

SPACE-FILLING CURVES

SHIHAN KANUNGO

ABSTRACT. We examine space-filling curves, which are surjective continuous maps from $[0, 1]$ to some higher-dimensional space, usually the unit square $[0, 1]^2$. In particular, we define Peano's curve and Lebesgue's curve, and state some of their properties. We also discuss the Hahn-Mazurkiewicz theorem, which characterizes those subsets of \mathbb{R}^n that are the image of a space-filling curve. Finally, we discuss real-world applications of Hilbert curves, in particular Google's *S2 Cells*.

1. INTRODUCTION

A space-filling curve is a surjective continuous map $f : [0, 1] \rightarrow [0, 1]^n$ for some $n \geq 2$, which can be thought of as a curve inside a unit square that somehow manages to hit every single point. Since a curve has no thickness, it is extremely counterintuitive that it could fill a square. The first indication that this could even be possible came in the mid 19th century, when Georg Cantor (1845 – 1918) proved that $[0, 1]$ and $[0, 1]^2$ have the same cardinality, i.e. there is a way to bijectively map the unit line segment to the unit square. However, his map flings points that are very close in $[0, 1]$ to far-away points in $[0, 1]^2$; so is not continuous. Even so, this result was already so unbelievable that in his 1877 correspondence of this result to his friend and fellow mathematician, Richard Dedekind, Cantor remarked, “I see it, but I don't believe it!”. Naturally, the next question to ask was: is there a *continuous* surjective map $[0, 1] \rightarrow [0, 1]^2$. In other words, is there a space-filling curve?

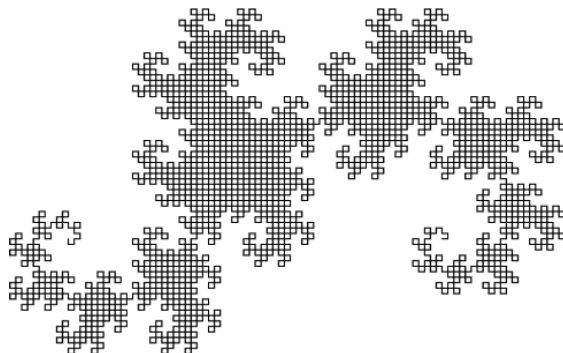


FIGURE 1. The Dragon Fractal

Date: December 13, 2024.

MATH 231A: REAL ANALYSIS I.

Motivated by Cantor's work, Giuseppe Peano (1858 – 1932), an Italian mathematician and professor at the University of Turin, first discovered a space-filling curve in 1890 by constructing a surjective continuous function from $[0, 1]$ to the unit square $[0, 1]^2$. Soon after, many other space-filling curves were discovered, including the Hilbert curve (a variant of Peano's curve), Lebesgue's curve, and the famous example of the *dragon fractal*, shown in Figure 1.

In this paper, we will look at Hilbert's simplified version of the Peano curve, and Lebesgue's curve, another space filling curve $[0, 1] \rightarrow [0, 1]^2$. We also discuss the Hahn-Mazurkiewicz Theorem, which classifies which subsets of \mathbb{R}^n can be the image of some continuous map $f : [0, 1] \rightarrow \mathbb{R}^n$.

2. BASIC CONCEPTS

In this section, we define the two important space-filling curves we will be studying: Peano's curve and Lebesgue's curve.

2.1. Peano's curve. There are many variants of Peano's curve. The one we present is the simplest variant, called the *Hilbert curve*. We will give a visual description of the Hilbert curve, which is constructed iteratively as shown in Figure 2.

