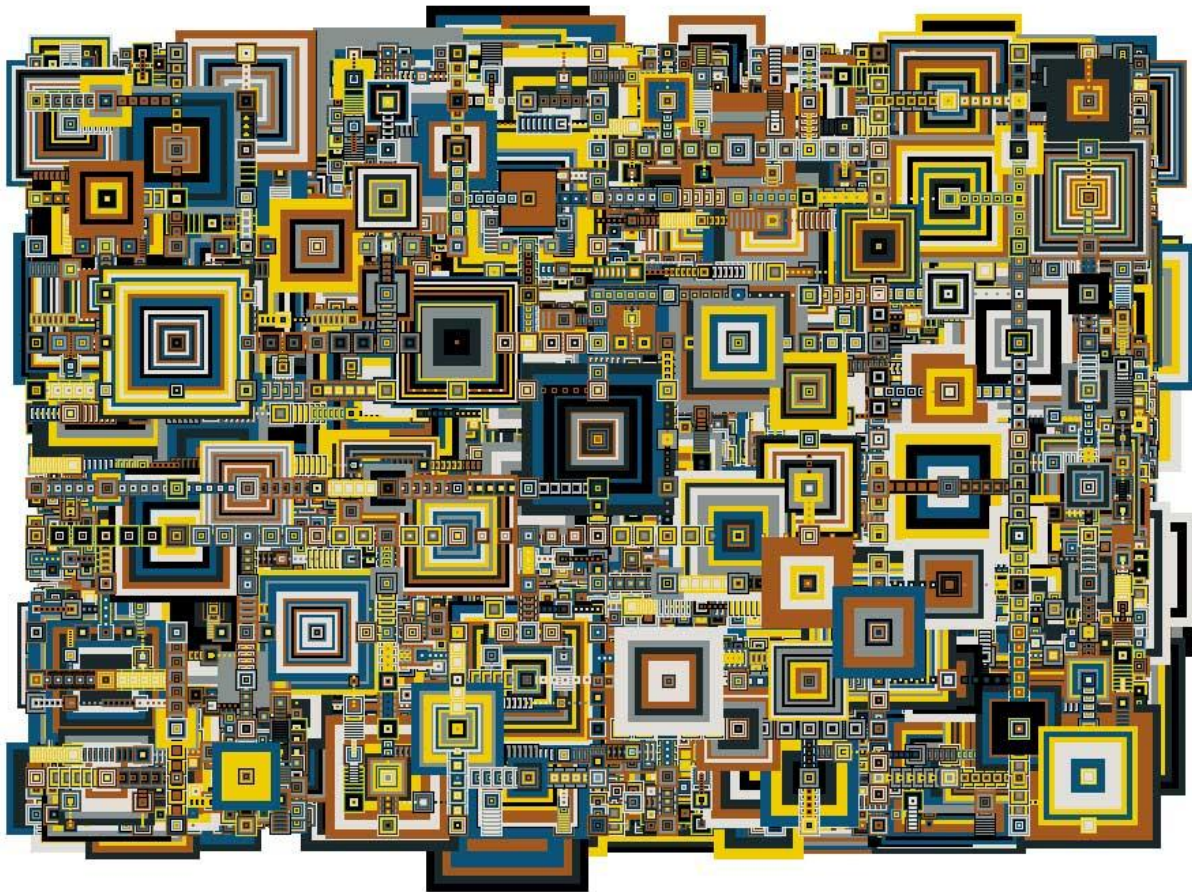


Space-Filling Curves

Frieda Parker ■ Math 635 ■ Fall 2008



In 1878, George Cantor published a remarkable finding: there exists a bijective function between any two finite-dimensional smooth manifolds. For example, it was possible to map the points in the unit interval $[0,1]$ onto the unit square $[0,1] \times [0,1]$. As for most people, this was not intuitive for Cantor. He spent three years trying to disprove this idea before attempting to prove it true. Even upon finding a proof he wrote in a letter to Dedekind, “I see it, but I don’t believe it.” Remarkably, the idea underlying the proof is rather simple. Assume you want to find a mapping between a coordinate pair (x,y) and a real number z . Represent x and y in their decimal forms: $x = 0.abcd\dots$ and $y = 0.ABCD\dots$. The mapping is accomplished by setting $z = 0.aAbBcCdD\dots$

