0$, we infer that $\bar{H}_\varepsilon^\delta(\bigcup_{n \in \mathbb{N}} B_n) \leq \sum_{n \in \mathbb{N}} \bar{H}_\varepsilon^\delta(B_n)$.

Now, we show that \bar{H}^δ is an outer measure. As before, $\bar{H}^\delta(\emptyset) = 0$ and monotonicity trivially follows from the definition. To show countable sub-additivity, we use the recently shown countable sub-additivity of $\bar{H}_\varepsilon^\delta$ and the fact that $\bar{H}^\delta(B_n) = \sup_{\varepsilon > 0} \bar{H}_\varepsilon^\delta(B_n) \geq \bar{H}_\varepsilon^\delta(B_n) \forall n \in \mathbb{N}$ to deduce that

$$\bar{H}_\varepsilon^\delta(\bigcup_{n \in \mathbb{N}} B_n) \leq \sum_{n=1}^{\infty} \bar{H}_\varepsilon^\delta(B_n) \leq \sum_{n=1}^{\infty} \bar{H}^\delta(B_n)$$

Taking the limit $\varepsilon \rightarrow 0$ then yields the desired inequality, thus finishing the proof. \square

Now that we know that \bar{H}^δ is an outer measure on X for each $\delta > 0$, we can immediately obtain the corresponding δ -Hausdorff measure H^δ on the σ -algebra of \bar{H}^δ -measurable sets in X by using Theorem 2.7. Moreover, Theorem 2.8 guarantees that the \bar{H}^δ -measurable sets are sufficiently rich in the sense that all Borel subsets in X are \bar{H}^δ -measurable.

Remark 3.3. *The k -dimensional Lebesgue measure and the k -dimensional Hausdorff measure on \mathbb{R}^k are equivalent, upto multiplication by a constant, see [Edg08, Theorem 6.3.6].*

In the next lemma, we describe a very useful scaling property of Hausdorff measures.

Lemma 3.4. *Let $B \subseteq \mathbb{R}^k$ and let $S : B \rightarrow \mathbb{R}^l$ be Lipschitz, i.e. there exists a $\lambda > 0$ such that $\|S(x) - S(y)\| \leq \lambda \|x - y\| \forall x, y \in B$. Then for all $\delta > 0$,*

$$\bar{H}^\delta(S(B)) \leq \lambda^\delta \bar{H}^\delta(B),$$

where $S(B) := \{S(x) : x \in B\} \subseteq \mathbb{R}^l$.

Proof. If $\mathcal{C} = \{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}_\varepsilon$ is a countable ε -cover of B , then $\lambda\mathcal{C} = \{\lambda A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}_{\lambda\varepsilon}$ is a countable $\lambda\varepsilon$ -cover for $S(B)$. Hence, we have $\bar{H}_{\lambda\varepsilon}^\delta(S(B)) \leq \sum_{n=1}^{\infty} [\text{diam}(\lambda A_n)]^\delta = \lambda^\delta \sum_{n=1}^{\infty} (\text{diam } A_n)^\delta$. Since this is true for all countable ε -covers $\mathcal{C} \subseteq \mathcal{A}_\varepsilon$ of B , taking infimum yields $\bar{H}_{\lambda\varepsilon}^\delta(S(B)) \leq \lambda^\delta \bar{H}_\varepsilon^\delta(B)$. Taking limit $\varepsilon \rightarrow 0$ then proves the desired result. \square

Corollary 3.5. *Let $S : \mathbb{R}^k \rightarrow \mathbb{R}^k$ be a similarity transformation, i.e. $\forall x, y \in \mathbb{R}^k$, $\|S(x) - S(y)\| = \lambda \|x - y\|$ for some $\lambda > 0$. Then, $\bar{H}^\delta(S(B)) = \lambda^\delta \bar{H}^\delta(B)$ for all $\delta > 0$ and $B \subseteq \mathbb{R}^k$.*

Proof. Observe that since S is a similarity transformation, we have $\|S^{-1}(x) - S^{-1}(y)\| = \lambda^{-1} \|x - y\|$ for all $x, y \in \mathbb{R}^k$. The desired result is then obtained by applying Lemma 3.4 on S and S^{-1} . \square

Before proceeding to the second and final step in defining the notion of Hausdorff dimension, we must study the behaviour of $\bar{H}^\delta(B)$ as a function of δ for an subset $B \subseteq X$.

