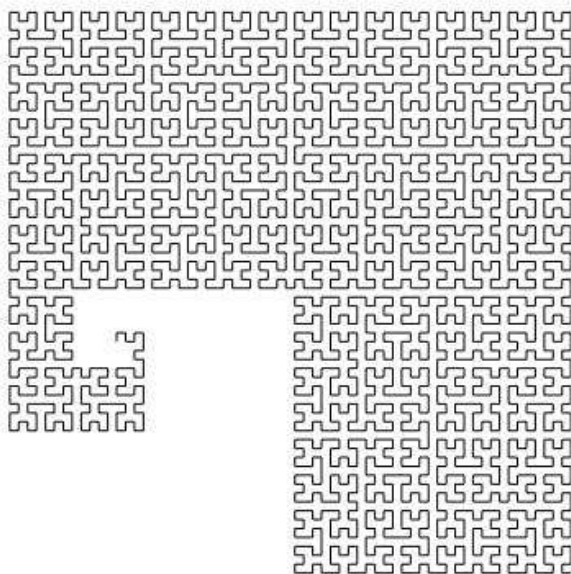


An Introduction to Space-Filling Curves

1407493



Contents

1	Introduction	3
1.1	Notation and Definitions	4
2	Peano's Space-Filling Curve	5
2.1	Definition	5
2.2	Properties	5
3	Hilbert's Space-Filling Curve	12
3.1	Definition	13
3.2	Properties	13
3.3	Complex and Arithmetic Represenation	14
3.4	Generalisation to 3-Dimensions	16
4	Lebesgue's Space-Filling Curve	17
4.1	Definition	18
4.2	Properties	19

Abstract

This paper should be interpreted as an introduction to the concept of space filling curves. Looking at examples created by Peano, Hilbert and Lebesgue. Such curves have motivation in computer science as an efficient way to traverse multi-dimensional data. The contents throughout is heavily influenced by what is one of the only books to really study the subject, Han's Sagan's *Space Filling Curves* [1].

1 Introduction

We start with a definition fundamental to geometry.

Definition 1. [2] An *n-dimensional manifold* is a topological space \mathcal{M} for which every point $x \in \mathcal{M}$ has a neighbourhood homeomorphic to Euclidean space \mathbb{E}^n .

That is for each point in \mathcal{M} we can find a neighbourhood \mathcal{X} such that there exist's a continuous mapping

$$f : \mathcal{X} \rightarrow \mathbb{E}^n$$

Where the inverse f^{-1} is also continuous.

The following description adapted from Wikipedia [3] I believe helps understand the idea. If we describe these invertible maps from \mathbb{E}^n to each of our subspaces \mathcal{X} as a *chart*, with the collection of them being an atlas. We can consider the transition maps between any two charts, i.e. the composition of one chart with the inverse of another. If the transition maps for every chart are differentiable then the manifold \mathcal{M} is said to be a differentiable manifold, further specialisation of this idea is the smooth manifold for which every transition map is C^∞ (the derivatives of all orders exist). Examples of such smooth manifolds are the closed unit-hypercubes (the n-dimensional analogue of the unit square), the Möbius Strip, or even circles and lines in 2 dimensions.

In 1878 the German mathematician George Cantor proved that for any two finite-dimensional smooth manifolds there exists a bijective mapping from one to the other. That is the cardinalities of any two finite-dimensional smooth manifolds are equal. This has a rather remarkable consequence that the interval $[0, 1]$ can be mapped bijectively onto the square $[0, 1] \times [0, 1]$. This can be demonstrated easily as I've done below. A more in depth construction can be found at [4, p. 3]

Theorem 1. *There exists a bijection $f : [0, 1] \rightarrow [0, 1]^2$*

Proof. We represent $(x, y) \in [0, 1]^2 \setminus \{(0, 0)\}$ by it's unique infinite decimal expansion.

$$(x, y) = (0.x_1x_2x_3\dots, 0.y_1y_2y_3\dots)$$

Such that we dont end in a string of infinite zeros, i.e $\forall N \in \mathbb{N}, \exists p, q > N$ such that $x_p \neq 0$ and $y_q \neq 0$. In this way 0.36 would be represented 0.359999...

Then consider the function

$$f(x, y) = \begin{cases} 0 & \text{if } (x, y) = (0, 0) \\ 0, x_1y_1x_2y_2x_3y_3\dots & \text{otherwise} \end{cases}$$

Then clearly as each decimal expansion is unique this is an injective function which by the Schröder-Bernstein theorem implies that $|[0, 1]^2| \leq |[0, 1]|$, but as $[0, 1] \subset [0, 1]^2$ then we also have that $|[0, 1]| \leq |[0, 1]^2|$. Hence $|[0, 1]| = |[0, 1]^2|$ and so there must exist a bijection from $[0, 1] \rightarrow [0, 1]^2$ \square

The immediate question after this result is whether such a mapping can be continuous, it was Netto in 1879 who showed this couldn't be so.

Theorem 2 (Netto). [1, p. 6] *If f represents a bijective map from an m -dimensional smooth manifold μ_m onto an n -dimensional smooth manifold μ_n and $m \neq n$ then f is necessarily discontinuous.*

The next progression was to consider whether it would be possible to find a continuous surjective mapping (losing the condition of injectivity). In 1890 Peano constructed the first such curve, these curves are now referred to as space filling curves.

1.1 Notation and Definitions

The following notation has been used as to mimic Sagan [1].

For the purpose of this text we will use the following letters to denote common subsets of \mathbb{R}^n

$\mathcal{I} = [0, 1]$ the closed unit interval

$\mathcal{Q} = [0, 1]^2$ the closed unit square

$\mathcal{W} = [0, 1]^3$ the closed unit cube

We denote binaries by

$$0_2 b_1 b_2 b_3 \dots = b_1/2 + b_2/2^2 + b_3/2^3 + \dots, \quad b_j = 0 \text{ or } 1$$

,ternaries by

$$0_3 t_1 t_2 t_3 \dots = t_1/3 + t_2/3^2 + t_3/3^3 + \dots, \quad t_j = 0, 1 \text{ or } 2$$

and quaternaries by

$$0_4 q_1 q_2 q_3 \dots = q_1/4 + q_2/4^2 + q_3/4^3 + \dots, \quad q_j = 0, 1, 2 \text{ or } 3$$

Definition 2. [1, p. 4] If $f : \mathcal{I} \rightarrow \mathbb{E}^n$ is continuous then the image of f , $f(\mathcal{I})$ is called a curve. $f(0)$ is called the beginning point of the curve and $f(1)$ is called its endpoint.

