Introducing the Cantor Set

Gary M. Lewis*

(Dated: November 17, 2015; last revision: April 11, 2016)

Two things happening here. One, I wanted to better understand the Cantor set. And, two, I'm trying to learn LaTeX. So I combined the two activities into this document. Consider it an exploration of both the Cantor set and LaTeX. Or, better yet, consider it as maxing out my brain.

Keywords: Cantor set; fractals; chaos.

Georg Cantor (1845–1918) was a German mathematician who polarized the mathematics world near the end of the nineteenth century when he introduced various notions of infinity into mathematics. Detractors included Henri Poincaré, who considered Cantor's ideas a "mathematical malady ... that would one day be cured," while David Hilbert believed Cantor had created a "new paradise for mathematicians" [1]. Not much room there to reach common ground. And Poincaré and Hilbert were only two of the luminaries lined up on either side.

My task here is much less controversial. I merely want to better understand a concoction associated with Georg Cantor's name, even though it was first introduced independently by two other mathematicians several years before Cantor gave it his imprimatur. It's an amazing set of numbers between 0 and 1, now known as Cantor's ternary set, that challenges our understandings of infinite, infinitesimal, continuous, and discrete.

Here I intend only to explore the Cantor set and to identify it's principal features.

Geometric Construction

The Cantor set can be created using the following recursive operations on the unit interval [0,1]:

1. Start with the unit interval.
2. Remove the middle one-third from the unit interval, but not the end points created at 1/3 and 2/3. Two intervals remain: [0,1/3] and [2/3,1].
3. Remove the middle one-third from each of the remaining intervals, but, again, not the new end points. This leaves four intervals: [0,1/9], [2/9,1/3], [2/3,7/9], and [8/9,1].
4. Continue in this iterated manner ad infinitum.
5. The remaining points constitute the Cantor set.

Initially the points 0 and 1 are known members of the Cantor set. Removing the first middle-third adds the points at 1/3 and 2/3. Removing middle-thirds at the next iteration identifies 4 more points in the Cantor set: 1/9, 2/9, 7/9, and 8/9. And so on, throughout the infinite recursion.

It would seem that the Cantor set consists only of end points. But this turns out to be incorrect. End points are only the tip of an infinitely large iceberg. Before getting to this, however, let's look more closely at end points.

End Points

Already we can see several characteristics that identify end points in the Cantor set. For example, all new end points after the initial 0 and 1 are rational numbers (ie, fractions). But they're not just any fractions; there is a definite pattern. At iteration 1, we get 1/3 and 2/3. At iteration 2, we get 1/9, 2/9, 7/9, and 8/9. Because the recursive operation involves removing the middle-third, all end points have denominators that are a power of 3. So, $3^1 = 3$ is the denominator of end points created at iteration 1, and $3^2 = 9$ is the denominator of end points created at iteration 2. It probably won't come as a surprise that the denominator for new end points created at iteration 3 will be $3^3 = 27$; or that iteration 4 will involve fractions where the denominator is $3^4 = 81$. And so on.

Base-3 Representation

Since Cantor set end points all have denominators that are powers of 3, it's easy to express end points in base-3.

First, a bit of nomenclature. Numbers written in base-3 are sometimes referred to as a ternary representation or a triadic expansion or expression. Ternary and triadic both refer to the 3 that plays such a large role in the Cantor set as defined above.

Base-3 works just like base-10 decimals, but the positions to the right of the radix point are now 1/3, 1/9, 1/27 etc instead of 1/10, 1/100, 1/1000 etc. Just as the fraction 2/10 is represented in base-10 as the decimal 0.2, similarly, the fraction 1/3 is represented in base-3 as 0.1.

Base-3 provides a particularly helpful way to further characterize Cantor set end points. Above, we identified the following fractions as end points: 1/9, 2/9, 1/3, 2/3, 7/9, and 8/9. In base-3, these are: 0.01, 0.02, 0.1, 0.2, 0.21, and 0.22. There is a pattern developing in these numbers, but it helps to do one more iteration to see it emerge clearly.

* email: gml@garymlewis.com

In iteration 3, the following new end points get created[2]: 1/27, 2/27, 7/27, 8/27, 19/27, 20/27, 25/27, and 26/27. In base 3, these are 0.001, 0.002, 0.021, 0.022, 0.201, 0.202, 0.221, and 0.222.