Lemma 3.6. *Let X be a metric space, $B \subseteq X$ and $0 < \delta_1 < \delta_2$. Then,*

- $\bar{H}^{\delta_1}(B) < \infty \implies \bar{H}^{\delta_2}(B) = 0$.
- $\bar{H}^{\delta_2}(B) > 0 \implies \bar{H}^{\delta_1}(B) = \infty$.

Proof. For every $\varepsilon > 0$ and $A \subseteq X$ with $\text{diam } A \leq \varepsilon$, we have

$$|A|^{\delta_2} = |A|^{\delta_2 - \delta_1} |A|^{\delta_1} \leq \varepsilon^{\delta_2 - \delta_1} |A|^{\delta_1}.$$

Hence, for $B \subseteq X$, we infer that $\bar{H}_\varepsilon^{\delta_2}(B) \leq \varepsilon^{\delta_2 - \delta_1} \bar{H}_\varepsilon^{\delta_1}(B)$. Now, if $H^{\delta_1}(B) < \infty$, the desired conclusion follows immediately by taking the limit $\varepsilon \rightarrow 0$. The second conclusion follows similarly. \square

The above theorem tells us that for every $B \subseteq X$, there is a critical value of δ at which $\bar{H}^\delta(B)$ changes from ∞ to 0, see Figure 3. More precisely, there exists a *unique* $d \in [0, \infty]$ such that

$$\bar{H}^\delta(B) = \infty \quad \forall \delta < d \quad \text{and} \quad \bar{H}^\delta(B) = 0 \quad \forall \delta > d.$$

For $\delta = d$, $\bar{H}^\delta(B)$ can take any value from 0 to ∞ . We define this critical value as the *Hausdorff dimension* of B .

Definition 3.7. Let $B \subseteq X$ be a subset of a metric space X . Then, the Hausdorff dimension of B is defined as

$$\dim_H B = \sup\{\delta : \bar{H}^\delta(B) = \infty\} = \inf\{\delta : \bar{H}^\delta(B) = 0\}$$

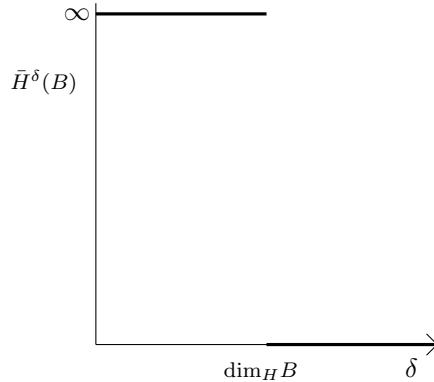


FIGURE 3. At $\delta = \dim_H B$, the value of $\bar{H}^\delta(B)$ drops from infinity to zero.

Let us now pause for a moment to reflect upon the above definition. We are now capable of assigning dimensions to subsets of arbitrary metric spaces in a way which not only generalizes the topological notion of dimension for the usual geometric shapes (see Remark 3.9 in this regard), but at the same time also appreciates the objects' small-scale anatomy by exploiting the underlying distance function. One can think of the Hausdorff dimension as a mathematically rigorous way to encapsulate the traditional intuition that the volume of a surface, for instance, is zero or that the length of a surface is "infinite". If someone hands you a curve in a plane, it is absurd to talk about its (two-dimensional) area. On the other hand, it is perfectly meaningful to talk about the curve's (one-dimensional) length. The δ -Hausdorff measures capture this idea by allowing one to measure the δ -dimensional size of an object for all $\delta > 0$ and not just for integers. In this formulation, the Hausdorff dimension of an object is the right value of δ such that it becomes meaningful to talk about the δ -dimensional size of the object. Since there is no restriction on $\delta > 0$ to be an integer, one often obtains fractional dimensions for numerous exotic self-similar objects. We will investigate this phenomenon more closely in the next section.

To conclude this section, we collect some important properties of the Hausdorff dimension, which can be easily derived from Definition 3.7 and the properties of Hausdorff measures listed in Remark 3.3 and Lemma 3.4.