Note that here we have used the domain of \mathcal{I} , we could have used any closed interval on the real line, however for our purposes this is the only example we will be discussing.

An important concept for the definition of a space filling curve is that of Jordan content, this is simply a measure of size generalised to n -dimensional space, e.g. area for $n=2$, volume for $n=3$. Then we can define an SFC

Definition 3. [1, p. 5] If $f : \mathcal{I} \rightarrow \mathbb{E}^n$, $n \geq 2$, is continuous and its image $f(\mathcal{I})$ has strictly positive Jordan content then $f(\mathcal{I})$ is called a space-filling curve.

This definition can be rewritten in what I believe to be a more intuitive way using the idea of surjectivity from Bader [5, p. 17].

Definition 4 (Alternative SFC Definition). If $f : \mathcal{I} \rightarrow \mathbb{E}^n$, $n \geq 2$, is continuous and is surjective on some region $A \subset \mathbb{E}^n$ with positive jordan content then $f[\mathcal{I}]$ is called a space-filling curve.

It can clearly be seen that in this case surjectivity is a sufficient condition for positive jordan content. The curves we are studying have the image $[0, 1]^2$ clearly a region with positive Jordan content so any surjective and continuous map onto it is space-filling.

2 Peano's Space-Filling Curve

This definition is the one found in Valgaerts [6, p. 7]

2.1 Definition

It was Giuseppe Peano in 1890 who found the first such curve defined using ternaries and the operator

$$kt_j = 2 - t_j \quad (t_j = 0, 1, 2)$$

Where k^v is the v^{th} iterate of k , that is k composed with itself v times. He then defined $f_p : \mathcal{I} \rightarrow \mathcal{Q}$ as

$$f_p(0_3 t_1 t_2 t_3 t_4 \dots) = \begin{pmatrix} 0_3 t_1 (k^{t_2} t_3) (k^{t_2+t_4} t_5) \dots \\ 0_3 (k^{t_1} t_2) (k^{t_1+t_3} t_4) \dots \end{pmatrix}$$

2.2 Properties

¹ From this definition we can check that this does in fact follow the definition of an SFC, using the alternative definition we can first show that f_p is *surjective*:

Theorem 3. *[1, p. 30] f_p as defined above is surjective*

Proof. Consider

$$(0_3 \alpha_1 \alpha_2 \alpha_3 \dots, 0_3 \beta_1 \beta_2 \beta_3 \dots) \in \mathcal{Q}$$

Then we want to find $t = 0_3 t_1 t_2 t_3 t_4 \dots \in \mathcal{I}$ such that

$$f_p(t) = \begin{pmatrix} 0_3 \alpha_1 \alpha_2 \alpha_3 \dots \\ 0_3 \beta_1 \beta_2 \beta_3 \dots \end{pmatrix}$$

By our definition of f_p through comparing each ternary place we can see that

$$\alpha_n = k^{t_0+t_2+\dots+t_{2n-2}} t_{2n-1}, \quad \beta_n = k^{t_1+t_3+\dots+t_{2n-1}} t_{2n}$$

Then noting the fact that clearly by its definition k is its own inverse we get.

¹Through §2.2 I follow very closely Sagans writing on the properties and geometric interpretation of the Hilbert Curve

$$t_{2n-1} = k^{t_0+t_2+\dots+t_{2n-2}} \alpha_n, \quad t_{2n} = k^{t_1+t_3+\dots+t_{2n-1}} \beta_n$$

Which we can solve iteratively for each t_i .

□

This original definition by Peano gave no geometric representation of the curve, it wasn't in fact until Hilbert devised his curve that any sort of geometric interpretation was given for a space filling curve. Through Hilbert's generating principle however we can derive such a visualisation for Peano's original curve now, and in doing so the proof that the map is in fact continuous will become much simpler.

If we consider any ternary of the form

$$t = 0_3 00 t_3 t_4 t_5 \in \left[0, \frac{1}{9}\right]$$

Then by its definition

$$f_p(t) = \begin{pmatrix} 0_3 0 \xi_2 \xi_3 \xi_4 \dots \\ 0_3 0 \eta_2 \eta_3 \eta_4 \dots \end{pmatrix}$$

i.e. $f_p(t) \in \left[0, \frac{1}{3}\right]^2$ the subsquare labelled 1 in figure 1 below.

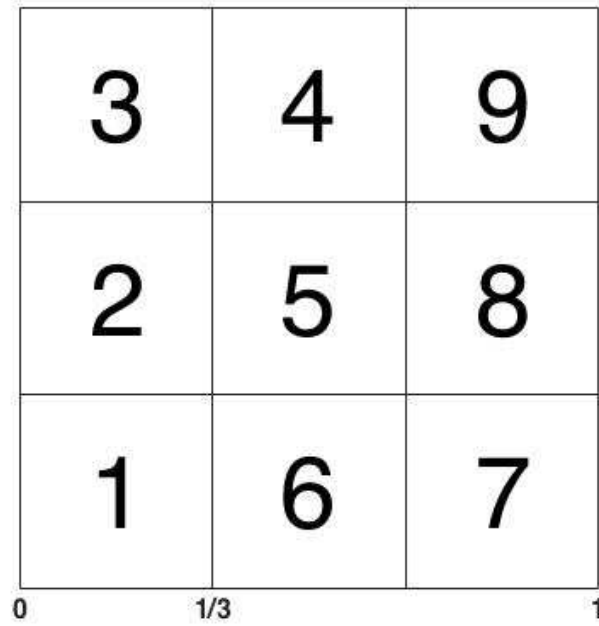


Figure 1: Ordering of the peano curve

Similarly

$$f_p(0_3 01 t_3 t_4 t_5 \dots) = \begin{pmatrix} 0_3 0 \xi'_2 \xi'_3 \xi'_4 \dots \\ 0_3 1 \eta'_2 \eta'_3 \eta'_4 \dots \end{pmatrix}$$

so the interval $[\frac{1}{9}, \frac{2}{9}]$ is mapped into the second subsquare $[0, \frac{1}{3}] \times [\frac{1}{3}, \frac{2}{3}]$ continuing in this way we note that the interval $[\frac{j-1}{9}, \frac{j}{9}]$, $j = 1, 2, 3, \dots, 9$ is mapped into the subsquare j . In fact if we subdivide \mathcal{I} into 3^{2n} congruent subintervals then we can map them each into 3^{2n} subsquares of \mathcal{Q} in the order we have described below.

