THE PEANO CURVE

JEFFERY CAVALLARO

Abstract. In the late 19^{th} century, the field of mathematics was on the verge of a major explosion. Most of the intuitive problems had already been solved, and mathematicians were starting to pose questions regarding certain non-intuitive problems that seemed to shake the mathematical orthodoxy at its core. The trio of Cantor, Peano, and Hilbert offered the mathematical world methods to wrestle with these problems, setting the foundations for the practice of mathematical logic that we take for granted today. One such discovery that epitomizes this movement was Peano's formulation of a "space-filling" curve, and Hilbert's method for producing such a curve. Starting with an intuitive isomorphism between the unit interval and the unit square, one can describe a sequence of functions that map curves in the unit square. Then, employing some straightforward (by today's standards) analysis, one arrives at the fairly non-intuitive result that the sequence of functions converges to one whose map is surjective on the unit square.

1. Historical Background

Up until the mid-19^{th} century, mathematics was still limited by the extreme existentialism of the Hellenistic world view that so enamored the West: a strong reliance on intuition and an open hostility to the concept of non-intuitive logic, especially problems involving infinity. But by that time, most of the "intuitive" problems had been solved based on earlier work by notables such as Euler and Gauss. As a result, mathematicians were starting to come up with a whole new class of problems whose solutions did not seem to be solvable using the past intuitive methods.

One popular reaction was to dismiss such problems, usually with a near-religious zeal, and probably not altogether disconnected from the nationalism and drive for empire that was sweeping across Europe at the time. Straight into this fray walked a Russian by the name of Georg Cantor (1845–1918), whose non-intuitive set theory and formal treatment of infinity were something completely new and unforeseen. As a result, Cantor was lambasted by the most vocal in the mathematics community and spent most of his later life in a sanatorium.

However, Cantor had a soulmate in the person of Guiseppe Peano (1858–1932), a soft-spoken Italian who attempted to formalize mathematical logic in education as part of his so-called *Formulario Project*. Today's analysts may recall that one of the first things they learned was the Peano axioms on the natural numbers. Peano was so well-liked by his students and peers that he didn't receive the same scorn as Cantor; however, his work was largely ignored until David Hilbert (1862–1943), a Prussian, started solving many of those pesky non-intuitive problems using Peano and Cantor's methods.

As the new methods in mathematical logic gained popularity, they were used by other budding mathematical giants such as Borel and Lebesque in France, and Riemann and Dedekind in Germany (Gauss's students) to make similar discoveries. In fact, it was found

Date: May 9, 2017.

Math 231A: Real Analysis I.

that many of the so-called non-intuitive results had concrete applications in the blossoming fields of electrical engineering and communication theory, born of recent developments in electricity transmission, the telegraph, and the laying of trans-Atlantic cables.

One of Peano's discoveries that so epitomizes this movement is his formulation in 1890 of a continuous, "space-filling" curve. Once formulated, it was then left to Hilbert to come up with a method for generating such a curve. It is this formulation and method that we will examine in this paper.

2. Overview

The goal is to find a curve $t \mapsto \mathcal{P}(t)$ with the following properties:

(1) $\mathcal{P} : [0,1] \to [0,1] \times [0,1]$.
(2) $\mathcal{P}$ is continuous.
(3) $\mathcal{P}$ is surjective.
(4) The image under $\mathcal{P}$ of any $[a,b] \subset [0,1]$ is a square $S_{[a,b]} \subset [0,1] \times [0,1]$ such that $m_2(S_{[a,b]}) = b - a$.

The gist of property 4 is demonstrated in Figure 1.