Theorem 3.8. The Hausdorff dimension enjoys the following properties:

- $\dim_H B = 0$ for any countable B .
- $\dim_H A \leq \dim_H B$ if $A \subseteq B$.
- $\dim_H(A \cup B) = \max\{\dim_H A, \dim_H B\}$.
- $\dim_H B \leq k$ for all subsets $B \subseteq \mathbb{R}^k$.
- $\dim_H f[B] \leq \dim_H B$ for Lipschitz $f : B \mapsto \mathbb{R}^l$, where $B \subseteq \mathbb{R}^k$.
- $\dim_H B = k$ for any subset $B \subseteq \mathbb{R}^k$ which has a positive k -dimensional Lebesgue measure. In particular, $\dim_H B = k$ for all open sets $B \subseteq \mathbb{R}^k$ and $\dim_H \mathbb{R}^k = k$.

Remark 3.9. The above theorem tells us that the topological dimension coincides with the Hausdorff dimension for the familiar geometric shapes like lines, planes, disks, cubes, and so on.

4. EXAMPLES

In this section, we lay out a general protocol to compute the Hausdorff dimension of a large class of self-similar objects. Before delving into the standard method, it is instructive to first look at the quintessential example of a self-similar set: the *Cantor set*.

Example 4.1 (*Cantor set*). Let us first recall the definition of the Cantor set. Beginning with the closed unit interval $C_0 = [0, 1] \subset \mathbb{R}$, we first delete the open middle third interval $(1/3, 2/3)$ from C_0 to obtain $C_1 = [0, 1/3] \cup [2/3, 1]$. Then, the open middle third interval of each of these two closed intervals is deleted to obtain C_2 . Proceeding in this way, we define

$$C_n = \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right) \quad n \in \mathbb{N}.$$

The Cantor set C is then defined as the infinite intersection $C = \bigcap_{n \in \mathbb{N}} C_n$.



FIGURE 4. The first few steps in the construction of the Cantor set.

Being a countable intersection of closed (Borel) sets, the Cantor set itself is a Borel subset of \mathbb{R} and thus lies in the domain of all δ -Hausdorff measures. Moreover, it is clear that C can be written as the disjoint union $C = C' \cup C''$, where $C' = C \cap [0, 1/3] = C/3$ and $C'' = C' + 2/3$. Now, if we consider some δ -Hausdorff measure on \mathbb{R} for $\delta > 0$, we can use the additivity (since H^δ is a measure) and the scaling property of H^δ (see Corollary 3.5) to write

$$H^\delta(C) = H^\delta(C') + H^\delta(C'') = H^\delta(C/3) + H^\delta(C/3 + 2/3) = \frac{2}{3^\delta} H^\delta(C)$$

Assuming $H^\delta(C)$ is finite, we deduce that $2 = 3^\delta \implies \delta = \log 2 / \log 3$. Indeed, it can be checked that for $\delta = \log 2 / \log 3$, $H^\delta(C) = 1$. It is hence evident that $\dim_H C = \log 2 / \log 3$. The fractional Hausdorff dimension of the Cantor set tells us that the set is neither a countable collection of points (i.e. it is uncountably infinite) nor is it a continuous curve. It lies somewhere in between these two, hence the fractional dimension. It would be instructive for the reader to show that the topological dimension of C is zero.

Taking inspiration from the above example, we now make the notion of a self-similar set in \mathbb{R}^k more precise as follows. Consider a closed set $D \subseteq \mathbb{R}^k$ and a set of contractions (also called *iterated function systems*) $\{S_i : D \rightarrow D\}_{i=1}^n$ such that $\|S_i(x) - S_i(y)\| \leq r_i \|x - y\| \forall x, y \in D$, where $0 < r_i < 1$ are known as the *scale factors* of the corresponding contractions S_i . Then, a compact set $F \subseteq D$ is said to be an *attractor* if $F = \bigcup_{i=1}^n S_i(F)$. Since F can be written as a union of shrunken copies of itself, it is clear that F is self-similar. Notice that in Example 4.1, if D is any closed set containing the unit interval $[0, 1]$ and $S_1(x) = x/3$, $S_2(x) = x/3 + 2/3 \forall x \in D$, the Cantor set C becomes the unique attractor $C = S_1(C) \cup S_2(C)$ of the system $\{S_1, S_2\}$.