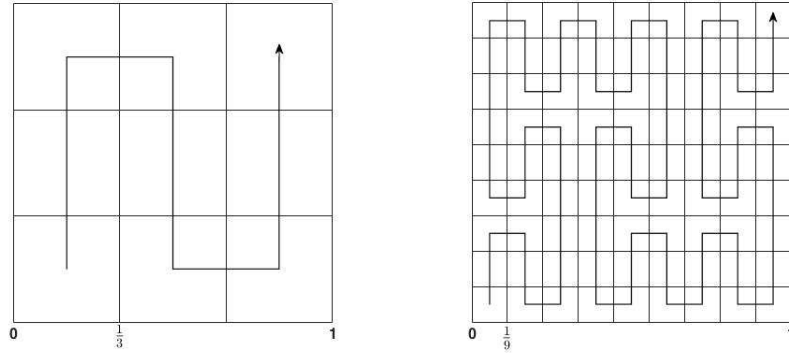


Figure 2: Generated Peano curves of order 1 and 2

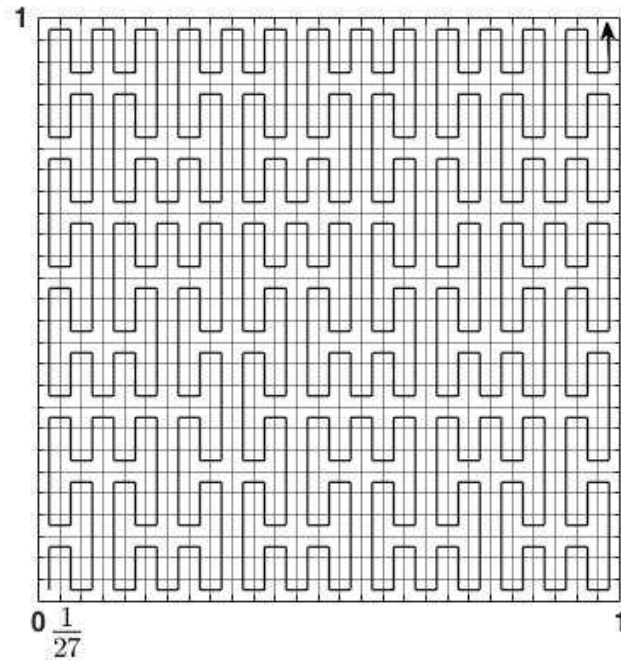


Figure 3: Generated Peano curve of order 3

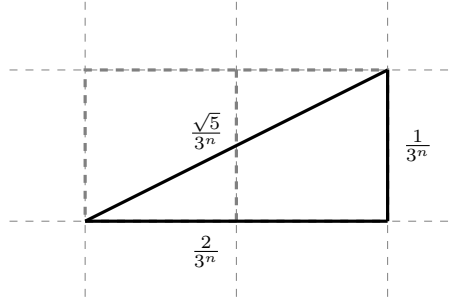
Then this maps adjacent intervals to adjacent subsquares sharing an edge, and each iteration preserves the previous one. By Hilberts generating process

this suggests that the mapping we've just defined does in fact represent an SFC, in fact it represents the peano curve. The proof of this is a rather simple but tedious induction which I have opted to ommit for the sake of brevity, it can be found in Sagan [1, p. 37]

Knowing that these maps are in fact equivalent makes the proof that the Peano curve is continuous much simpler.

Theorem 4. *The Peano curve is continuous*

*Proof.*² Considering the n^{th} iteration in the generation of the Peano curve, denoted f_n we've seen that \mathcal{I} is subdivided into 3^{2n} subintervals. So if we consider $x_1, x_2 \in \mathcal{I}$ such that $|x_1 - x_2| < \frac{1}{3^{2n}}$ that is the length of one of these subintervals then the interval $[x_1, x_2]$ will intersect at most 2 subintervals and hence its image will be contained in 2 adjacent subsquares of side length $\frac{1}{\sqrt{3^{2n}}} = \frac{1}{3^n}$



Then the greatest distance between $f(x_1)$ and $f(x_2)$ is the diagonal of the rectangle formed by adjoining the two squares as shown above. So by Pythagoras

$$|f_n(x_1) - f_n(x_2)| < \frac{\sqrt{5}}{3^n}$$

We have seen that

$$f_p(x) = \lim_{n \rightarrow \infty} f_n(x)$$

then

$$|f_p(x_1) - f_p(x_2)| = \lim_{n \rightarrow \infty} |f_n(x_1) - f_n(x_2)| = \lim_{n \rightarrow \infty} \frac{\sqrt{5}}{3^n} = 0$$

hence the Peano curve is continuous. \square

And so finally we have shown that the Peano curve is both surjective and continuous and hence is a Space Filling Curve.

We can now investigate further properties of the curve. It is possible, with not too much thought, to alter the definition of differentiability given in MA131 Analysis II [7] to involve sequences as below.

Definition 5 (Condition for Non-Differentiability). Let $f : \mathcal{I} \rightarrow \mathbb{R}$. If there are two sequences $\{a_n\}, \{b_n\}$ with $0 < a_n \leq t < b_n < 1$, and $\{a_n\} \rightarrow t, \{b_n\} \rightarrow t$, such that

$$\lim_{n \rightarrow \infty} \frac{f(a_n) - f(b_n)}{a_n - b_n}$$

²This proof is adapted from the one Sagan used to show continuity of the Hilvert Curve. [1, p. 12]

does not exist, then f is *not* differentiable at t .

