Hausdorff dimension and Kolmogorov complexity

Patrick Lutz

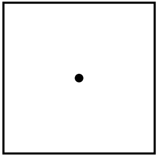
University of Michigan

UM Dearborn Junior Colloquium

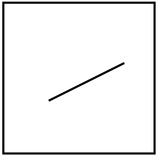
Hausdorff dimension

Question. How can we define the dimension of sets $A \subseteq \mathbb{R}^2$?

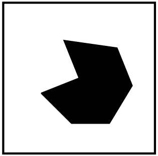
Some easy cases:



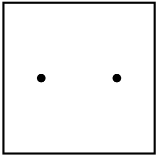
Dimension 0



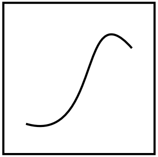
Dimension 1



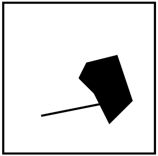
Dimension 2



Dimension 0



Dimension 1

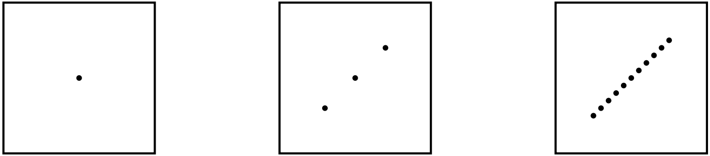


Dimension 2

Question. How can we define the dimension of sets $A \subseteq \mathbb{R}^2$?

Surprisingly difficult to answer in general

A problem. How many points does it take to form a set of dimension 1?



Dimension 0 Dimension 0 Dimension 0?

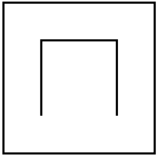
Question. How can we define the dimension of sets $A \subseteq \mathbb{R}^2$?

Surprisingly difficult to answer in general

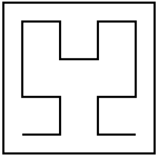
Another problem. Space filling curves

Definition. A curve is a continuous function $g\colon [0,1] \to [0,1]^2$

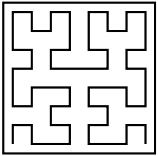
Definition. A space filling curve is a curve whose image is all of $[0,1]^2$



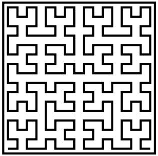
Dimension 1



Dimension 1



Dimension 1

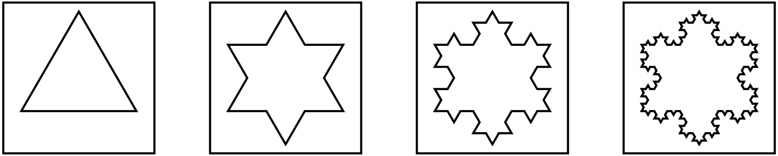


Dimension 1?

Question. How can we define the dimension of sets $A \subseteq \mathbb{R}^2$?

Surprisingly difficult to answer in general

A related problem. Fractals which look neither look "normal" curves nor like space filling curves



The Koch snowflake

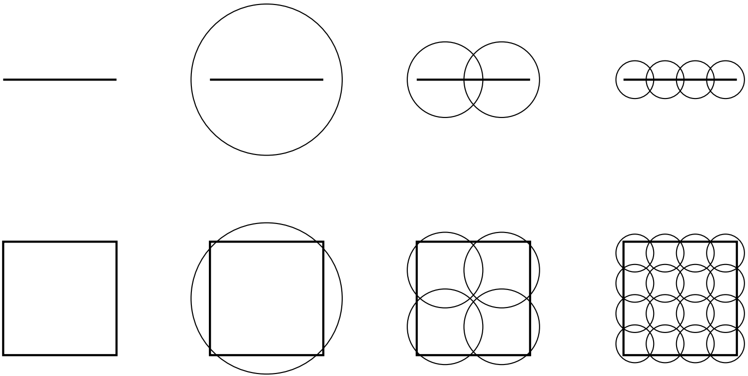
Definitions of "dimension" in mathematics:

- Asymptotic dimension
- Upper Minkowski dimension
- Lower Minkowski dimension
- Packing dimension
- Lebesgue covering dimension
- Small inductive dimension
- Large inductive dimension
- Hausdorff dimension

Hausdorff dimension:

- Matches intuitive notion of dimension in simple examples
- Sets can have fractional dimension, e.g. can have dimension 1.45672
- Assigns plausible fractional dimensions to various fractals
- Leads to lots of interesting mathematics

Idea of Hausdorff dimension. If A has dimension d then it takes $\sim (1/r)^d$ balls of radius r to cover A



Line: $\sim (1/r)$ balls of radius r needed

Square: $\sim (1/r)^2$ balls of radius r needed

Idea of Hausdorff dimension. If A has dimension d then it takes $\sim (1/r)^d$ balls of radius r to cover A

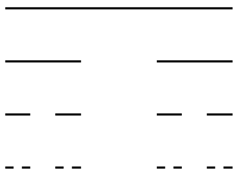
Another example. The Cantor set.

Start with a unit interval

Remove the middle third, giving two intervals

Remove the middle third of each one, giving four intervals

Repeat and take the intersection of all the levels



Comment. Many interesting properties. E.g. uncountable but nowhere dense

Question. What is the dimension of the Cantor set?

Answer. $\log_3(2)$

For balls of radius 3^{-n}, only 2^n are needed to cover the Cantor set

$$(3^n)^d = 2^n \implies d = \log(2)/\log(3) = \log_3(2)$$

Idea of Hausdorff dimension. If A has dimension d then it takes $\sim (1/r)^d$ balls of radius r to cover A

Question. How can we turn this into a formal definition?

A naive attempt. Given a set $A \subseteq \mathbb{R}^2$ and a real number $r > 0$, define $N_r(A)$ to be the smallest number of balls of radius r needed to cover A

Define the dimension of A to be

$$\lim_{r \to 0} \frac{\log(N_r(A))}{\log(1/r)}.$$

Explanation. If it takes $(1/r)^d$ balls of radius r to cover A then

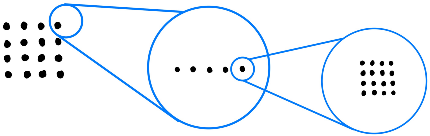
$$\frac{\log(N_r(A))}{\log(1/r)} = \frac{d \log(1/r)}{\log(1/r)} = d.$$

A problem. What if the limit does not exist?

A naive attempt. Given a set $A \subseteq \mathbb{R}^2$ and a real number $r > 0$, define $N_r(A)$ to be the smallest number of balls of radius r needed to cover A Define the dimension of A to be

$$\lim_{r \to 0} \frac{\log(N_r(A))}{\log(1/r)}.$$

A problem. What if the limit does not exist?



A problem. What if the limit does not exist?

Solution. When finding the minimum number of balls needed to cover A, allow balls of different radii and weight each one differently depending on how large its radius is

Definition. A set $A \subseteq \mathbb{R}^2$ has measure 0 in dimension d if for all $\varepsilon > 0$, there is a countable collection of balls $\{B_n\}_{n\in\omega}$ with radii $\{r_n\}_{n\in\omega}$ such that

- $A \subseteq \bigcup_n B_n$
- and $\sum_{n\in\omega} r_n^d < \varepsilon$

Important point. If $d < d'$ and A has measure 0 in dimension d then A also has measure 0 in dimension d'

Definition. The Hausdorff dimension of A, denoted $\dim_H(A)$, is

$$\dim_H(A) = \inf\{d \mid A \text{ has measure 0 in dimension } d\}.$$

Definition. A set $A \subseteq \mathbb{R}^2$ has measure 0 in dimension d if for all $\varepsilon > 0$, there is a countable collection of balls $\{B_n\}_{n\in\omega}$ with radii $\{r_n\}_{n\in\omega}$ such that

- $A \subseteq \bigcup_n B_n$
- and $\sum_{n\in\omega} r_n^d < \varepsilon$

Definition. The Hausdorff dimension of A, denoted $\dim_H(A)$, is

$$\dim_H(A) = \inf\{d \mid A \text{ has measure 0 in dimension } d\}.$$

Intuition: Suppose that for every r, A can be covered with $(1/r)^d$ balls of radius r

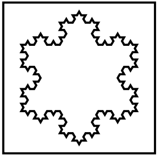
For any $d' > d$,

$$(1/r)^d \cdot r^{d'} = r^{d'-d}$$

which goes to 0 as $r \to 0$. So A has measure 0 in dimension d'.