Now, the fundamental property of such a set of contractions is that it uniquely determines a non-empty compact set $F \subseteq D$ as its attractor, see the remarkable paper [Hut81] or [FF90, Theorem 9.1]. Moreover, if the sets $S_i(F)$ are also pairwise disjoint, then we can apply the reasoning of Example 4.1 to deduce that,

$$H^\delta(F) = \sum_{i=1}^n r_i^\delta H^\delta(F). \quad (4)$$

Notice that since all compact sets are Borel, F lies in the domain of definition of H^δ for every $\delta > 0$. Moreover, if we assume that $H^\delta(F)$ is finite, we can cancel it from both sides of the above equation to infer that the Hausdorff dimension of F is the unique solution d of the following equation:

$$\sum_{i=1}^n r_i^d = 1. \quad (5)$$

Now, in real world scenarios, the attractors F don't usually give rise to pairwise disjoint sets $S_i(F)$ and hence the above reasoning cannot be directly applied in such cases. However, one can relax this *exact pairwise disjointness* condition by an *approximate* one, which amounts to saying that for a given system of contractions $\{S_i : D \rightarrow D\}_{i=1}^n$, there exists a bounded, non-empty open set V such that $\bigcup_{i=1}^n S_i(V) \subseteq V$, where the union is over disjoint sets (this is known as *Moran's open set condition*, see [Edg08, Section 6.5] or [FF90, Section 9.2]). This also guarantees that the δ -Hausdorff measure takes a finite value when $\delta = \dim_H F$. We formulate the above discussion more succinctly in the following theorem and refer the reader to [FF90, Theorem 9.3] for a (rather technical) proof.

Theorem 4.2. *Let $D \subseteq \mathbb{R}^k$ be a compact set. Consider a system of contractions $\{S_i : D \rightarrow D\}_{i=1}^n$ with the corresponding scaling factors $r_i \in (0, 1)$ such that there exists a bounded no-empty open set V with the property that $\bigcup_{i=1}^n S_i(V)$ is a disjoint union contained in V . Then, the Hausdorff dimension of the associated attractor F is given by the unique solution d of the equation*

$$\sum_{i=1}^n r_i^d = 1.$$

Moreover, $H^\delta(F)$ takes a finite value for $\delta = d$.

To conclude this article, we compute the Hausdorff dimension of several self-similar structures to show how Theorem 4.2 is employed in practice.

Example 4.3 (Koch curve). Start with the unit interval $K_0 = [0, 1]$. divide it into 3 equal parts and replace the middle part with an equilateral triangle with its base fixed in the position of the removed interval. Remove the base of the added triangle again to obtain the curve K_1 as a union of four equal intervals. Repeat this step for each of these intervals to obtain K_2 , and iterate this process indefinitely to construct the *Koch curve*, see Figure 5.

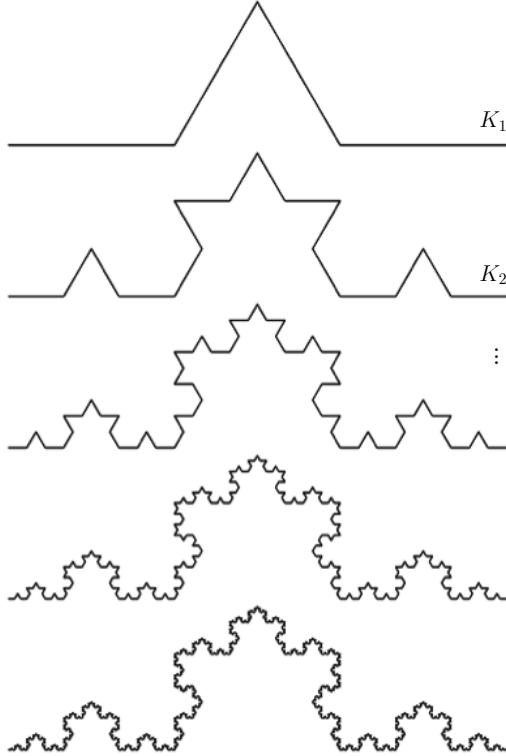
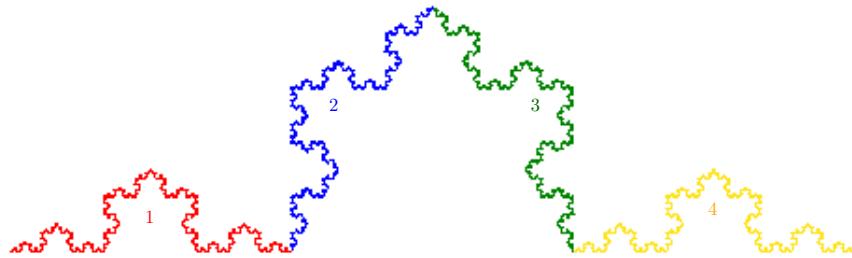