With this definition in mind it is quite simple to show that the Peano curve is non-differentiable everywhere, the proof follows Sagan [1]

Theorem 5. *The Peano curve is nowhere differentiable.*

Proof. Consider $t = 0_3 t_1 t_2 t_3 \dots t_{2n} t_{2n+1} t_{2n+2} \dots \in [0, 1]$. Then define $t_n = 0_3 t_1 t_2 t_3 \dots t_{2n} \tau_{2n+1} t_{2n+2} \dots$, where $\tau_{2n+1} = t_{2n+1} + 1 \pmod{2}$. We now have two sequences $\{t\}$ and $\{t_n\}$ satisfying the conditions in the definition given above. Note that $|t - t_n| = \frac{1}{3^{2n+1}}$. Consider now

$$\varphi_p(t) = 0_3 t_1 (k^{t_2} t_3) (k^{t_2+t_4} t_5) \dots$$

Then $\varphi_p(t)$ and $\varphi_p(t_n)$ only differ in the $(n+1)$ -th ternary place and so,

$$|\varphi_p(t) - \varphi_p(t_n)| = |k^{t_2+\dots+t_{2n}} t_{2n+1} - k^{t_2+\dots+t_{2n}} \tau_{2n+1}| / 3^{n+1} = \frac{1}{3^{n+1}}$$

then clearly

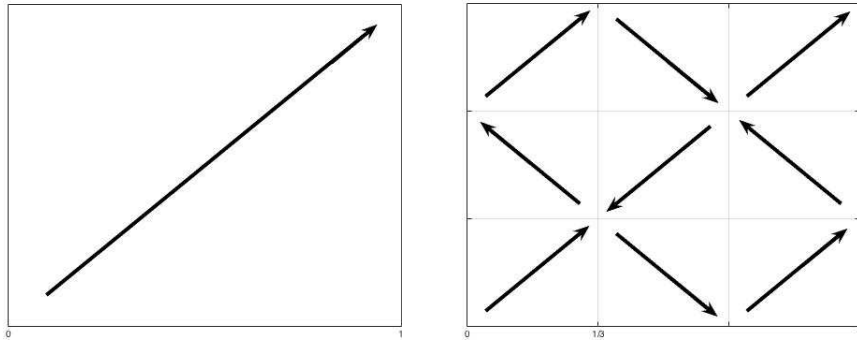
$$\frac{|\varphi_p(t) - \varphi_p(t_n)|}{|t - t_n|} = 3^n$$

And so as $\lim_{n \rightarrow \infty} 3^n$ doesn't exist φ_p is nowhere differentiable. We can then consider $\psi_p(t) = 0_3 (k^{t_1} t_2) (k^{t_1+t_3} t_4) \dots$. Then it isn't hard to convince yourself that $\psi_p(t) = 3\varphi(\frac{t}{3})$ and so ψ_p is also nowhere differentiable. From these two results we can note that

$$f_p(t) = \begin{pmatrix} \varphi_p(t) \\ \psi_p(t) \end{pmatrix}$$

and so the Peano curve is nowhere differentiable. \square

Now we have a geometric definition of Peano's curve it will be a useful exercise to formalise this in terms of transformations in the complex plane. Doing this we can easily create figures as those above and when we come to study the Hilbert curve it will give us a means to actually calculate the image of a point. The first thing we have to notice is that the geometric curve described above has its start point at $(0, 0)$ and then concludes at the corner of \mathcal{Q} , $(1, 1)$. This fact will dictate the orientation of our subsquares as the exit point for each subsquare must coalesce with the entry point on the proceeding subsquare. This idea is illustrated below using the ordering from earlier.



From this we can start to come up with the transformations that will map the square $[0, 1] \times [0, i]$ to the subsquares described in the figure. The reason we can describe this in the complex plane is due to the isomorphism that exists between \mathbb{C} and \mathbb{R}^2 , and so $[0, 1] \times [0, i]$ is isomorphic to \mathcal{Q} . I include a second copy of figure to act as reference of the ordering.

3	4	9
2	5	8
1	6	7

0 1/3 1

We first see that subsquares 1,3,7, and 9 are simply obtained by scaling \mathcal{Q} by a factor of $\frac{1}{3}$ about the origin and translating accordingly so no reorientation is necessary. Subsquares 5 can be derived by a similar scaling and translation and also a rotation of 180° . Subsquares 2 and 8 require a reflection in the imaginary axis along with the scale and translation, and similarly 4 and 6 are obtained via a reflection in the real axis, scaling and translation. We can now summarise these transformations below where \mathfrak{B}_i represents the transformation necessary to acquire subsquare $i + 1$ from $[0, 1] \times [0, i]$. The reason for this mismatched labelling of our transformations to our subsquares should become apparent soon.

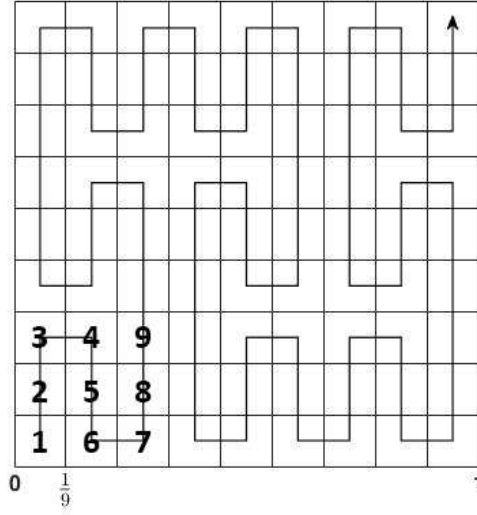
$$\begin{array}{lll}
\mathfrak{B}_0 z = \frac{1}{3} z & \mathfrak{B}_1 z = -\frac{1}{3} \bar{z} + \frac{1}{3} + \frac{i}{3} & \mathfrak{B}_2 z = \frac{1}{3} z + \frac{2i}{3} \\
\mathfrak{B}_3 z = \frac{1}{3} \bar{z} + i + \frac{1}{3} & \mathfrak{B}_4 z = -\frac{1}{3} z + \frac{2}{3} + \frac{2i}{3} & \mathfrak{B}_5 z = \frac{1}{3} \bar{z} + \frac{1}{3} + \frac{i}{3} \\
\mathfrak{B}_6 z = \frac{1}{3} z + \frac{2}{3} & \mathfrak{B}_7 z = -\frac{1}{3} \bar{z} + 1 + \frac{i}{3} & \mathfrak{B}_8 z = \frac{1}{3} z + \frac{2}{3} + \frac{2i}{3}
\end{array}$$

From this what we can notice that if we consider the *nonary* form of t (the base 9 form of t) that is $t = 0_9 q_1 q_2 q_3 \dots$ then in the first iteration of the geometric Peano curve t lies in the $(q_1 - 1)th$ subsquare and in the second iteration it will lie in the $(q_2 - 1)th$ subsquare of the $(q_1 - 1)th$ subsquare of the first iteration. Whilst this may seem confusing at first hopefully an example should clear it up.

$$\text{Consider } t = \frac{1}{27} = 0_9 03$$

$$\text{Then } q_1 = 0, q_2 = 3$$

And so $f_1(t)$ lies in the 1st subsquare in figure and then $f_2(t)$ will lie in the 4th subsquare in the figure below.