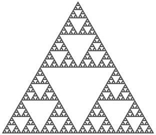
Hausdorff dimension of some interesting sets

Koch snowflake



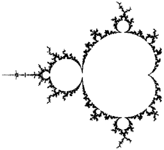
Dimension $= \log_3(4)$

Sierpinski triangle



Dimension $= \log_2(3)$

Boundary of Mandelbrot set

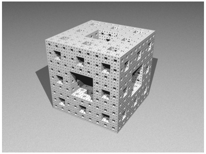


Dimension $= 2$

Coastline of Britain

Dimension ≈ 1.25

Menger sponge



Dimension $= \log_3(20)$

Lung surface

Dimension ≈ 2.97

Most images from Wikipedia

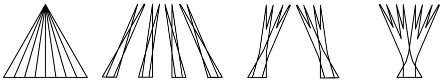
Idea. Hausdorff dimension can be used to measure the "size" of sets whose area is 0

This has led to many interesting conjectures

One example. The Kakeya conjecture

Question (Kakeya). What is the smallest area of a set $A \subseteq \mathbb{R}^2$ which contains a unit line segment in every direction?

Theorem (Besicovitch). There is such an A which has area 0



Kakeya Conjecture. If a set $A \subseteq \mathbb{R}^n$ contains a line segment in every direction, it must have Hausdorff dimension n

Theorem (Davies, 1971). Kakeya conjecture holds $n = 2$

Theorem (Wang and Zahl, 2025). Kakeya conjecture holds for $n = 3$

Kolmogorov complexity

Question. Which of these two strings is more random?

(1) 01010101010101010101010101010101

(2) 00111110101000100011001001011101

Answer. Obviously 2. But what does that even mean?

Question. Is there a way to precisely define what it means for some individual strings to be more random than others?

Idea of Kolmogorov complexity. Define the randomness of a string in terms of how much it can be compressed

Non-random strings have short descriptions, random ones do not

Key idea. Description = computer program

Idea of Kolmogorov complexity. Define the randomness of a string in terms of how much it can be compressed

Definition. The Kolmogorov complexity of a string $\sigma \in 0,1^n$ is the length of the shortest computer program which outputs σ

Denoted $C(\sigma)$

Comment. Assume all programs are encoded in binary

Question. What is the Kolmogorov complexity of $0^n = \underbrace{00\ldots0}_{n}$?

Answer. The following program prints 0^n

```
For i in 1, ..., n:
    Print(0)
```

Length of program

$$= (\text{bits needed to write } n) + (\text{bits needed to write everything else})$$
$$= \log(n) + O(1)$$

So $C(0^n) \leq \log(n) + O(1)$

Definition. The Kolmogorov complexity of a string $\sigma \in 0, 1^n$ is the length of the shortest computer program which outputs σ

Think of a string σ as "random" if $C(\sigma) \approx |\sigma|$

Observation 1. For any string σ, the Kolmogorov complexity of σ cannot be much more than the length of σ

I.e. $C(\sigma) \leq |\sigma| + O(1)$

Because we can just hard-code σ into the program

Observation 2. For most strings σ of length n, $C(\sigma) \geq n - 1$.

I.e. most strings are random

Proof.

Number of strings of length n: 2^n
Number of programs of length at most $n-2$: $\leq \sum_{i=0}^{n-2} 2^i = 2^{n-1} - 1$
Number of strings of length n with complexity at least $n-1$:
$2^n - (2^{n-1} - 1) = 2^{n-1} + 1$

In general: The fraction of length n strings with complexity less than $n - k$ is at most $\frac{1}{2^k}$.

Definition. The Kolmogorov complexity of a string $\sigma \in 0, 1^n$ is the length of the shortest computer program which outputs σ

Almost all strings have high Kolmogorov complexity. But...

Surprising fact. There is no program p such that for any n, $p(n)$ outputs a string with complexity at least n

Proof. Suppose such a program existed. For each n, consider the following program

```
Print(p(n))
```

Length of this program: $\approx |p| + \log(n)$
Let n be large enough that $n > |p| + \log(n)$
Then $C(p(n)) < n$ and $C(p(n)) \geq n$, a contradiction

Even more surprising fact. There is a constant $k \in \mathbb{N}$ such that for no string σ can ZFC prove that $C(\sigma) > k$

It is possible to compute upper bounds for k. E.g. $k \leq 1000000$

Almost every string of length ≥ 20 has complexity ≥ 1000000, but there is no string at all for which this can be proved in ZFC