FIGURE 5. The first few iterations in the evolution of the Koch curve

Alternatively, one can define K as the attractor of the contractions mapping the unit interval $[0, 1]$ onto each of the four intervals in K_1 , see Figure 6. Taking V to be the (topological) interior of the isosceles triangle with base $[0, 1]$ and height $1/2\sqrt{3}$, it is clear that these contractions satisfy Moran's open set condition. Finally, since the scaling factor $r = 1/3$ is the same for all 4 contractions, we deduce that the Koch curve has Hausdorff dimension $\dim_H K = \log 4 / \log 3$, see Theorem 4.2.

FIGURE 6. The Koch curve K , expressed as the attractor of the set of 4 contractions $\{S_i\}_{i=1}^4$, each of which maps the unit interval onto one of the 4 intervals in K_1 . Different colored parts of the curve denote the different images $S_i(K)$ of the Koch curve under these contractions, so that $K = \cup_{i=1}^4 S_i(K)$

Example 4.4 (Sierpiński triangle). Start with an equilateral triangle T_0 with base equal to the unit interval $[0, 1]$. Shrink it to an equilateral triangle with half the base length. Make 3 copies of this shrunken triangle and arrange them so that each triangle touches the other two at the corners. (Notice the emergence of the inverted triangle hole in the middle). Call this arrangement T_1 . Repeat the above step for each of the 3 triangles in T_1 to obtain T_2 and iterate this process indefinitely to obtain the *Sierpiński triangle* T , see Figure 7.

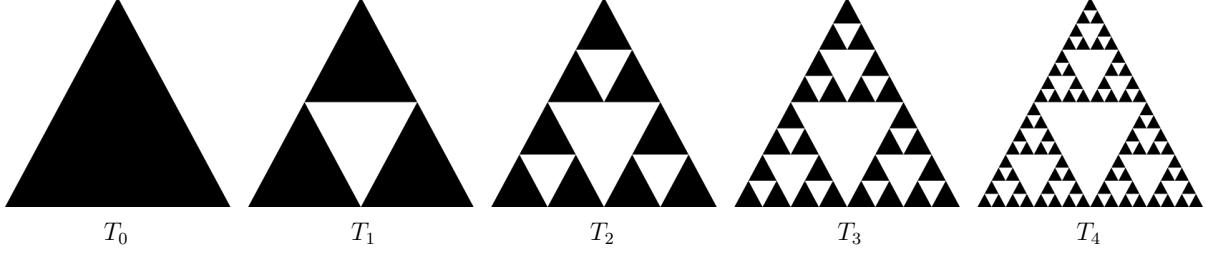
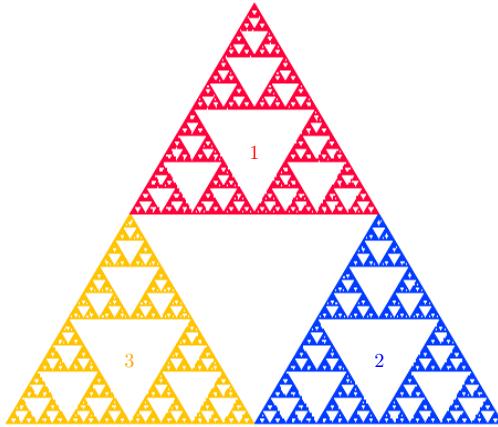


FIGURE 7. The first few iterations in the evolution of the Sierpiński triangle.

It is clear that T can also be defined as the attractor of the set of contractions with scaling factor $r = 1/2$ that map the T_0 triangle onto each of the 3 smaller ones in T_1 . Moreover, Moran's open set condition holds, as can be seen by taking V to be the (topological) interior of T_0 triangle. Hence, we can immediately apply Theorem 4.2 again to obtain the Hausdorff dimension $\dim_H T = \log 3 / \log 2$.