Now for the pattern that's emerging in the base-3 representation of Cantor set end points. One, they all terminate. That is, they all have a finite length. In base-10, the fraction 1/3 repeats indefinitely as the decimal 0.333..., but in base-3 the ternary representation is 0.1. Second, and this is more subtle, all of the Cantor set end points include 0 or 2 at any position, but can only include the digit 1 if it's in the last position. For example, 0.1 (1/3) is an end point, as are 0.01 (1/9), 0.001 (1/27), 0.21 (7/9), 0.021 (7/27), and 0.201 (19/27). But a triadic expression that includes the digit 1 at any position other than the last one is not a Cantor set end point. Consider the base-3 number 0.12 (5/9). The digit 1 is not in the last position, and, sure enough, 5/9 is not one of the end points created during our middle-third iterations.

Redefining the Cantor Set

The geometric construction of the Cantor set provides a useful visualization, but it also introduces a recursive operation that never ceases. Number theory provides a cleaner definition.

You'll often see the following definition used to identify *all* points in the Cantor set: "The Cantor set C is the set of points in [0,1] for which there is a triadic expansion that does not contain the digit 1."[3]

This means that only the digits 0 and 2 can appear in a Cantor set point. But, wait, what about all those end points that have a digit 1 in the last position (eg, 1/3 is 0.1)? Well, they're still ok because any terminating 1 can be re-expressed as an infinite series of 2s. For example 0.1 = 0.02222... I prefer to say that a Cantor set point is any triadic expression containing any number of 0s and 2s in any position, but the digit 1 can only appear in the last position, if it appears at all.

So now we have a convenient way to identify Cantor set points. And, note specifically, that this refers to *all* points in the Cantor set, not just the end points created by removal of middle-third segments.

Points Excluded By Dropping Middle-Thirds

By dropping middle-third intervals, we're declaring that these points do not meet the requirements for inclusion in the Cantor set. Specifically, all these points should have a ternary representation that contains the digit 1 in some location other than the last position. Let's see.

The first middle-third that was dropped was the open interval (1/3, 2/3). Open means that the values 1/3 and 2/3 are not included in the points dropped, so we're really dropping $1/3 < x < 2/3$. In base 3, the number 1/3 is 0.1 and the number 2/3 is 0.2. So we're dropping $0.1 < x < 0.2$. For example, 0.11, 0.12, 0.111, 0.112, etc. All of these contain a 1 somewhere other than in the last position, so do not qualify as points in the Cantor set.

Now consider the number 1/2. In base 3, the number 1/2 is 0.11111..., where the digit 1 repeats forever. Based on its ternary representation, the number 1/2 obviously does not qualify for the Cantor set.

At the second iteration, two middle-thirds were dropped: (1/9, 2/9) and (7/9, 8/9). Let's consider (7/9, 8/9). In base 3, the number 7/9 is 0.21 and the number 8/9 is 0.22. So we're dropping $0.21 < x < 0.22$. For example, 0.211, 0.212, 0.2111, 0.2112, etc. All of these also contain a 1 somewhere other than in the last position.

We've only considered a few examples here, but it seems apparent that dropping middle-thirds is consistent with the number theoretic definition of the Cantor set.

All End Points are Included

The geometric visualization of the Cantor set makes clear that all end points, created by recursively removing middle-third intervals, are valid members of the Cantor set.

Let's consider the ternary representation of new end points after iteration 3. Here we're left with 16 total end points and 8 new end points at 1/27, 2/27, 7/27, 8/27, 19/27, 20/27, 25/27, and 26/27.

Their ternary representations are: 1/27 = 0.001; 2/27 = 0.002; 7/27 = 0.021; 8/27 = 0.022; 19/27 = 0.201; 20/27 = 0.202; 25/27 = 0.221; and 26/27 = 0.222.

Again, all these points qualify as members of the Cantor set according to the number theoretic criteria. So it seems apparent that all end points created by removing middle-thirds would also qualify.

End Points are Not the Only Points

Based on the geometric visualization discussed above, it certainly appears that only end points are members of the Cantor set. This, in fact, is not the case. Simple examples illustrate that there are very many points in the Cantor set that are not end points.

Consider the number 3/4. The ternary representation of 3/4 is 0.2(02), meaning that the 02s repeat forever (ie, 3/4 = 0.2020202...). Recall that Cantor set end points, when expressed in base-3, all have a finite length. Meaning no infinite repeats, and that 3/4 is not an end point.