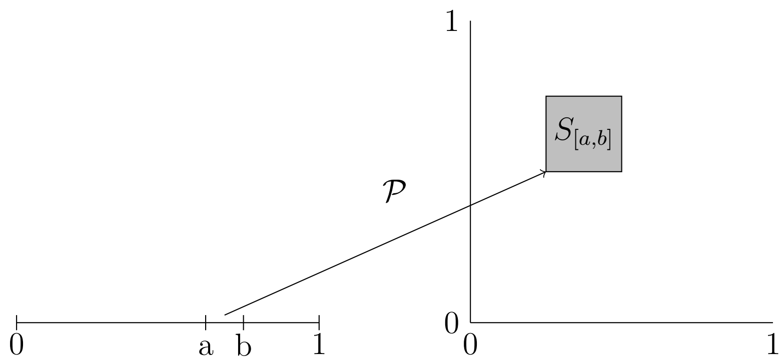


Figure 1. Demostration of Property 4

Such a mapping is called a *Peano Mapping* and its corresponding curve is called a *Peano Curve*. Note that this is a parameterized curve from domain $\mathbb{R}$ to codomain $\mathbb{R}^2$; it is not a mapping *on* $\mathbb{R}$ and thus is not subject to our normal intuitive notions of a curve in $\mathbb{R}^2$.

The steps in the formulation of $\mathcal{P}$ are very similar to the construction of a Cantor set and the terms in the Cantor-Lebesgue function. We start with a treatment of the domain that continually subdivides each interval into four equal parts, called *quartic intervals*. Thus, the n^{th} generation contains 4^n almost-disjoint intervals. The codomain receives a similar treatment, this time into four equal squares, called *dyadic squares*. Likewise, the n^{th} generation contains 4^n almost-disjoint squares.

The matching cardinality between a generation of quartic intervals and its corresponding generation of dyadic squares invites the definition of a one-to-one correspondence Φ, referred to as the *dyadic correspondence*. As we continually refine each quartic interval and dyadic square, we can see that each collapses to a single point. This gives rise to an *induced mapping* Φ^* between the points in the domain and codomain. Although this mapping is not well-defined on the boundaries, the locus of the boundaries has measure zero. By eliminating the boundaries, the induced mapping is well-defined, continuous, and surjective.

To generate the actual curve, we take each generation of dyadic squares and connect their centers with line segments. We will call the curve for the k^{th} generation $\mathcal{P}_k$. Thus, the $\mathcal{P}_k$ comprise a sequence of uniformly continuous curves that converge to a continuous curve that is surjective on the unit square.

Each of these points is explained in detail in the following sections.

3. Quartic Intervals

Start with the domain $[0, 1]$. For each generation, subdivide each interval into four equal parts. This is demonstrated in Figure 2.

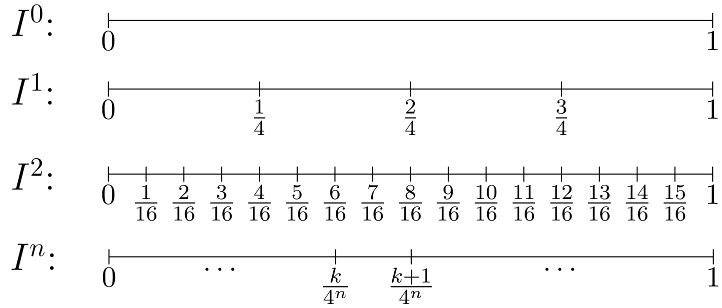


Figure 2. Generations of Quartic Intervals

Depending on context, the notation I^k represents either the entire k^{th} generation or a quartic interval from the k^{th} generation. Each generation has 4^n intervals of length (measure): $m_1(I^n) = \frac{1}{4^n}$.

We define a refinement of contained intervals as follows:

Definition 3.1. *A chain of quartic intervals is a decreasing sequence:*

$$I^0 \subset I^1 \subset I^2 \subset \ldots \subset I^k \subset \ldots$$

where I^k is a quartic interval of the k^{th} generation.

Properties 3.2. *Chains of quartic intervals have the following properties:*

1. *If (I^k) is a chain of quartic intervals then there exists a unique $t \in [0, 1]$ such that $t \in \bigcap_k I^k$.*
2. *Conversely, given $t \in [0, 1]$, there is a chain (I^k) of quartic intervals such that $t \in \bigcap_k I^k$.*
3. *The set of t for which the chain in (2) is not unique is a countable set of measure zero.*

Proof. Assume that Q is a chain of quartic intervals. Denote the interval in the k^{th} generation of Q by $[a_k, b_k]$. Then:

$$[0, 1] = [a_0, b_0] \supset [a_1, b_1] \supset [a_2, b_2] \supset \cdots$$

or, in general, $[a_k, b_k] \supset [a_{k+1}, b_{k+1}]$. So $a_k \leq a_{k+1} \leq b_{k+1} \leq b_k$ and in particular, $a_0 \leq a_k \leq b_k \leq b_0$. Thus (a_k) and (b_k) are both monotone, bounded sequences ($a_k \nearrow$ and $b_k \searrow$) and therefore they both converge. Let $a = \lim a_k = \sup(a_k)$ and $b = \lim b_k = \inf(b_k)$. Since $\forall k, a_k \leq b_k$, it is the case that $a \leq b$. Let $t = a$. So, $\forall k, a_k \leq t \leq b_k$ and:

$$t \in \bigcap_k [a_n, b_n] = \bigcap_k I^k$$

Therefore such a t exists.