FIGURE 8. The Sierpiński triangle T , expressed as the attractor of the set of 3 contractions $\{S_i\}_{i=1}^3$, each of which maps the T_0 triangle onto one of the 3 triangles in T_1 . Different colored parts of the triangle denote the different images $S_i(K)$ of the Sierpiński triangle under these contraction, so that $T = \cup_{i=1}^3 S_i(T)$.

Remark 4.5. *It would be edifying for the reader to compute the topological dimensions of the objects that have been considered in this section and check that these are strictly less than the corresponding Hausdorff dimensions. In fact, this property is a fundamental feature of fractals, a class of objects with an inherent local roughness and structure built into them which stays undetected through the topological lens of dimension. While some perfectly self-similar sets – like the ones considered in this section – can be classified as fractals, this is not necessarily true in general (consider a line or a plane for instance, which are certainly self-similar but have the same topological and Hausdorff dimensions). Unquestionably, there's more to the theory of fractals than what has been said here, and we refer the interested readers to Mandelbrot's marvellous book [MFC83] for further details.*

REFERENCES

- [Bad12] M. Bader. *Space-Filling Curves: An Introduction with Applications in Scientific Computing*. Texts in Computational Science and Engineering. Springer Berlin Heidelberg, 2012. [1](#)
- [BBE11] Richárd Balka, Zoltán Buczolich, and Marton Elekes. A new fractal dimension: The topological hausdorff dimension. *Advances in Mathematics*, 274, 08 2011. [3](#)
- [Bil86] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 1986. [3](#)
- [Can77] G. Cantor. Ein Beitrag zur Mannigfaltigkeitslehre. *Journal für die reine und angewandte Mathematik*, 84:242–258, 1877. [1](#)

- [Edg08] G. Edgar. *Measure, Topology, and Fractal Geometry*. Undergraduate Texts in Mathematics. Springer New York, 2008. [2](#), [3](#), [5](#), [7](#), [10](#)
- [FF90] K. Falconer and K.A. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. Wiley, 1990. [3](#), [10](#)
- [FM16] M. Fernández-Martínez. A survey on fractal dimension for fractal structures. *Applied Mathematics and Nonlinear Sciences*, 1(2):437 – 472, 01 Jul. 2016. [3](#)
- [Gra85] P. Grassberger. Generalizations of the hausdorff dimension of fractal measures. *Physics Letters A*, 107(3):101 – 105, 1985. [3](#)
- [Hal76] P.R. Halmos. *Measure Theory*. Graduate Texts in Mathematics. Springer New York, 1976. [3](#), [4](#)
- [Hil91] David Hilbert. Ueber die stetige abbildung einer line auf ein flächenstück. *Mathematische Annalen*, 38(3):459–460, September 1891. [1](#)
- [Hut81] John E. Hutchinson. Fractals and self similarity. *Indiana University Mathematics Journal*, 30(5):713–747, 1981. [10](#)
- [Man05] Yuri I. Manin. The notion of dimension in geometry and algebra. *Bulletin of the American Mathematical Society*, 43:139–161, 2005. [3](#)
- [MFC83] B.B. Mandelbrot, W.H. Freeman, and Company. *The Fractal Geometry of Nature*. Einaudi paperbacks. Henry Holt and Company, 1983. [3](#), [12](#)
- [NS14] Samuel Nicolay and Laurent Simons. Building cantor’s bijection, 2014. [1](#)
- [Pea90] G. Peano. Sur une courbe, qui remplit toute une aire plane. *Mathematische Annalen*, 36(1):157–160, March 1890. [1](#)
- [PJS⁺92] H.O. Peitgen, H. Jürgens, D. Saupe, E. Maletsky, and L. Yunker. *Chaos and Fractals: New Frontiers of Science*. Springer-Verlag, 1992. [3](#)
- [RRC87] W. Rudin, W.A. RUDIN, and Tata McGraw-Hill Publishing Company. *Real and Complex Analysis*. Higher Mathematics Series. McGraw-Hill Education, 1987. [3](#)
- [Sch05] Dierk Schleicher. Hausdorff dimension, its properties, and its surprises. *American Mathematical Monthly*, 114, 06 2005. [3](#)

Email address: satviksingh2@gmail.com

DEPARTMENT OF PHYSICAL SCIENCES,
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH (IISER) MOHALI, PUNJAB, INDIA.