With this remark in mind we can see that

$$f_p(t) = \lim_{n \rightarrow \infty} \mathfrak{B}_{q_1} \mathfrak{B}_{q_2} \mathfrak{B}_{q_3} \dots \mathfrak{B}_{q_n} \mathcal{Q}$$

We can in fact modify this further to keep in line with our ternary representation we have been using by noting that

$$t = 0_3 t_1 t_2 t_3 \dots t_{2n-1} t_{2n} \dots = 0_{\mathfrak{g}} (3t_1 + t_2)(3t_3 + t_4) \dots (3t_{2n-1} + t_{2n}) \dots$$

and so

$$f_p(t) = \lim_{n \rightarrow \infty} \mathfrak{B}_{3t_1+t_2} \mathfrak{B}_{3t_3+t_4} \mathfrak{B}_{3t_5+t_6} \dots \mathfrak{B}_{3t_{2n-1}+t_{2n}} \mathcal{Q}$$

We can also notice that if we only consider the decimal of a ternary to a finite place then

$$f_p(0_3 t_1 t_2 t_3 \dots t_{2n-1} t_{2n}) = \mathfrak{B}_{3t_1+t_2} \mathfrak{B}_{3t_3+t_4} \mathfrak{B}_{3t_5+t_6} \dots \mathfrak{B}_{3t_{2n-1}+t_{2n}} \mathfrak{B}_0 \mathfrak{B}_0 \mathfrak{B}_0 \dots \mathcal{Q}$$

$$f_p(0_3 t_1 t_2 t_3 \dots t_{2n-1} t_{2n}) = \mathfrak{B}_{3t_1+t_2} \mathfrak{B}_{3t_3+t_4} \mathfrak{B}_{3t_5+t_6} \dots \mathfrak{B}_{3t_{2n-1}+t_{2n}} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

As the \mathfrak{B}_0 transformation only involves scaling.

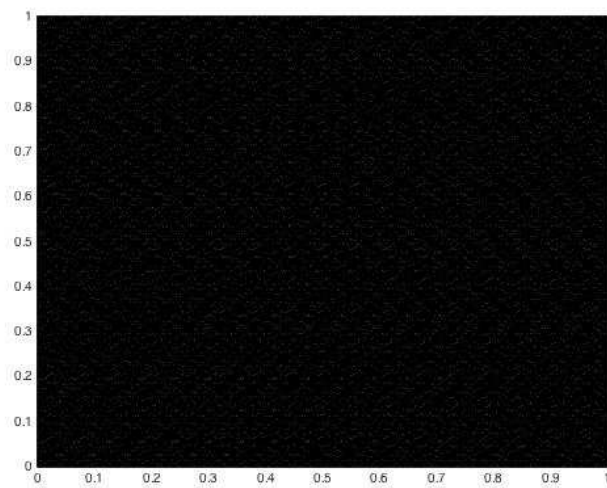
So we have found an alternative way for computing the image in the Peano curve of any point in \mathcal{I} in fact this method is what allows us to create the plots as seen throughout this text. Matlab does better than most programming language when it comes to the manipulation of complex numbers and so it is quite a trivial task to write a program which will produce such a plot as seen below.

```

z = 0.5+0.5i; %our initial point for transforming, the chosen value
%just insures that our line goes through the centre of each subsquare
iteration_level = 4; %which iteration we want to plot
for k = 1:iteration_level
    w = conj(z); %The conjugate value of z, for brevity in further code
    z = [z;
        -w+1+i;
        z+2i;
        w+3i+1;
        -z+2+2i;
        w+1+i;
        z+2;
        -w+3+i;
        z+2+2i]/3;
end
plot(z, 'clipping', 'off', 'color', 'black')
axis([0 1 0 1])
axis equal

```

This sort of program allows us very quickly to see the space filling nature of the Peano curve for example if we run this with our `iteration_level` variable set to a higher value for example 7 gives the result below.³



3 Hilbert's Space-Filling Curve

We now come to what is probably the most famous of the space filling curves⁴. It was Hilbert who first gave space filling curves formulated by a geometric recursion. This generating process can in fact be applied in any dimension , constructing a whole class of SFC's.

³The fact this plot seems like an almost black square demonstrates the space filling nature of the curve.

⁴This fact is probably due to its geometric definition allowing it to be easily visualised.

3.1 Definition

Following the thought process illustrated in sagan [1] we first consider the fact that the unit interval \mathcal{I} can be mapped continuously onto the unit square \mathcal{Q} , then by partitioning \mathcal{I} into 4 congruent subintervals it should be true that we can continuously map each subinterval into similarly partitioned subsquares of \mathcal{Q} . If we extend this process to the infinite recursion level then at each stage both \mathcal{I} and \mathcal{Q} are partitioned into 2^{2n} congruent duplicates, $n = 1, 2, 3, \dots$.

When performing a recursion like this ad infinitum we have to make sure that to preserve continuity adjacent subintervals are mapped to adjacent subsquares, it was Hilbert that showed such an arrangement of the subsquares is possible. We also require that inclusions remain through further iterations, that is if a square corresponds to some interval then its subsquares correspond to subintervals of that interval. The arrangement that Hilbert discovered also satisfies this property. The first 3 iterations of this ordering are shown below.