It's easy to find many more such points. For example, consider 0.(0220) (ie, 3/10). This qualifies as a member of the Cantor set, yet does not terminate, and so is not an end point.

It turns out there are many points in the Cantor set that are not end points. Here's how Peitgen, Jürgen, and Saupe[3] conclude): "... a number in which we pick digits 0 and 2 at random will belong to the Cantor set but are not end points, and those are, in fact, more typical for the Cantor set. In other words, if one picks a number from C at random, then with probability 1, it will not be an end point."

Wow! There are an infinite number of end points in the Cantor set. But there are even more points that are not end points.

Take-Aways at This Juncture

As an interim summary, here are the key features described above.

1. Despite introducing huge gaps in the unit interval [0,1] to produce the ternary Cantor set, an infinite number of points remain after the iterative removal of middle-thirds.
2. All the points discussed above are rational numbers (ie, fractions). As we'll see later, the Cantor set also includes irrational numbers.
3. When represented in a base-3 expression, all rational numbers in the Cantor set can include 0 or 2 at any digit position, but a 1 can only appear in the very last position.
4. When represented in a base-3 expression, all rational numbers in the Cantor set either terminate or repeat indefinitely.
5. Base-3 expressions that terminate are all end points created by removing middle-thirds from intervals. There are an infinite number of end points.
6. Base-3 expressions that repeat are not end points. There are an infinite number of these points, and they vastly outnumber the number of end points.

Another Way to Identify Cantor Set Points

In the material above, I discussed two ways to generate the Cantor Set. The so-called geometric approach involved the iterative removal of middle-third segments from the unit interval [0,1]. This was visually appealing, but left the mistaken impression that only end points would remain in the Cantor set. A second approach, based on number theory, clarified the situation by using base-3 representations of points on the unit interval. A rational number in the Cantor set is just any fraction in the unit interval whose base-3 expression contains only 0s and 2s, or a single 1 in the very last position, and that either terminates with a finite length or repeats indefinitely.

Mandelbrot[4] suggests a third way to identify points in the Cantor set. It, too, involves an appealing visual component. You choose any initial point in the unit interval, and then iterate that starting point to a new point according to a functional transformation. You then take that new point and again use the transformation function to get yet another point. And so on. This process can be plotted graphically, providing an orbit of points. Mandelbrot asserts that most starting points will rapidly fall off the graph by approaching negative infinity. Starting points whose orbits remain on the graph, however, are members of the Cantor set.

Mandelbrot called his function an "inverted V" transformation. It is given by the following equation:

$f(x) = (\frac{1}{2} - |x - \frac{1}{2}|)/r$, with $r = \frac{1}{3}$

Example 1: A Point Not in the Cantor Set

As discussed previously, the number 1/2 is clearly not a member of the Cantor set. If we let the initial starting point be 1/2, it should quickly head toward negative infinity when iterated with Mandelbrot's inverted V transformation. As shown below, indeed that does happen.

{1/2, 3/2, -(3/2), -(9/2), -(27/2), -(81/2), -(243/2), -(729/2), -(2187/2), -(6561/2)}

Figure 1 illustrates this more clearly in a time series plot. Note: All figures were prepared using matplotlib[5].

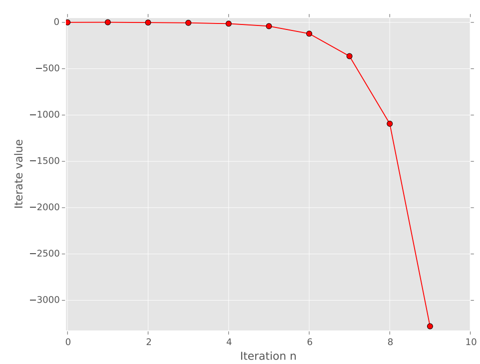

FIG. 1. Time Domain Plot: Initial Condition 1/2

Example 2: An End Point

The number 1/3 is an end point in the Cantor set, created by dropping the first middle-third. According to Mandelbrot, starting with 1/3 and iterating the inverted V function should generate an orbit that stays on the Cantor set. Below are the first 10 numbers in the orbit. They do remain on the Cantor set, by moving to and remaining on the fixed point at 0.