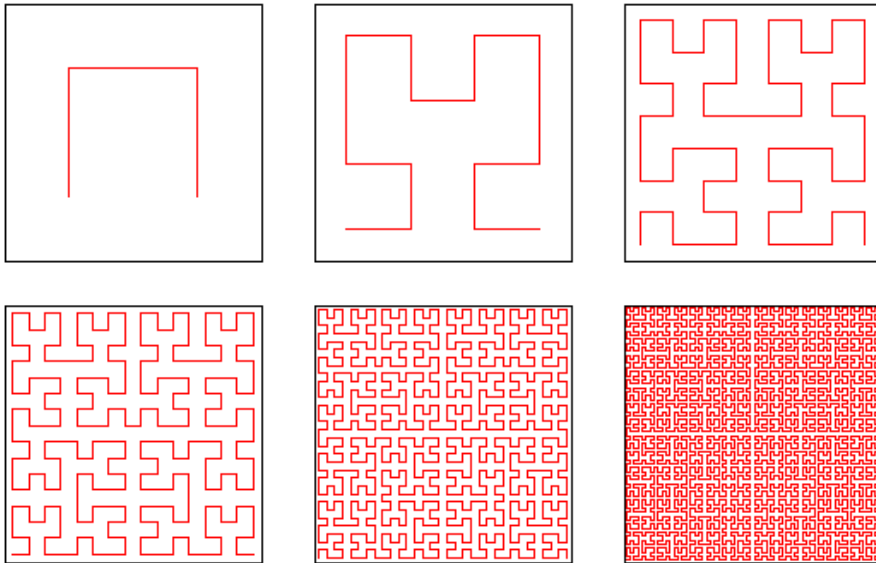


FIGURE 2. The Hilbert Curve

The idea of the construction is similar to the construction of the Cantor set. We let *quartic intervals of generation k* be the intervals obtained when $[0, 1]$ is partitioned into 4^k intervals of equal length. We let I^k denote the set of quartic intervals of generation k . So there are 4^k intervals, each of length 4^{-k} , in I^k , and they all have the form $[\frac{i}{4^k}, \frac{i+1}{4^k}]$.

Each interval in I^k is the union of four intervals in I^{k+1} . Conversely, every interval in I^k is a subset of a unique interval in I^{k-1} . Generalizing this, we make the following definition.

2.1. Definition. A chain of quartic intervals is a sequence $(I_i)_{i=0}^\infty$ of intervals with $I_i \in I^i$ such that

$$I_0 \supset I_1 \supset I_2 \supset \cdots .$$

We have the following elementary properties related to chains.

2.2. Proposition. Chains of quartic intervals satisfy the following properties:

- (a) If $(I_i)_{i=0}^\infty$ is a chain, then there exists a unique $t \in [0, 1]$ such that $t \in I_i$ for all i ,
- (b) For each $t \in [0, 1]$ there is a chain $(I_i)_{i=0}^\infty$ satisfying (a), and
- (c) For all but countably many t , the chain in (b) is unique.

Chains are related to base-4 representation in the following way: Each quartic interval $I \in I^k$ consists of all numbers x that have a base-4 expansion starting with $(\overline{0.a_1a_2 \dots a_k})_4$, where a_1, \dots, a_k are fixed digits between 0 and 3. So we can think of each I_k in the chain as fixing the k -th digit in the base-4 expansion of t . Notice that this argument is similar to one used to prove that the Cantor set consists of those numbers with no 1 in their base-3 expansion.

Now, the Hilbert curve will be constructed iteratively, with each iteration mapping intervals in I^k to a 2-dimensional analog of quaternary intervals.

The *dyadic squares* of generation k are the set of 4^k squares obtained when $[0, 1]^2$ is partitioned into a $2^k \times 2^k$ grid of $2^{-k} \times 2^{-k}$ squares. We let S^k denote the set of dyadic squares of generation k . Mirroring Definition 2.1, we define

2.3. Definition. A chain of dyadic squares is a sequence $(S_i)_{i=0}^\infty$ with $S_i \in S^i$ such that

$$S_0 \supset S_1 \supset S_2 \supset \cdots .$$

We pause here to mention that we are assuming all dyadic squares, as well as all quartic intervals to be closed.

There is an analog of Proposition 2.2 for dyadic squares.

2.4. Proposition. Chains of dyadic squares satisfy the following properties:

- (a) If $(S_i)_{i=0}^\infty$ is a chain, then there exists a unique $x \in [0, 1]^2$ such that $x \in S_i$ for all i ,
- (b) For each $x \in [0, 1]^2$ there is a chain $(S_i)_{i=0}^\infty$ satisfying (a), and
- (c) For all but countably many x , the chain in (b) is unique.

Now, we have the following important definition.

2.5. Definition. A dyadic correspondence is a bijective map ϕ from the set of quartic intervals to the set of dyadic squares such that $\phi(I^k) = S^k$ and if $I \supset J$ then $\phi(I) \supset \phi(J)$.