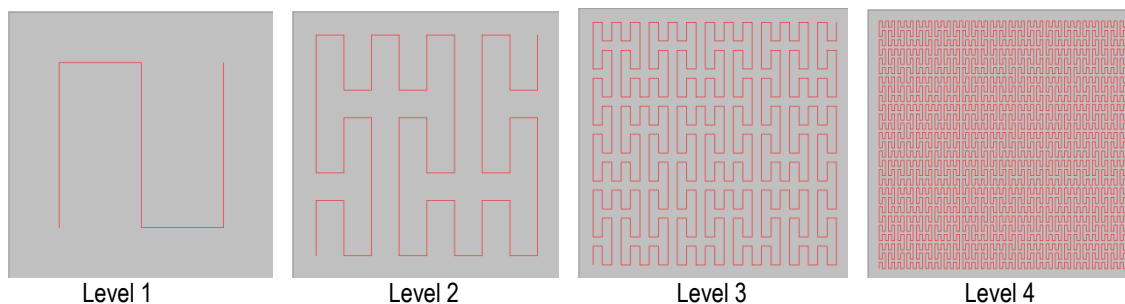
A question that immediately arose in response to Cantor’s result was whether or not such a bijection was continuous. In 1879, E. Netto proved it could not be. The outline of his proof is as follows.

- Assume there exists a continuous bijection f that maps an interval in the real numbers I onto an n -dimensional space (with n greater than 1).
- Show f^{-1} is continuous.
- Invoke the theorem that if f is continuous, then it maps connected sets to connected sets.
- Show f^{-1} does not do this by removing a point x_o from both the domain and its image $f(x_o)$ from the range. The set remaining in the domain is no longer compact, while the set in the range remains compact.
- Conclude f cannot be continuous.

After Netto’s result, some mathematicians began to look for continuous surjective mappings. In 1890, Peano found one. The Peano curve maps the unit interval $[0,1]$ into the plane with the

image having positive Jordan content. This curve has been called the first space-filling curve (SFC). Other SFCs soon followed with Hilbert's in 1891, Moore's in 1900, Lebesgue's in 1904, Sierpinski's in 1912, and Polya's in 1913. Despite the creation of these other curves, sometimes space-filling curves are still referred to as Peano curves.

A space-filling curve can be thought of as a map from one-dimensional space onto a higher-dimensional space. When working in Euclidean space, it is often thought of as a continuous map from the unit interval $[0,1]$ to \mathbf{R}^d ($d > 1$) whose image has positive Jordan content. Such curves are usually the limit of approximating curves where each approximating curve is bounded, but the length of these curves continues indefinitely. Because the mapping cannot be both continuous and bijective, the final curve is always self-intersecting, even if the approximating curves are not. Below are the first four approximating curves for the Peano curve.

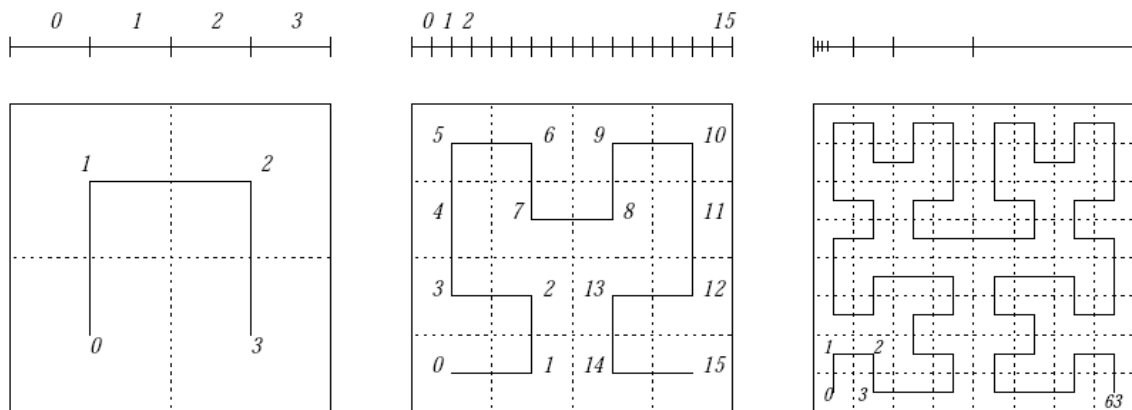


After exploring continuous surjective maps, mathematicians looked into continuous injective maps. In 1903, W. F. Osgood found a continuous injective map from the unit interval $[0,1]$ into \mathbf{R}^n ($n > 1$) with positive Lebesgue measure whose limiting arc is the Peano curve. Similarly, K. Knopp in 1917 found a map whose limiting arc is the Sierpinski curve.