Now by way of contradiction, assume that there are two such distinct t: t_1 and $t_2 \in Q$. Let $d = |t_1 - t_2|$. There exists a generation k such that length of one of the intervals in Q is $\frac{1}{4^k} < d$. Assuming that t_1 is in this interval, it is clear that t_2 is not, resulting in a contradiction. Therefore t exists and is unique, proving (1).

Conversely, given $t \in [0, 1]$, by construction, for each k there is at least one I^k such that $t \in I^k$, proving (2). In fact, non-uniqueness occurs at the boundaries of each interval at the points $\{\frac{j}{4^k} | 0 < j < 4^k\}$, which is referred to as the set of *dyadic rationals*. Since the set of dyadic rationals is a countable subset of $\mathbb{Q}$, it is also countable and has measure zero, thus proving (3). ☐

Similar to the Cantor-Lebesgue function, we can represent each chain of quartic intervals with a base-4 string of the form $0.a_1 a_2 \cdots a_k \cdots$ where each digit selects one of the four subintervals in an interval from the previous generation. Thus, each $t \in [0, 1]$ can be expressed as:

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

which is well-defined except at the dyadic rationals, which correspond to numbers of the form:

$$0.a_1 a_2 \cdots a_k 0000 \cdots = 0.a_1 a_2 \cdots (a_k - 1) 3333 \cdots$$

on the interval boundaries. This corresponds to the fact that each boundary point is contained in two different chains.

4. Dyadic Squares

The codomain gets a similar treatment. Start with the codomain $[0, 1] \times [0, 1]$. For each generation, subdivide each square into four equal parts. This is demonstrated in Figure 3.

Depending on context, the notation S^k represents either the entire k^{th} generation or a dyadic square from the k^{th} generation. Each generation has 4^n squares ($2^n \times 2^n$) of area (measure): $m_2(S^n) = \frac{1}{4^n}$ and sides of length $\frac{1}{2^n}$.

We define a refinement of contained intervals as follows:

Definition 4.1. *A* chain *of dyadic squares is a decreasing sequence:*

$$S^0 \subset S^1 \subset S^2 \subset \ldots \subset S^k \subset \ldots$$

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

Similar to quartic chains:

Properties 4.2. *Dyadic chains have the following properties:*

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

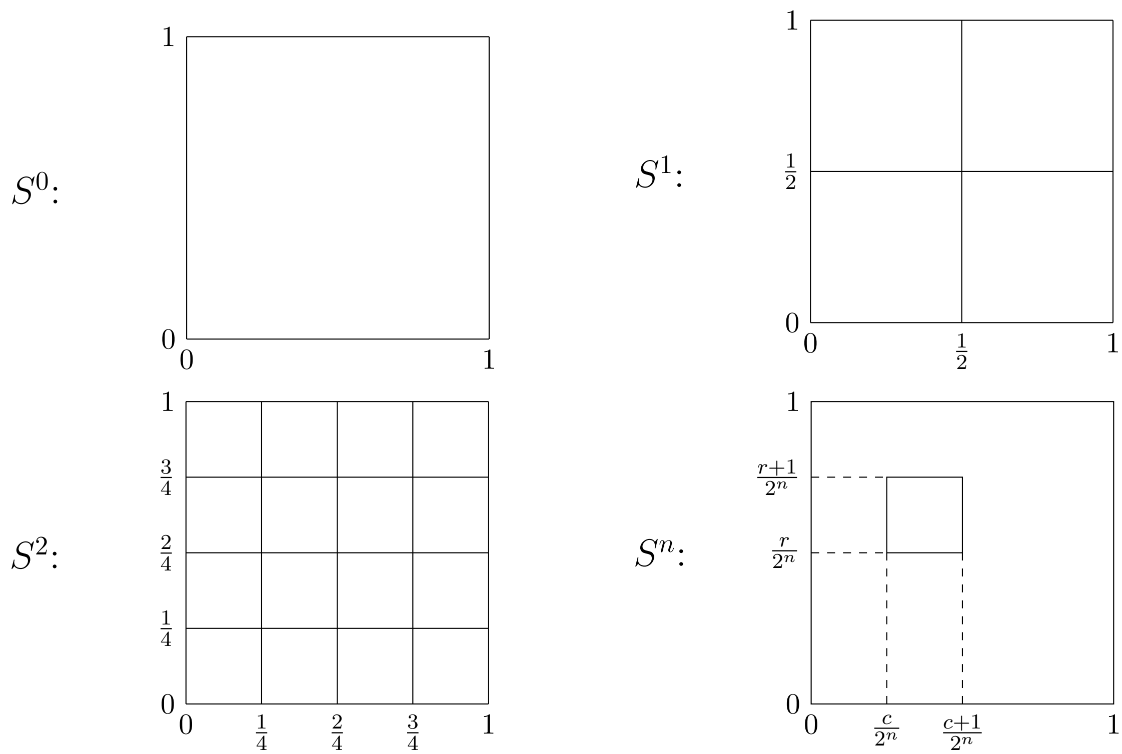


FIGURE 3. Generations of Dyadic Squares

(2) *Conversely, given* $x \in [0,1] \times [0,1]$, *there is a chain* (S^k) *of dyadic squares such that* $x \in \bigcap_k S^k$.
(3) *The set of* x *for which the chain in (2) is not unique is a countable set of measure zero.*

This time, the non-uniqueness occurs at all points (x_1, x_2), where at least one of the coordinates is a dyadic rational ($4^k = 2^{2k}$). The locus of these points compose the vertical and horizontal line segment boundaries. Since each vertical segment has measure zero, and since vertical segments occur at a subset of the dyadic rationals, the vertical segments comprise a set of measure zero. Similarly, the horizontal segments comprise a set of measure zero. Thus, their union is also a set of measure zero.

Similar to the Cantor-Lebesgue function, we can represent each chain with a base-4 string of the form $0.b_1 b_2 \cdots b_k \cdots$ where each digit selects one of the 4 sub-squares in a square from the previous generation. Thus, each $x \in [0,1] \times [0,1]$ can be expressed as:

$$x = \sum_{k=1}^{\infty} \frac{\overline{b_k}}{4^k} \tag{2}$$

where

$$\overline{b_k} = \begin{cases} (0,0) & b_k = 0 \\ (0,1) & b_k = 1 \\ (1,0) & b_k = 2 \\ (1,1) & b_k = 3 \end{cases}$$

which is well-defined except along the boundaries. Note that each non-corner is contained in two different chains and each corner is contained in four different chains.

5. Dyadic Correspondence

For each generation, $|I^n| = |S^n|$ and within a generation, $m_1(I^n) = m_2(S^n)$. So, it is possible and desirable to establish a bijection between the quartic intervals in the domain and the dyadic squares in the co-domain.

Definition 5.1. *A dyadic correspondence is a mapping Φ from collections of quartic intervals to collections of dyadic squares such that:*

1. Φ *is bijective.*
2. Φ *respects generations:* $I \in I^k \implies \Phi(I) \in S^k$.
3. Φ *respects inclusion:* $I \subset J \implies \Phi(I) \subset \Phi(J)$.

The trivial correspondence is given by $a_k = b_k$ in equations (1) and (2); however, this results in a discontinuous mapping of intervals to rectangles. A proper bijection ensures that adjacent intervals correspond to adjacent squares. The existence of such a bijection will be proved in section 6 during the construction of an actual curve.

Assuming that we have a proper bijection, we can then define the *induced mapping* $\Phi^* : [0,1] \to [0,1] \times [0,1]$.