Note that these exist because $|I^k| = |S^k| = 4^k$. Now, from each dyadic correspondence, we can get a map from $[0, 1]$ to $[0, 1]^2$ as follows: for each $t \in [0, 1]$ there is a chain $(I_i)_{i=0}^\infty$ with $t \in I_i$ for all i . Since ϕ respects inclusion, $(\phi(I_i))_{i=0}^\infty$ is also a chain of dyadic squares. So there is a unique $x \in [0, 1]^2$ with $x \in \phi(I_i)$ for all i . We define $\mathcal{P}(t) = x$. Since ϕ is a bijection, it follows that $\mathcal{P}([0, 1])$ differs from $[0, 1]^2$ by a countable set (as countably many $t \in [0, 1]$ have multiple chains corresponding to t).

However, we don't know whether $\mathcal{P}(t)$ is continuous or not. We will show in the following section that there exists a dyadic correspondence such that the corresponding curve \mathcal{P} is continuous, and this is the Hilbert curve.

2.2. Lebesgue's curve. Lebesgue's space-filling curve is defined using the Cantor set \mathcal{C} , which contains all elements of $[0, 1]$ which don't have a 1 in their ternary expansion. Then, in base-3, each element of \mathcal{C} can be written as $\overline{0.(2a_1)(2a_2)(2a_3)\dots_3}$ where $a_i \in \{0, 1\}$. Recall that the Cantor-Lebesgue function $L : \mathcal{C} \rightarrow [0, 1]$ is defined as

$$L(\overline{0.(2a_1)(2a_2)(2a_3)\dots_3}) = \overline{0.a_1a_2a_3\dots_2} \quad (2.1)$$

for elements of \mathcal{C} and then extending linearly to all of $[0, 1]$. Then L is an example of a continuous function which has derivative 0 almost everywhere, but goes from 0 to 1.

Similarly, we can define $\ell : \mathcal{C} \rightarrow [0, 1]^2$

$$\ell(\overline{0.(2a_1)(2a_2)(2a_3)\dots_3}) = (\overline{0.a_1a_3a_5\dots_2}, \overline{0.a_2a_4a_6\dots_2}), \quad (2.2)$$

which is clearly surjective. Moreover, similarly to the proof that L is continuous, we have:

2.6. Proposition. *The function $\ell : \mathcal{C} \rightarrow [0, 1]^2$ is continuous.*

Finally, we can extend ℓ linearly to a function $\tilde{\ell} : [0, 1] \rightarrow [0, 1]^2$. More formally, for each gap (a_n, b_n) in the Cantor set \mathcal{C} , we define

$$\tilde{\ell}(x) = \ell(a_n) + \frac{x_n - a_n}{b_n - a_n}(\ell(b_n) - \ell(a_n))$$

for $x \in [a_n, b_n]$. Just like the extension of L to $[0, 1]$, this linear extension is also continuous.

2.7. Proposition. *The function $\tilde{\ell} : [0, 1] \rightarrow [0, 1]^2$ is continuous.*

This function $\tilde{\ell}$ is Lebesgue's space-filling curve, which can also be described iteratively like the Hilbert curve. This is shown in Figure 3.

3. MAIN RESULTS

In this section, we look at the existence and some properties of the Hilbert Curve. Then, we state the Hahn-Mazurkiewicz theorem and discuss the ideas of the proof.

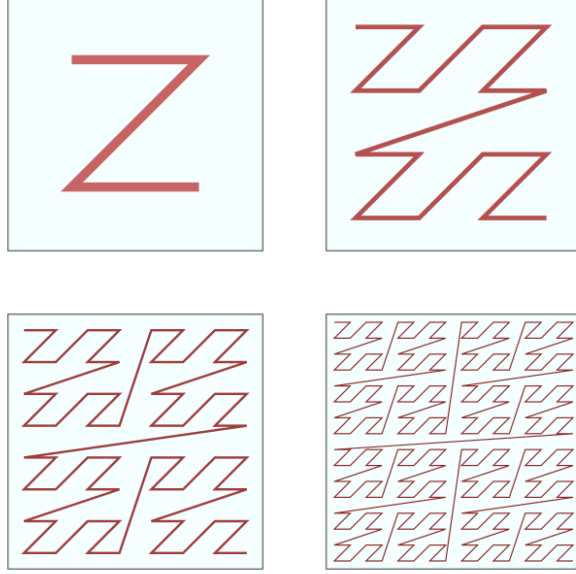


FIGURE 3. The Lebesgue Curve

3.1. Properties of the Hilbert curve. It turns out that there exists a choice of the dyadic correspondence ϕ such that the corresponding map $\mathcal{P} : [0, 1] \rightarrow [0, 1]^2$ satisfies the following nice properties.