The next logical question was what is the general structure of a continuous image of a line segment? In 1913, Mazurkiewica and H. Hahn found that a set is the continuous image of a

line segment if and only if it is compact, connected, and locally connected. This applies to both n -dimensional Euclidean spaces and more generally to Hausdorff spaces. This finding is now known as the Mazurkiewicz-Hahn Theorem.

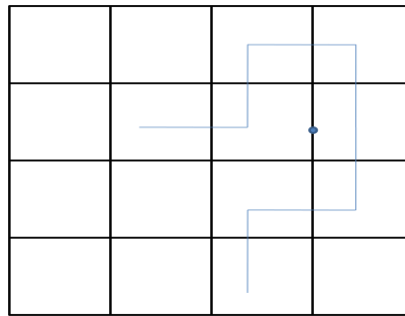
Hilbert's curve, while not being the first SFC, is a significant SFC because Hilbert's method led to a class of SFCs. Also, the Hilbert curve has been used in real-world applications. The Hilbert curve is constructed by first dividing the unit interval into four equal segments. For a 2-dimensional version of the Hilbert curve, the image space is the unit square. Similarly, the unit square is subdivided into four equal squares. The points that mark the subdivisions of the unit interval are each placed into one sub-square and then connected so that consecutive points in the interval lie in consecutive sub-squares. This process is repeated for the next level. Each sub-interval is divided into four equal intervals and each sub-square is divided into four equal squares and a similar mapping of points into the squares occurs. Below is a visual of this construction.



This construction shows that every point t in the unit interval is uniquely determined by a sequence of nested closed intervals. This sequence corresponds to a unique sequence of nested closed squares. The fact that this curve is not one-to-one can be visualized from the image below.

The point in the image lies on two different sub-squares. Therefore, this point is part of two sequences of nested sub-squares and so has more than one pre-image in the unit interval.

Therefore, the curve will intersect at this point.

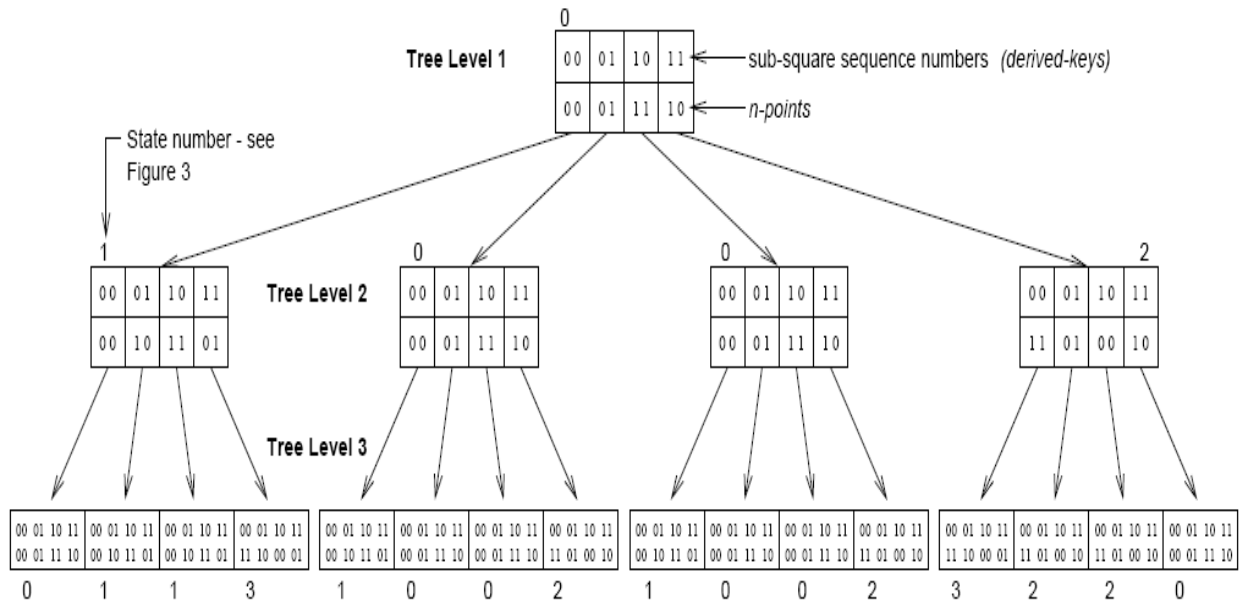


Outlines of proofs that the Hilbert curve is continuous, but nowhere differentiable are included in the Appendix.