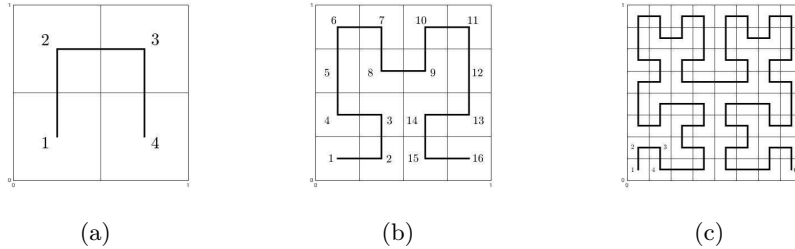


Figure 5: First 3 iterations of the Hilbert curve

We then give the formal definition of the Hilbert curve

Definition 6. [1, p. 11] Every $t \in \mathcal{I}$ is uniquely determined by a sequence of nested closed intervals (that are generated by our successive partitioning), the lengths of which shrink to 0. With this sequence corresponds a unique sequence of nested closed squares, the diagonals of which shrink into a point, and which define a unique point in \mathcal{Q} , the image $f_h(t)$ of t . We call $f_h(\mathcal{I})$ the Hilbert Curve.⁵

In the figure above we have joined the midpoints of each interval to show the convergence to the hilbert curve. From this definition we can now realise why it is important for adjacent intervals to be mapped to adjacent squares. Consider any point $t \in \mathcal{I}, t \neq 0, 1$ that is the endpoint of one of the subintervals, then t belongs to two sequences of nested intervals, so the adjacency of the squares these intervals are mapped to implies that the images will be the same.

3.2 Properties

Theorem 6. *The Hilbert Curve is surjective.*

Proof. Any point in \mathcal{Q} will lie in at least one sequence of nested closed subsquares, then each of these sequences corresponds to a sequence of nested subintervals which tend to some point in \mathcal{I} and so the curve is indeed surjective. \square

⁵I find this definition to be quite vague however was unable to find a more succinct version.

It is also quite trivial to confirm Netto's Theorem for the Hilbert Curve that it cannot be injective. Consider any point on the corner of a square in \mathcal{Q} then that point will belong to at least two squares (due to their closed nature) that do not correspond to adjacent intervals and so will have two or more distinct preimages in \mathcal{I} .

Finally to show the Hilbert curve does in fact fall in the class of an SFC we must show it to be continuous, to do this we can follow the same procedure as for the Peano curve.

Theorem 7. *The Hilbert Curve is continuous.*

Proof. I omit this proof for the sake of not repeating myself, as with the Peano curve it is an identical argument using Pythagoras to calculate the maximum distance between any two points in adjacent squares. It can be found in Sagan [1, p. 12]. \square

From these properties we can in fact say that the Hilbert Curve does represent an SFC. As with the Peano curve we can also show that the Hilbert curve isn't a differentiable function, and so $f_h(\mathcal{I}) = \mathcal{Q}$.

Theorem 8. *The Hilbert Curve is nowhere differentiable.*

Proof. For $n \geq 3$ and any $t \in \mathcal{I}$ we can choose a $t_n \in \mathcal{I}$ such that $|t - t_n| \leq 10/2^{2n}$ and that the coordinates of t and t_n under the image of the Hilbert curve are separated by at least a square of sidelength $1/2^n$ (This can be clearly seen to be always possible by inspecting the figure showing iterations of the Hilbert Curve above.) \square

3.3 Complex and Arithmetic Representation

⁶ This definition of the Hilbert curve has allowed us to consider all properties of the curve but doesn't give us an explicit computation to calculate the image of any point $t \in \mathcal{I}$. To do this we will need to do come up with a set of transformations which will allow us to compute such an image. It was the German mathematician W. Wunderlich who first came up with the idea of such a procedure with space filling curves⁷. We start as we did with the Peano curve and produce a complex representation, with \mathcal{Q} now denoting the complex unit square we first shrink it in the ratio 2 : 1 towards the origin, we rotate the resulting square 90° about the origin and then reflect in the imaginary axis, the composition of these transformations produces the subsquare $[0, 0.5] \times [0, 0.5]$ and will be named

$$\mathfrak{H}_0 z = \frac{1}{2} \bar{z} i$$

The subsquare $[0, 0.5] \times [0.5, 1]$ comes from the similar scaling of \mathcal{Q} shifted half a unit upwards. This time rotation and reflection is not necessary as the orientation of this subsquare is the same as that of \mathcal{Q} in the original ordering. The transformation which produces this is

$$\mathfrak{H}_1 z = \frac{1}{2} z + \frac{i}{2}$$

⁶§3.3 is again almost straight out of Sagan [1] with a few modifications to explain some of the ideas in greater detail

⁷Whilst Wunderlich had the idea it was Hilbert who formalised it with his curve and applied it to others

The top right subsquare $[0.5, 1] \times [0.5, 1]$ is produced similarly, no rotation necessary.

$$\mathfrak{H}_2 z = \frac{1}{2}z + \frac{i}{2} + \frac{1}{2}$$

The final lower right subsquare require a similar rotation to the first to preserve the necessary orientation of the entry and exit points and so we obtain the transformation

$$\mathfrak{H}_3 z = -\frac{1}{2}\bar{z} + \frac{i}{2} + 1$$

Then applying these transformations to Figure 5(a) will result in Figure 5(b) and application to Figure 5(b) will give Figure 5(c) etc. So if we consider a point $t \in \mathcal{I}$ in its quaternary form

$$t = 0_4 q_1 q_2 q_3 \dots, \quad q_j = 0, 1, 2, 3$$

Then we have the property that in the first iteration t will lie in the $(q_1 + 1)th$ subinterval of \mathcal{I} and so its image will lie in the $(q_1 + 1)th$ subsquare in the first partitioning of \mathcal{Q} , i.e $f(t) \in \mathfrak{H}_{q_1} \mathcal{Q}$. Then in the second iteration $f(t)$ will lie in the $(q_2 + 1)th$ subsquare of the $(q_1 + 1)th$ subsquare from the first iteration, so in terms of transformations $f(t) \in \mathfrak{H}_{q_1} \mathfrak{H}_{q_2} \mathcal{Q}$ (This is referred to as the localisation property). Continuing this sort of reasoning we can deduce that

$$f_h(t) = \begin{pmatrix} \mathcal{Re} \\ \mathcal{Im} \end{pmatrix} \lim_{n \rightarrow \infty} \mathfrak{H}_{q_1} \mathfrak{H}_{q_2} \mathfrak{H}_{q_3} \dots \mathfrak{H}_{q_n} \mathcal{Q}$$