3.1. Theorem. *There exists a continuous and surjective map $\mathcal{P} : [0, 1] \rightarrow [0, 1]^2$ such that the following two properties holds*

(a) \mathcal{P} satisfies a Lipschitz condition of exponent $1/2$:

$$\|\mathcal{P}(t) - \mathcal{P}(s)\| \leq M|t - s|^{1/2} \quad (3.1)$$

for some constant M .

(b) For any $[a, b] \subset [0, 1]$, $\mathcal{P}([a, b])$ is a compact subset of $[0, 1]^2$ that has Lebesgue measure of exactly $b - a$.

Then, we can remove sets of measure zero from $[0, 1]$ and $[0, 1]^2$ such that \mathcal{P} becomes a bijection, and then we have that $E \subseteq [0, 1]$ is measurable if and only if $\mathcal{P}(E) \subseteq [0, 1]^2$ is. Additionally, $m(E) = m(\mathcal{P}(E))$. It follows that from a measure-theoretic point of view, $[0, 1]$ and $[0, 1]^2$ are “isomorphic” in some sense. We can also use space-filling curves like these to turn an integral over $[0, 1]^2$ to a one-dimensional integral over $[0, 1]$. In practice however, evaluating such a integral would not be easy.

To prove Theorem 3.1, we use the following result.

3.2. Lemma. *There is a unique dyadic correspondence ϕ such that*

- The explicit construction for ϕ is shown in Figure 4.

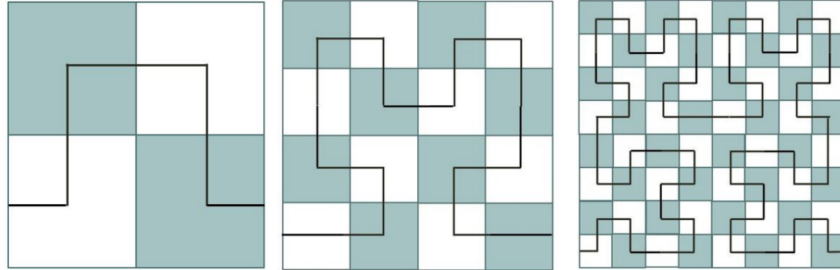


FIGURE 4. Construction of ϕ

To specify ϕ , we simply need to give an ordering of the squares in S^k . Then the i 'th interval in I^k maps to the i 'th square in the ordering. In the figure, we start at the bottom left and then follow the line to get the order of the squares till we end up at the end at the bottom right (Note how this is similar to Figure 2, but with the dyadic squares overlaid under the curve).

With the square (and all its continuous images) revealed as a continuous image of $[0, 1]$, the question arose as to the general structure of a continuous image of a line segment. In 1908, A. Schoenflies found a criterion which, by its very nature, only applies to subsets of the two-dimensional plane and never entered the mathematical mainstream. The further pursuit of this question led to the new topological concept of local connectedness and to a complete answer found independently by S. Mazurkiewicz and H. Hahn in 1913. It turned out that a set is the continuous image of a line segment if and only if it is compact, connected, and locally connected, and this criterion applies not only to \mathbb{R}^n but more generally to Hausdorff spaces.

4. THE HAHN-MAZURKIEWICZ THEOREM

Let's start by defining the key concepts related to the Hahn-Mazurkiewicz theorem.

4.1. Definition. A subset A of \mathbb{R}^n is compact if every sequence has a convergent subsequence with limit in A . By the Bolzano-Weirstrass theorem, this is equivalent to A being closed and bounded. We say that A is connected if there does not exist non-empty disjoint $U, V \in \mathbb{R}^d$ such that $A = A \cap (U \sqcup V)$. Finally, we say A is weakly locally-connected if for all $a \in A$ and $\eta > 0$, there exists $0 < \varepsilon < \eta$ such that if $x \in B_\varepsilon(a)$, then x and a lie in the same connected component of $A \cap B_\eta(a)$.