The Hilbert curve, and other space-filling curves, has been used in a variety of applications. The three I will discuss are multi-dimensional indexing in computer database management systems, geographic research, and art. Database management systems (DBMS) are the software and firmware that manage the storage, querying, and retrieval of data files. Initially, DBMS were one-dimensional. That is, every file has a primary key that uniquely identifies each record in the file and this key is used to place the records in physical storage. In a multi-dimensional DBMS, primary keys may be composed of more than one element and any element may be used for retrieval. The entire key set must represent a unique identifier for the record, but individual elements in the key do not have to be unique for the record. For example, a student file may have a primary key composed of student ID, last name, and home state. The last name of the student may not be unique, but a multi-dimensional DBMS would allow retrieval of records by last name. DBMS developers have examined ways to construct multi-dimensional

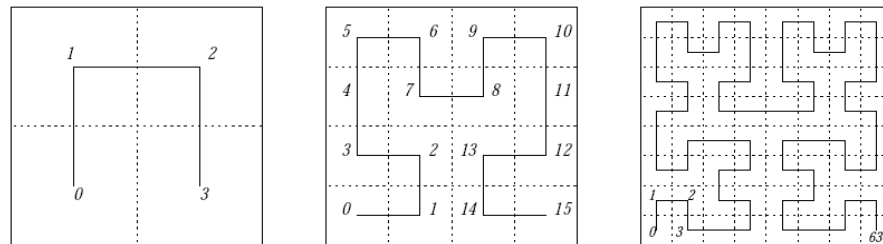
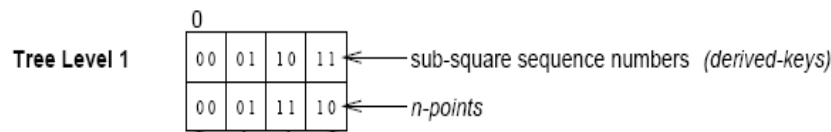
indexing that would like on top of the one-dimensional retrieval system. That is, the multi-dimensional index takes the n-dimensional key spaces and maps them onto a one-dimensional space and then the one-dimensional retrieval system is activated. Space-filling curves can be employed to accomplish this mapping.

The details of such a mapping, including the necessary algorithms, are too extensive to explicate in this paper, but I will provide an overview of the process. First, the multiple elements of the primary key set are mapped into an n-dimensional space. Then a space-filling curve is used to establish a mapping between the n-dimensional coordinates and the real numbers. An essential element of this mapping is how to store it. For the Hilbert curve, a tree structure is created. The structure for a 2-dimensional curve is diagrammed below.



In tree level 1, the top row is the real number in the unit interval. The bottom row contains the corresponding coordinates of the sub-square in the first approximating curve. For example, the number 2 (01_2) corresponds to the sub-square in the upper right-hand corner (1,1). To find points

in each sub-square, the next level is retrieved from the tree. A nice property of the Hilbert curve is that there are only four ways in which any four points are connected between the sub-squares. For DBMS purposes, these are called four *states*. Using the tree structure established above, the middle two columns always correspond to columns of the same state in the level below them.



For example, the first level (consisting of just the point 0, 1, 2, and 3) has the state with the side open at the bottom. The 0- and 3-sub-squares in level 1 correspond to states that are different in level 2 (the middle square above). However, the 1- and 2-sub-squares in level 1 correspond to sub-squares in level 2 that have the same state, open at the bottom. This pattern in the Hilbert curve allows for more efficient storage of the tree elements. This use of the Hilbert curve in multi-dimensional indexing was patented in 2007. See Lawder and King for more information on this topic.

Other computer scientists have explored the properties of many SFCs. They have analyzed these curves by classifying each segment according to its behavior. Classifications include Jump, Contiguity, Reverse, Forward, and Still. It turns out that there are identifiable relationships between the types of segments in each SFC. This analysis has allowed computer

scientists to evaluate the applicability of different SFCs to different applications. See Mokbel, Aref, and Kamel for more information.

The process of partitioning space into regions is a central theme in geographic research. Often the goal of such partitioning is to optimize a function. For example, there may be a need to determine the placement of several health centers in a region in such a way as to minimize the distance a person who lives in the region has to travel to a health center. Traditionally, the location-allocation method has been employed to tackle such problems. However, this method is computationally complex and is sometimes prohibitive in projects with a short lead time. The space-order method, which involves the use of a space-filling curve, provides a simpler way of partitioning space.