This is progress but still doesn't allow us to find the image without having omniscient knowledge of the infinite expansion of a quaternary. However there are points in our curve where we have such knowledge. The points which lie at the beginning or end of any of the subintervals when partitioning \mathcal{I} will have a finite quaternary expansion. Consider such a t

$$t = 0_4 q_1 q_2 q_3 \dots q_n = 0_4 q_1 q_2 q_3 \dots q_n 000 \dots$$

and so from our previous formulation we have

$$f(t) = \begin{pmatrix} \mathcal{Re} \\ \mathcal{Im} \end{pmatrix} \mathfrak{H}_{q_1} \mathfrak{H}_{q_2} \mathfrak{H}_{q_3} \dots \mathfrak{H}_{q_k} \mathfrak{H}_0 \mathfrak{H}_0 \mathfrak{H}_0 \dots \mathcal{Q}$$

Then it seems natural to think that $\mathfrak{H}_0 \mathfrak{H}_0 \mathfrak{H}_0 \dots \mathcal{Q}$ would simply be 0 as we discussed above it is effectively a scale towards the origin along with reorientation, and indeed

$$\mathfrak{H}_0 \mathfrak{H}_0 \mathfrak{H}_0 \dots z = \lim_{n \rightarrow \infty} \mathfrak{H}^n = \lim_{n \rightarrow \infty} \frac{1}{2^n} \begin{cases} i\bar{z} & \text{if } n \text{ is even} \\ z & \text{if } n \text{ is odd} \end{cases} = 0$$

and so for such points we have

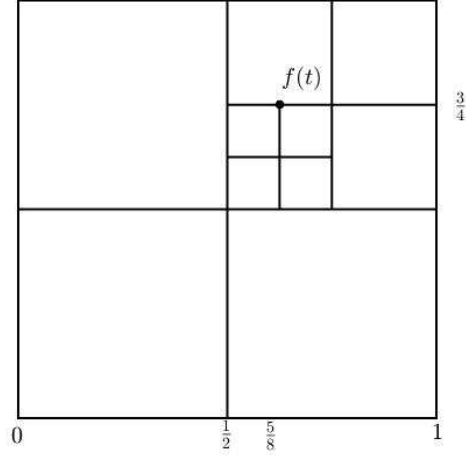
$$f(t) = \begin{pmatrix} \mathcal{Re} \\ \mathcal{Im} \end{pmatrix} \mathfrak{H}_{q_1} \mathfrak{H}_{q_2} \mathfrak{H}_{q_3} \dots \mathfrak{H}_{q_k} 0$$

It then becomes a trivial task to compute the images of any such points and indeed very close approximations to those points not at the endpoint of an interval through a programming language such as MATLAB. For example consider the point $t = 0_4 203 = 0.546875$

$$q_1 = 2, q_2 = 0, q_3 = 3$$

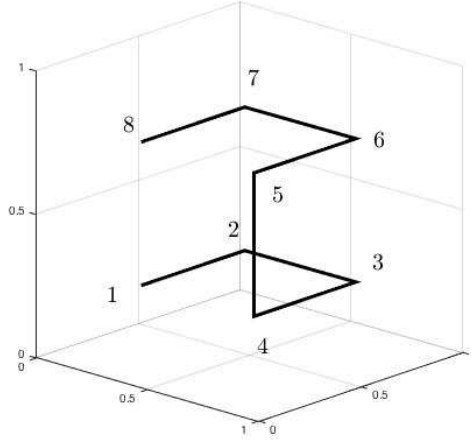
So we have the image

$$\begin{aligned}
f(0_4 203) &= \begin{pmatrix} \mathcal{R}e \\ \mathcal{I}m \end{pmatrix} \mathfrak{H}_2 \mathfrak{H}_0 \mathfrak{H}_3 0 \\
&= \begin{pmatrix} \mathcal{R}e \\ \mathcal{I}m \end{pmatrix} \mathfrak{H}_2 \mathfrak{H}_0 \left(\frac{i}{2} + 1 \right) \\
&= \begin{pmatrix} \mathcal{R}e \\ \mathcal{I}m \end{pmatrix} \mathfrak{H}_2 \left(\frac{1}{4} + \frac{i}{2} \right) \\
&= \begin{pmatrix} \mathcal{R}e \\ \mathcal{I}m \end{pmatrix} \left(\frac{1}{8} + \frac{i}{4} + \frac{i}{2} + \frac{1}{2} \right) \\
&= \begin{pmatrix} \mathcal{R}e \\ \mathcal{I}m \end{pmatrix} \left(\frac{5}{8} + \frac{3i}{4} \right)
\end{aligned}$$

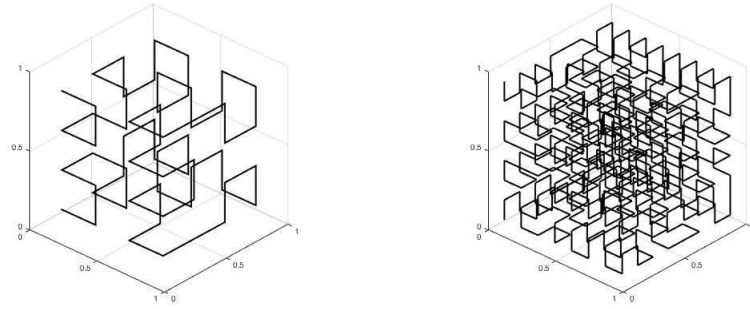


3.4 Generalisation to 3-Dimensions

The construction of the 3-D Hilbert curve is very similar to its 2-dimensional counterpart however instead of dividing the unit interval \mathcal{I} into 2^{2n} congruent subintervals we divide into 2^{3n} smaller intervals, and alongside this we divide \mathcal{W} the unit cube into 2^{3n} congruent subcubes. We then desire an arrangement of such subcubes similar to that of the subsquares in the original Hilbert curve. The figure below shows such an arrangement for the first iteration.



Then this arrangement described in Sagan [1] satisfies the condition we place of adjacent subintervals mapping to adjacent subcubes. Continuing as we did with the 2-D Hilbert curve iterating within nested intervals this defines a space filling curve. Below the 2^{nd} and 3^{rd} iterations of the 3-D Hilbert curve can be seen.