Definition 5.2. *Let $\{t\} = \bigcap_{k=1}^{\infty} I^k$ where (I^k) is a chain of quartic intervals. Then $(\Phi(I^k))$ is a chain of dyadic squares such that $\bigcap_{k=1}^{\infty} \Phi(I^k) = \{x\}$ and:*

$$\Phi^*(t) = x \tag{3}$$

Note that if I is a quartic interval of the k^{th} generation then $\Phi^*[I] = \Phi(I)$, where $\Phi(I)$ is a dyadic square of the k^{th} generation and: $m_1(I) = m_2(\Phi^*[I])$.

The problem with Φ^* is that it is not well-defined on the quartic interval boundary points, since each is representable by two different quartic chains, resulting in two different dyadic chains. This is addressed in the following key theorem:

Theorem 5.3. *Given a dyadic correspondence Φ, there exists sets $Z_1 \subset [0,1]$ and $Z_2 \subset [0,1] \times [0,1]$, both of measure zero, such that:*

1. $\Phi^* : [0,1] \setminus Z_1 \to [0,1] \times [0,1] \setminus Z_2$ *is a bijection.*
2. $E \subset [0,1]$ *is measurable iff* $\Phi^*[E] \subset [0,1] \times [0,1]$ *is measurable.*
3. $m_1(E) = m_2(\Phi^*[E])$.

Before proving this key theorem, the following lemma for quartic intervals is needed:

Lemma 5.4. *Let (f_k) be a sequence of base-4 digits as in equation (1) and define:*

$$E = \left\{ t = \sum_{k=1}^{\infty} \frac{a^k}{4^k} \mid a_k \neq f_k, \text{ for all sufficiently-large } k \right\}$$

Then $m(E) = 0$.

Proof. Assume that (a_k) and (f_k) differ completely starting at digit r. Then the preceding $r-1$ digits of (a_k) select a particular quartic interval, call it E_0. Since $a_r \neq f_r$, one of the four possible sub-intervals in the next generation is eliminated. Call the remaining three intervals E_1. Continuing this process, we get the following sequence:

$$E_0 \supset E_1 \supset E_2 \supset \dots \searrow E$$

where $m(E_k) = \frac{3}{4}m(E_{k-1})$ and $m(E_0) < \infty$. Therefore:

$$m(E) = \lim m(E_0) = \lim \left(\frac{3}{4}\right)^n = 0$$

□

There is a similar statement for dyadic squares:

Lemma 5.5. *Let (f_k) be a sequence of base-4 digits as in equation (2) and define:*

$$E = \left\{ x = \sum_{k=1}^{\infty} \frac{\overline{b^k}}{4^k} \mid \overline{b_k} \neq f_k, \text{ for all sufficiently-large } k \right\}$$

Then $m(E) = 0$.

With the above lemma, we can now prove part (1) of theorem 5.3.

Proof. Let $\mathcal{N}_1$ be the collection of quartic chains for all points $t \in [0,1]$ that are not uniquely representable. Also, let $\mathcal{N}_2$ be the collection of dyadic chains for all points $x \in [0,1] \times [0,1]$ that are not uniquely representable. By viewing Φ more broadly as a bijection between collections of chains, Φ is a bijection between $\mathcal{N}_1 \cup \Phi^{-1}(\mathcal{N}_2)$ and $\Phi(\mathcal{N}_1) \cup \mathcal{N}_2$, as well as between their complements. By letting Z_1 be the locus of points in $\mathcal{N}_1 \cup \Phi^{-1}(\mathcal{N}_2)$ realizable by equation (1), and Z_2 be the locus of points in $\Phi(\mathcal{N}_1) \cup \mathcal{N}_2$ realizable by equation (2), then the constraint of Φ^* becomes a well-defined bijection from $[0,1] \setminus Z_1$ to $[0,1] \times [0,1] \setminus Z_2$.

To show that Z_1 has measure zero, we use lemma 5.4 to select appropriate (f_k) such that for all chains in $\mathcal{N}_1$ and $\Phi^{-1}(\mathcal{N}_2)$, $a_k \neq f_k$. For example, since each such chain occurs in $\mathcal{N}_1$ with trailing digit sequences of all 0's and all 3's, choose (f_k) to be all 1's. A similar argument using lemma 5.5 shows that Z_2 also has measure zero. □

Now, with a well-defined version of the bijection Φ^*, parts (2) and (3) of theorem 5.3 can be proved as follows:

Proof. Assume that $E \subset [0,1] \setminus Z_1$ is a subset of measure zero. Assume $\epsilon > 0$. E can be covered with a countable set of quartic intervals: $E = \bigcup_k I_k$ such that $\sum_k m(I_k) < \epsilon$. Now, since $\Phi^*[E] \subset \bigcup_k \Phi^*[I_j]$ and because of the fact that $m_1(I^k) = m_2(S^k)$:

$$m_2(\Phi^*[E]) \leq \sum_k m_2(\Phi^*[I_k]) = \sum_k m_2(S_k) = \sum_k m_1(I_k) < \epsilon$$

Thus, $\Phi^*(E)$ is measurable and has measure zero. Similarly, $(\Phi^*)^{-1}$ maps sets of zero measure in $[0,1] \times [0,1]$ to sets of measure zero in $[0,1]$. Therefore, theorem 5.3 holds for sets of measure zero.

Recall that any open subset of $[0,1]$ can be written as a countable union of closed intervals with disjoint interiors. The quartic intervals are such closed intervals, and so any open set in $[0,1]$ can be written as a countable union of quartic intervals. Similarly, any open set in $[0,1] \times [0,1]$ can be written as a countable union of dyadic squares. Thus, any open set

$\mathcal{O} \subset [0,1]$ can be written as a countable union of quartic intervals: $\mathcal{O} = \bigcup_k I_k$, the set $\mathcal{O} \setminus Z_1$ is measurable, and

$$m_2(\Phi^*[\mathcal{O} \setminus Z_1]) = \sum_k m_2(\Phi^*[I_k]) = \sum_k m_2(S_k) = \sum_k m_1(I_k) = m_1(\mathcal{O} \setminus Z_1)$$

Similarly, $(\Phi^*)^{-1}$ maps open sets in $[0,1] \times [0,1]$ to open sets in $[0,1]$. Therefore, theorem 5.3 holds for G_δ sets.

Finally, since every measurable set differs from some G_δ set by a set of measure zero, theorem 5.3 holds for all measurable sets E. $\square$

6. Dyadic Correspondence with Adjacency

With the Peano mapping properly defined, it is time to turn to Hilbert for the construction of a Peano curve. Before doing so, we must prove the existence of a dyadic correspondence that preserves adjacency. To do so, we must first characterize the possible traversals through the dyadic squares of a given generation.

Given a dyadic square S in generation k, consider its four sub-squares in generation $k+1$. Assuming we can enter at any square S_1, we can always traverse adjacent squares in a sequence (S_1, S_2, S_3, S_4) and exit at any edge σ. The four possibilities are shown in Figure 4.