Hausdorff dimension and Kolmogorov complexity

Definition. Given a real number x, the effective Hausdorff dimension of x, denoted $\mathrm{edim}_H(x)$, is

$$\mathrm{edim}_H(x) = \liminf_{n\to\infty} \frac{C(\text{first } n \text{ bits of } x)}{n}$$

Intuition. $\mathrm{edim}_H(x) =$ information rate of x

I.e. how many bits of information are conveyed by each bit of the binary expansion of x

Example. $x = 0$

For each n, $C(\text{first } n \text{ bits of } x) = C(0^n) \leq \log(n) + O(1)$

So $\mathrm{edim}_H(x) = \liminf_{n\to infty} \frac{\log(n)+O(1)}{n} = 0$.

Example. Form x by picking even bits randomly and setting odd bits to 0

E.g. $x = 0.010000010100\ldots$

Possible to show that $\mathrm{edim}_H(x) = 1/2$.

Only half the bits carry new information.

Definition. Given a real number x, the effective Hausdorff dimension of x, denoted $\mathrm{edim}_H(x)$, is

$$\mathrm{edim}_H(x) = \liminf_{n\to\infty} \frac{C(\text{first } n \text{ bits of } x)}{n}$$

Question. What does this have to do with Hausdorff dimension?

Answer. Originally called "effective Hausdorff dimension" because it has many analogous properties to Hausdorff dimension

Later, an explicit connection was discovered

Theorem (Hitchcock). For many sets $A \subseteq \mathbb{R}$,
$\dim_H(A) = \sup_{x\in A} \mathrm{edim}_H(x)$

The definition of effective Hausdorff dimension can be extended to points in $\mathbb{R}^n$

Hitchcock's theorem still holds and applies to all of the sets we saw earlier

Question. Why does this work?

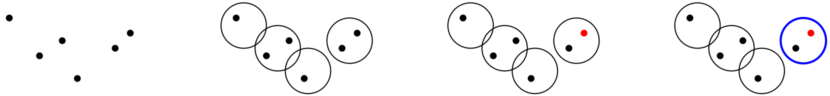
Another way to view Hausdorff dimension: You are playing a game with your friend using the set A:

(1) First you pick an arbitrary point $x \in A$ and an arbitrary $r > 0$

(2) Your goal is to describe x to your friend as concisely as possible

(3) More precisely: you need to give some information to your friend to allow them to guess a point that is within distance r of x and you want to give as little information as possible

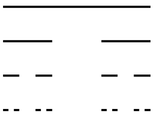
The point: If A can be covered by $(1/r)^d$ balls of radius r, you only need to give your friend $\log((1/r)^d) = d\log(1/r)$ bits of information



Guessing x within distance 2^{-n} corresponds to guessing the first n bits of x's binary expansion, which we can describe using at most dn bits

Theorem (Hitchcock). For many sets $A \subseteq \mathbb{R}$, $\dim_H(A) = \sup_{x \in A} \mathrm{edim}_H(x)$

A concrete example. Recall the Cantor set



Observation. Cantor set = all real numbers in $[0, 1]$ whose ternary expansion does not include the digit 1

For such an x, we can describe each digit of x's ternary expansion using one bit

First n bits of x's binary expansion $\approx$ first $\log_3(2)n$ digits of x's ternary expansion

So $C(\text{first } n \text{ bits of } x) \leq \log_3(2)n$

$\implies \mathrm{edim}_H(x) \leq \log_3(2)n$

Theorem (Hitchcock). For many sets $A \subseteq \mathbb{R}$,
$\dim_H(A) = \sup_{x \in A} \operatorname{edim}_H(x)$

There is an even more general connection between Hausdorff dimension and effective Hausdorff dimension

Idea. We can modify the definition of Kolmogorov complexity to allow programs to ask questions to (and receive answers from) an "oracle" which knows all the answers to some countable set of questions

Kolmogorov complexity relative to an oracle O is denoted $C^O(\sigma)$

Can likewise modify the definition of $\operatorname{edim}_H(x)$ to get $\operatorname{edim}_H^O(x)$

Theorem (J. Lutz and N. Lutz). For every set $A \subseteq \mathbb{R}$,

$$\dim_H(A) = \inf_O \sup_{x \in A} \operatorname{edim}_H^O(x)$$

Called the point-to-set principle: to find the dimension of a set, it is enough to find the effective dimension of each of its individual points