It turns out that weakly locally-connectedness is equivalent to the more famous notion of locally-connectedness. However, weakly locally-connectedness at a point is *not* the same as locally connectedness at a point. Now, we may state the Hahn-Mazurkiewicz theorem.

4.2. Theorem (Hahn-Mazurkiewicz). *A subset $A \subseteq \mathbb{R}^n$ is the image of some continuous map $f : [0, 1] \rightarrow \mathbb{R}^n$ if and only if A is compact, connected, and weakly locally-connected.*

So this theorem classifies what kinds of sets a space-filling curve can occupy. Proving that the image of any space filling curve $f : [0, 1] \rightarrow \mathbb{R}^n$ is compact, connected, and weakly locally-connected is relatively straightforward. To prove the other direction, we use ideas coming from the construction of Lebesgue's curve.

First, we have Hausdorff's theorem, which is stated as follows.

4.3. Theorem (Hausdorff). *Every compact subset of \mathbb{R}^n is the continuous image of the Cantor set.*

To prove this, we create some sort of analog of the dyadic squares in $[0, 1]^2$. More precisely, we have a series of finer and finer covers of our compact set A such that each ball in some cover is a subset of a ball in a previous cover. Then, we can describe each element of A by a chain of increasingly smaller balls in the covers. Then, for each base-3 representation of $x \in \mathcal{C}$ we can assign an element of A , which gives a surjective map, which is also continuous. This process is similar to (2.1) and (2.2).

So we have a continuous and surjective map $g : \mathcal{C} \rightarrow A$. Then the final step in the proof is to extend this linearly to $[0, 1]$. The idea is that for each gap (a, b) in the Cantor set, instead of simply extending our map linearly between $g(a)$ and $g(b)$ like we did for the Lebesgue curve, we instead find a (continuous) path from $g(a)$ to $g(b)$ that always stays inside our set A . The existence of such a path is made possible by the conditions that A is connected and weakly locally-connected. Then this process yields a continuous function $f : [0, 1] \rightarrow A$, proving the theorem.

However, while we now know many surjective and continuous functions $[0, 1] \rightarrow [0, 1]^2$, and in general from $[0, 1]$ to any compact, connected, and weakly locally-connected subset of \mathbb{R}^n , none of them are injective. It turns out that this is for good reason—it's impossible!

4.4. Theorem (Netto). *Any bijective map $f : [0, 1] \rightarrow [0, 1]^2$ is necessarily discontinuous.*

Of course, the same result holds if we replace $[0, 1]^2$ with \mathbb{R}^n or $[0, 1]^n$ for $n \geq 2$.

5. APPLICATIONS

At first, space-filling curves might seem simply like a nice mathematical construction with some theoretical implications. It definitely doesn't seem like that they have any use in the real world, as after all, it's impossible to make a physical copy of one. But suprisingly, it turns out that they have an extremely important application: in Google maps!

In Google maps, the programmers want a way to have *cache locality*. This means that there is a way to index all the positions on the earth such that if you move a little bit on the earth, then you only move a little bit in the memory. The reason that cache locality is so important is that it speeds up computing time immensely. In fact, it turns out that using the Hilbert curve, we can represent any square centimeter of the earth by a 64-bit integer!

The formal name for this tool is “*S2 cells*,” which were developed by Eric Veach at Google. The idea is as follows. First, we put the earth inside a massive cube. Then, we put a Hilbert curve on each face of the cube such that the ends of curves on adjacent faces connect.