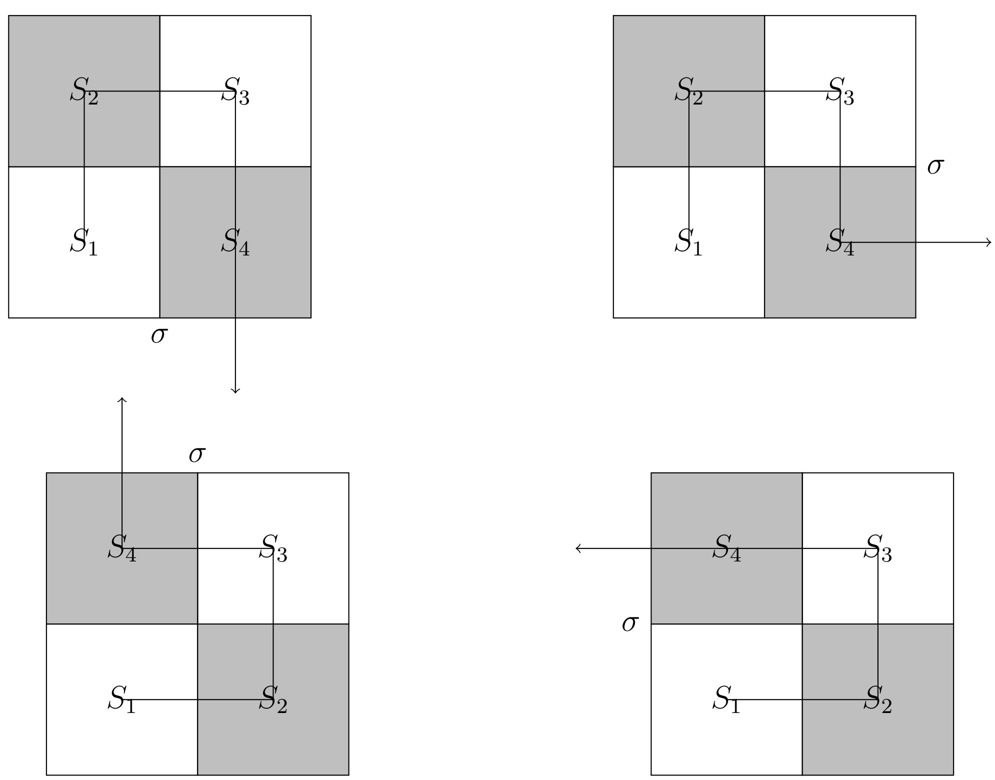


Figure 4. Traversal Through a Dyadic Square

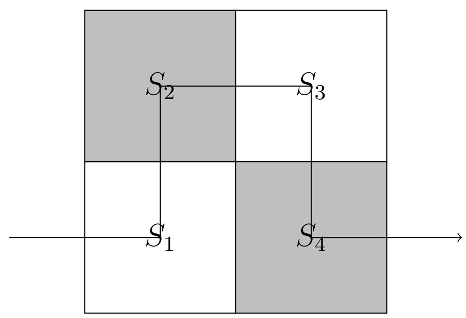
Note that each entry square is colored white and each exit square is colored black. With these possible traversals in hand, we are ready for the next theorem.

Theorem 6.1. *There exists a dyadic correspondence Φ such that:*

(1) *In generation k, Let I_- be the leftmost interval and I_+ be the rightmost interval. $\Phi(I_-)$ is the lower left square and $\Phi(I_+)$ is the lower right square.*
(2) *If I and J are two adjacent intervals in generation k then $\Phi(I)$ and $\Phi(J)$ are adjacent squares in generation k.*

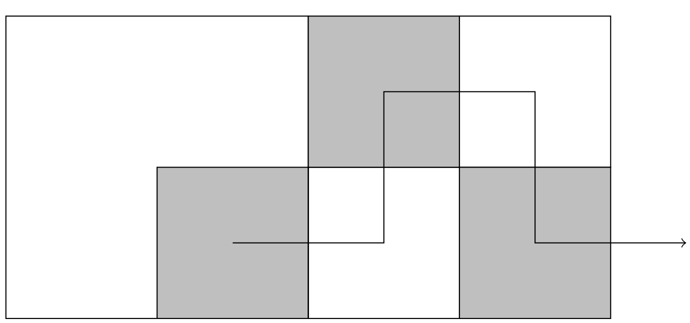
Proof. For condition (1), note that any generation of dyadic squares can be checkerboarded such that the lower-left square is white (an entry square) and the lower-right square is black (an exit square). By definitions, we force Φ to map the first interval to the lower-left (white) square and the last interval to the lower-right (black) square.

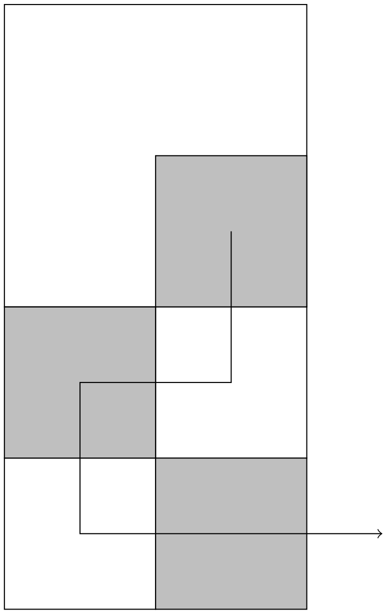
Condition (2) is proved by induction on generation number k. For the base case let $k = 1$:



Assume Φ has been defined for all generations $\leq k$ and consider generation $k+1$. Assume that S_j is a square from generation k that has been divided into squares $S_{j,1}, S_{j,2}, S_{j,3}, S_{j,4}$ in generation $k+1$. Square S_j is entered from adjacent square S_{j-1} by a traversal from a black square in S_{j-1} to a white square in S_j. Let σ be the edge between S_j and S_{j+1}. Traverse S_j using one of the four valid traversals to reach S_{j+1}.