4 Lebesgue's Space-Filling Curve

We've now seen two examples of space filling curves with what have been very similar properties, both were defined in terms of an ordering of subdivisions in the unit interval, both were nowhere differentiable, both had a fractal nature. To contrast these we will now briefly discuss an example which is none of these things⁸. Proofs will be omitted as this example is just to illustrate the variation in SFC's⁹.

⁸Another interesting SFC is the Sierpinski curve which actually maps to the triangle with vertices $(0,0)$, $(2,0)$, $(1,1)$ instead of \mathcal{Q}

⁹The proofs of these can be found in Kupers [p. 3-4]kupers

4.1 Definition

Before we go on to define Lebesgue's curve we must first introduce a very famous example in mathematics, Cantor's set. In 1833 Cantor introduced as an example of a perfect set which is not dense on any interval the set of all points which can be written in the form

$$\frac{2t_1}{3} + \frac{2t_2}{3^2} + \frac{2t_3}{3^3} \dots$$

Using our knowledge of ternaries we notice this can be written as in the following definition

Definition 7 (Cantor set of excluded middle thirds). [1, p. 69]

$$\Gamma = \{0_3(2t_1)(2t_2)(2t_3) \dots : t_i = 0 \text{ or } 1\}$$

The reason for the name is how the set can be constructed. If you first start with the unit interval \mathcal{I} then remove the inner third i.e. the interval $(\frac{1}{3}, \frac{2}{3})$ you result in the two closed intervals $[0, \frac{1}{3}]$, $[\frac{2}{3}, 1]$. Repeating the process on these two new intervals removing the inner thirds, now the intervals $(\frac{1}{9}, \frac{2}{9})$, $(\frac{7}{9}, \frac{8}{9})$, leaves us with the set of closed intervals $[0, \frac{1}{9}]$, $[\frac{2}{9}, \frac{1}{3}]$, $[\frac{2}{3}, \frac{7}{9}]$, $[\frac{8}{9}, 1]$. Carrying this removing process on ad infinitum will then result in the set defined above. It can then be shown that the mapping

$$f(0_3(2t_1)(2t_2)(2t_3) \dots) = \begin{pmatrix} 0.t_1t_3t_5 \dots \\ 0.t_2t_4t_6 \dots \end{pmatrix}$$

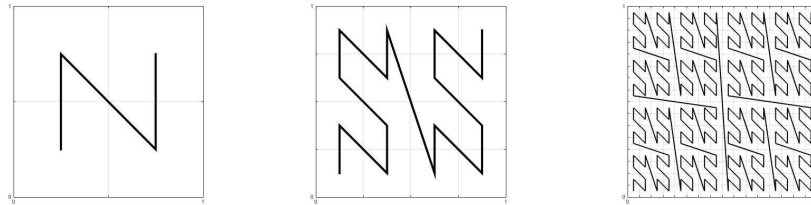
is a continuous, surjective mapping $f : \Gamma \rightarrow \mathcal{Q}$. Lebesgue's addition to this already studied map was its extension to $\Gamma^c(\mathcal{I} \setminus \Gamma)$ through linear interpolation. He did so as follows

Definition 8 (Lebesgue's Space filling Curve).

[8, p. 3] If (a_n, b_n) is an interval removed from \mathcal{I} in the n^{th} step in constructing Γ then define $f_l : \Gamma \cup \Gamma^c = \mathcal{I} \rightarrow \mathcal{Q}$

$$f_l(t) = \begin{pmatrix} f(t) & \text{if } t \in \Gamma \\ (f(b_n) - f(a_n)) \frac{t - a_n}{b_n - a_n} + f(a_n) & \text{if } t \in (a_n, b_n) \subset \Gamma^c \end{pmatrix}$$

Where f is the function defined above. The figure below shows respectively the 1st, 2nd and 4th iterations in generating the Lebesgue SFC.



4.2 Properties

As expected the function we just defined does in fact follow the definition of a space filling curve, i.e it is both continuous and surjective. The fun of the Lebesgue SFC comes when we consider differentiability. Whilst the other two curves studied in this paper have both been nowhere differentiable the Lebesgue SFC is almost everywhere differentiable, in fact everywhere on Γ^c it is differentiable and is nowhere differentiable on Γ .

Theorem 9. *Lebesgue's space filling curve is differentiable on Γ^c and is nowhere differentiable on Γ*

Seeing this fact poses the question "Is it possible to find a space filling curve differentiable everywhere?" the answer to this was given, only quite recently in 1980, by Polish mathematician Michal Morayne when he proved in fact that there are no differentiable space filling curves.

The other notable comparison that can be made between the Lebesgue SFC and the earlier examples is what's known as the fractal nature of the curves, that is every part of the earlier curves was itself a space filling curve. For example considering the Hilbert curve, if we take the interval $\pi = [(k-1)/2^{2n}, k/2^{2n}] \subset \mathcal{I}, n = 1, 2, 3, \dots, k = 1, 2, 3, \dots, 2^{2n}$ then applying the Hilbert curve $f_h : \pi \rightarrow \mathcal{Q}$ results in an exact copy of the Hilbert curve scaled in the ratio $2^n : 1$. This property can be seen to be lacking in the Lebesgue SFC simply by the amount in jumps across the unit square. The absence of this property changes some of the applications for which the Lebesgue SFC can be used.

References

- [1] *Space Filling Curves*, Hans Sagan, *Springer 1994* 1st edition.
- [2] *The Definition of a Manifold and First Examples*, Jenny Wilson, <http://web.stanford.edu/~jchw/WOMPtalk-Manifolds.pdf>
- [3] *Wikipedia*, Online encyclopedia, https://en.wikipedia.org/wiki/Differentiable_manifold
- [4] *Building Cantors Bijection*, Samuel Nicolay, Laurent Simons 2014, <http://arxiv.org/pdf/1409.1755.pdf>
- [5] *Space-Filling Curves: An Introduction with Applications in Scientific Computing*, Michael Bader, *Springer 2012*.
- [6] *Space-Filling curves an Introduction*, Levi Valgaerts, Department of Informatics, Technical University Munich
- [7] *MA131 Analysis II Lecture notes*, James Robinson, <http://www2.warwick.ac.uk/fac/sci/math/undergrad/ug handbook/lecturers/ma131t2/analysis>
- [8] *On Space-Filling Curves and the Hahn-Mazurikewicz theorem*, A.P.M Kupers, <http://web.stanford.edu/~kupers/spacefillingcurves.pdf>.