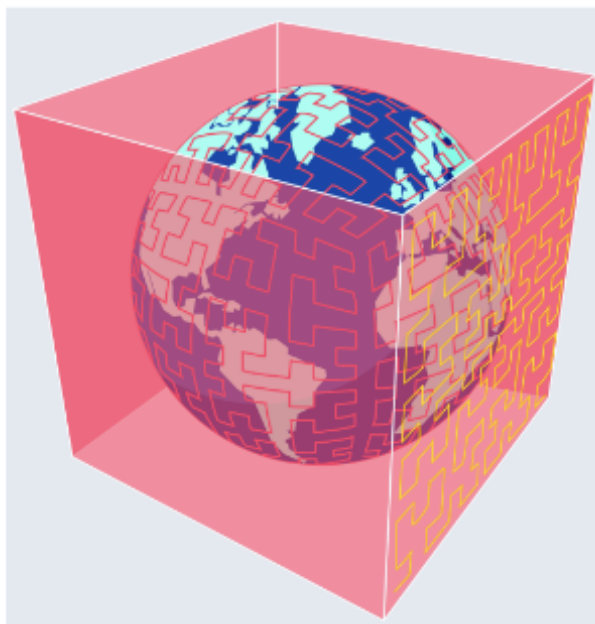


FIGURE 5. Google’s *S2 cells*

Thus, we get a continuous function from a line segment to the surface of the cube, depicted in Figure 5. Next, we use this to create a continuous function from the line segment to the surface of the earth as follows: for a point P on the surface of the earth, draw the ray from the center O of the earth to P , and it will intersect the cube at a point P' . Then, this point P' corresponds to some point on the line segment, and we assign this point to P .

Since the Hilbert curve is continuous, this new map is also continuous. Additionally, if two points are close together on the Hilbert curve, then their corresponding points are close together on the line segment—indeed, this follows from the Lipschitz condition (3.1).

Using the dyadic squares of generation k , we can divide the surface of the earth into $6 \cdot 4^k$ squares (4^k squares from each of the 6 squares), and it only takes $2k + 4$ bits to encode this: 3 bits to specify a number from 0 to 5, corresponding to one of the 6 faces, and then $2k$ bits to encode one of the 4^k squares, and finally a “1” at the end to encode that there are no more bits.

Now, we can use the Hilbert curve (and the corresponding dyadic correspondence) to pick the order of the squares so that close-by squares correspond to close-by numbers in the encoding. To do this, note that each quartic interval is $I = [\frac{i}{4^k}, \frac{i+1}{4^k}]$, where in base-2, i is $\overline{a_1 a_2 \dots a_{2k_2}}$ for $a_j \in \{0, 1\}$. So then, we set the square corresponding to the sequence of bits $a_1 a_2 \dots a_{2k}$ to be $\phi(I)$, where ϕ is our dyadic correspondence shown in Figure 4.

Finally, from each dyadic square on the face, we can assign a region on the earth by projecting it down: for each P' on the square, we send it to the point P on the sphere such that O, P, P' are collinear.

If we set $k = 30$, then each cell is represented by a $2k + 4 = 64$ -bit integer. If we work out the computations, it turns out that each cell has area between 0.48cm^2 and 0.93cm^2 , so we can represent any point on the earth with a 64-bit integer, with accuracy better than a centimeter!

A similar application is how GPUs use Z -order curves (higher-dimensional generalizations of the Lebesgue curve) to code multi-dimensional data into one dimension, while preserving locality (points close together in high dimensions are also close together on the line). Just like $S2$ cells, this makes computation much easier.

Space-filling curves are also used in load balancing: if computational domains need to be more precise in certain locations, space filling curves can deal with this by refining the curve in those places. For example, we might only need the generation 8 Hilbert curve for the top left, bottom left, and bottom right quadrants of the unit square, but in the top right we might need generation 15 instead. For an example of this used in practice, see [3].

REFERENCES

- [1] H. Haverkort, *Recursive tilings and space-filling curves*, [[Link](#)] 2024.
- [2] Alexander Kupers, *On space-filling curves and the Hahn-Mazurkiewicz theorem*, [[Link](#)] 2012.
- [3] Hui Liu, Kun Wang, Bo Yang, Min Yang, Ruijian He, Lihua Shen, He Zhong, and Zhangxin Chen, *Load balancing using Hilbert space-filling curves for parallel reservoir simulations*.
- [4] Christian S. Perone, *Google's S2, Geometry on the Sphere, cells and Hilbert curves*, [[Link](#)] Terra Incognita, 2015.
- [5] S. Kreiss, *S2 cells and space-filling curves: Keys to building better digital map tools for cities*, [[Link](#)] Medium, 2016.
- [6] H. Sagan, *Space-filling curves*, Springer New York, NY, 1994.
- [7] E M Stein and R Shakarchi, *Real analysis: measure theory, integration, and Hilbert spaces*, Princeton lectures in analysis, Princeton Univ. Press, Princeton, NJ, 2005.