The idea behind this method is straight-forward. First, identify centroids for sub-spaces within a region. For example, for larger regions, counties may be sub-spaces and centroids established for each county. Then use a space-filling curve to link these centroids. Therefore, the sub-spaces (counties) are now sequenced linearly. Close values represent centroids that are close geographically and distant values represent centroids that are distant geographically. However, it is possible for centroids to be close geographically, but have values that are not close. In the case of wanting to place HIV/AIDS treatment centers so that each center has to handle approximately the same number of cases, a case load could be attached to each centroid. Then as the sequence of centroids is traversed, the case load can be accumulated until some threshold is met. The centroids traversed before the threshold become a partition. The case load count is reset and the process is repeated. See Siu-Ngan and Liu for more information.

Finally, at least one artist has made creative use of the Hilbert curve. Don Relyea has written computer programs to generate pictures derived from the Hilbert curve. His web site contains a page about his work in this area (http://www.donrelyea.com/hilbert_algorithmic_art_menu.htm). It also contains a Flash program that takes user parameters, such as color palette, and generates pictures. The cover of this report is an example of Relyea's space-filling curve artwork.

Space-filling curves represent that class of theoretical mathematics that not only offers profound and interesting mathematical results, but that has multiple real-world applications as well. Perhaps the mathematical and application extensions of space-filling curves will continue to evolve.

Appendix

The Hilbert curve is continuous:

- In n^{th} interval, I is partitioned into 2^{2n} intervals of length $1/2^{2n}$ each.
- So, $|t_1 - t_2| < 1/2^{2n}$
- Images lie, at worst, in two consecutive squares with side length $1/2^n$, so can be at most $\sqrt{5}/2^n$ apart.
- So, $|f_h(t_1) - f_h(t_2)| \leq \sqrt{5}/2^n$ and f_h is continuous.

The Hilbert curve is nowhere differentiable:

- For $n > 3$, pick a t and t_n from I such that $|t - t_n| \leq 10/2^{2n}$ and the coordinates of the images of t and t_n are separated by at least a square side length of $1/2^n$.
- This is always possible.
- So, if φ_h is one coordinate function, then

$$\frac{|\varphi_h(t) - \varphi_h(t_n)|}{|t - t_n|} \geq 2^n/10$$

- The same argument applies to all coordinate functions.

References

- Buchin, K. (2007). *Organizing point sets: Space-filling curves, Delaunay Tessellations of random point sets, and flow complexes*. Retrieved October 20, 2008, from Institut für Informatik Freie Universität Berlin Web site: http://www.diss.fu-berlin.de/diss/servlets/MCRFileNodeServlet/FUDISS_derivate_000000003494/0_buchin.pdf?hosts
- Lawder, J. K., & King, P. J. H. (2000). Using space-filling curves for multi-dimensional indexing. In B. Lings & K. Jeffery, (Eds.), *Proceedings of the 17th British National Conference on Databases (BNCOD 17)*, Vol. 1832 of *Lecture Notes in Computer Science* (pp. 20–35). Berlin: Springer Verlag. Retrieved on October 20, 2008, from [http://wotan.liu.edu/docis/lib/sisl/rclis/dbl/bncodb/\(2000\)%253C20%253AUSCFMI%253E/www.dcs.bbk.ac.uk%252FTriStarp%252Fpubs%252Fbncod17.pdf](http://wotan.liu.edu/docis/lib/sisl/rclis/dbl/bncodb/(2000)%253C20%253AUSCFMI%253E/www.dcs.bbk.ac.uk%252FTriStarp%252Fpubs%252Fbncod17.pdf)
- Mokbel, M. F., & Aref, W. G. (2003). Analysis of multi-dimensional space-filling curves. *GeoInformatica*, 7(3), 179-209.
- Relyea, Don website: www.donrelyea.com
- Sagan, H. (1994). *Space-filling curves*. New York: Springer-Verlag.
- Siu-Ngan, N., & Liu, K. (1996). Use of space-filling curves in generating a national rural sampling frame for HIV/AIDS research. *Professional Geographer*, 48(3), 321-332.
- Space-filling curve. *Wikipedia*. Retrieved November, 8, 2008, from http://en.wikipedia.org/wiki/Space-filling_curve
- Space-filling curve modeling: <http://www.cs.utexas.edu/users/vbb/misc/sfc/Oindex.html>