By the inductive assumption, S_{4^k} is in the lower-right corner, and must be entered from adjacent square S_{4^k-1} from either the top or left. Since the lower-left square in generation $k+1$ is an exit square, it must be black. Since we must enter S_{4^k} on a white square, only one of the following two cases is possible:





Therefore, Φ is properly defined for generation $k+1$. $\square$

7. The Curve

With a proper dyadic correspondence Φ that preserves adjacency, we are now ready to construct the curve. For generation k, let t_j be the center of the j^{th} quartic interval:

$$t_j = \frac{j - \frac{1}{2}}{4^k}, 1 \leq j \leq 4^k$$

and let x_j be the center of the j^{th} dyadic square per the ordering imposed by Φ. Define:

$$\mathcal{P}(t) = \begin{cases} x_j & t = t_j \\ (0, \frac{1}{2^{k+1}}) = x_0 & t = t_0 = 0 \\ (1, \frac{1}{2^{k+1}}) = x_{4^k+1} & t = t_{4^k+1} = 1 \end{cases}$$

For each sub-interval $[t_j, t_{j+1}], 0 \leq j \leq 1$, define a linear mapping to the vertical or horizontal line segment joining the corresponding x_j and x_{j+1}. An example construction for $\mathcal{P}_2$ is shown in Figure 5.

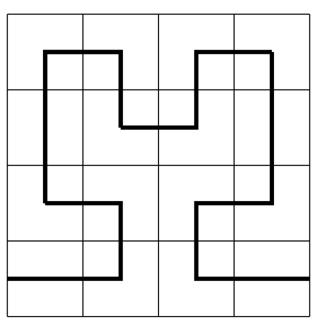


Figure 5. A Construction for $\mathcal{P}_2$

Theorem 7.1. *$P_k(t)$ is continuous.*

Proof. Assume $\epsilon > 0$ and let $\delta = \frac{\epsilon}{2^k}$. Note that the worst-case $|\mathcal{P}(t_1) - \mathcal{P}(t_2)|$ occurs along adjacent all vertical or all horizontal line segments. In such a case, $P_k'(t) = \frac{\frac{1}{2^k}}{\frac{1}{4^k}} = 2^k$, and thus $|\mathcal{P}(t_1) - \mathcal{P}(t_2)| \leq 2^k |t_1 - t_2| < 2^k \delta = 2^k \left(\frac{\epsilon}{2^k}\right) = \epsilon$ □

Theorem 7.2. *$\mathcal{P}$ exists, is continuous, and surjective.*

Proof. Note that since $\mathcal{P}_{k+1}(t)$ and $\mathcal{P}_k(t)$ are in the same dyadic square in generation k:

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

and thus the limit:

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

exists. Furthermore, $\mathcal{P}$ is continuous since the $\mathcal{P}_k$ are uniformly continuous. Moveover, since $\mathcal{P}_k$ visits every dyadic square in generation k, $\mathcal{P}$ is dense in $[0, 1] \times [0, 1]$. Since $\mathcal{P}$ is both continuous and dense, it is also surjective. □

The final step is to prove that $\mathcal{P} = \Phi^*$.

Theorem 7.3. $\Phi^*(t) = \mathcal{P}(t)$.

Proof. Assume $t \in [0, 1]$.

$$\Phi^*(t) = \Phi^* \left(\bigcap_k I_k \right) = \bigcap_k \Phi(I_k) = \bigcap_k S_k = x = \mathcal{P}(t)$$

□

References

1. Elias Stein and Rami Shakarchi, *Real analysis: Measure theory, integration, and hilbert spaces*, Princeton University Press, 2005.
2. Wikipedia, *David Hilbert — Wikipedia, the Free Encyclopedia*, 2016, [Online; accessed 27-May-2016].
3. ______, *Georg Cantor — Wikipedia, the Free Encyclopedia*, 2016, [Online; accessed 27-May-2016].
4. ______, *Guiseppe Peano — Wikipedia, the Free Encyclopedia*, 2016, [Online; accessed 27-May-2016].

Department of Mathematics, San José State University, San José, CA 95192-0103
E-mail address: jeffery.cavallaro@sjsu.edu