{1/3, 1, 0, 0, 0, 0, 0, 0, 0, 0}

Example 3: Any End Point

We know that end points all have denominators that are powers of 3. So we could randomly select an integer power, say 7. Then the point at $1/(3^7) = 1/2187$ should orbit on the Cantor set. Let's see. As shown below, it too heads for 0.

{1/2187, 1/729, 1/243, 1/81, 1/27, 1/9, 1/3, 1, 0, 0}

Figure 2 shows the corresponding time series plot.

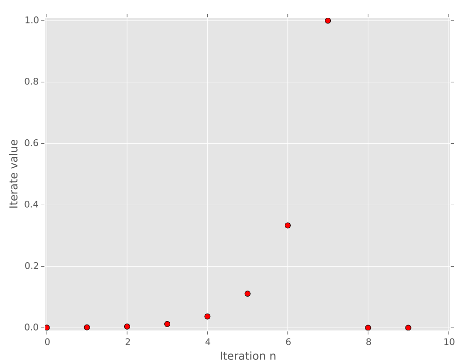

FIG. 2. Time Domain Plot: Initial Condition 1/2187

While we've looked at the orbits of only two end points, both got drawn to a fixed point at 0. It turns out that this is true generally and provides another way to identify end points. In this definition, an end point is any initial starting point in the unit interval that gets attracted to 0 when iteratively transformed under Mandelbrot's inverted V function.

Example 4: Non-End Points

As discussed above, the number 1/4 has a repeating base-3 representation of 0.(02); ie, 1/4 = 0.020202... . According to the number theoretic definition, 1/4 is a member of the Cantor set. In which case, Mandelbrot would say that an orbit starting with 1/4 and iterated forward using the inverted V function should stay on the Cantor set. That does happen. The orbit quickly moves to a fixed point at 3/4.

{1/4, 3/4, 3/4, 3/4, 3/4, 3/4, 3/4, 3/4, 3/4, 3/4}

Example 5: Periodic Non-End Points

As discussed above, there are an infinite number of base-3 representations that repeat. None of them are end points. Many of these will orbit under the inverted V function in periods of length 2, 3, 4 and so on. For example, starting with the fraction 1/10, which is 0.(0022) in base-3, the orbit quickly settles into a period of 2, alternating between 3/10 and 9/10. See initial iterations below.

{1/10, 3/10, 9/10, 3/10, 9/10, 3/10, 9/10, 3/10, 9/10, 3/10}

We can also see this in Figure 3, called a return graph because it plots each iterate versus its subsequent iterate, making it ideal for identifying periodic behavior. In this case, the graph eliminates transient behavior before the period 2 cycle begins.

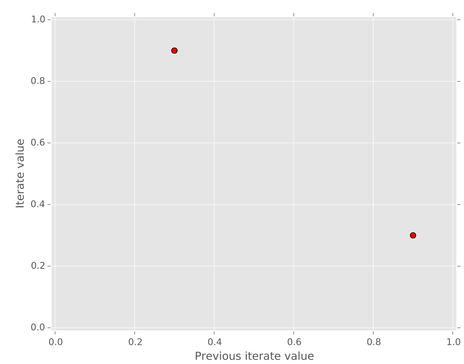

FIG. 3. Return Plot: Initial Condition 1/10

Example 6: Larger Periods

How about a period 6? Can that occur in the inverted V transformation? Well, yes, it can. Start with the number 1/91. As shown below, the orbit quickly settles into an orbit of period 6. Very nice.

{1/91, 3/91, 9/91, 27/91, 81/91, 30/91, 90/91, 3/91, 9/91, 27/91, 81/91, 30/91, 90/91, 3/91, 9/91}

Figure 4 clearly identifies the period 6 behavior with an initial condition of 1/91. Again, transient iterates preceding the periodic behavior are not shown in the graph.

Example 7: Periodic Orbits of Non-End Points, in General

It turns out that the existence of periodic orbits for non-end points can be summarized very simply. Consider the fractions $1/(3^k + 1)$, where k = 1,2,3, etc. These are the fractions 1/4, 1/10, 1/28, 1/82 and so on.

A starting point at 1/4 orbits to the fixed point at 3/4, as we saw above. As we also saw above, the starting point of 1/10 iterates to a period 2 orbit. Now consider the starting point 1/28. It will settle into a period 3 orbit, as shown below.

{1/28, 3/28, 9/28, 27/28, 3/28, 9/28, 27/28, 3/28, 9/28, 27/28}

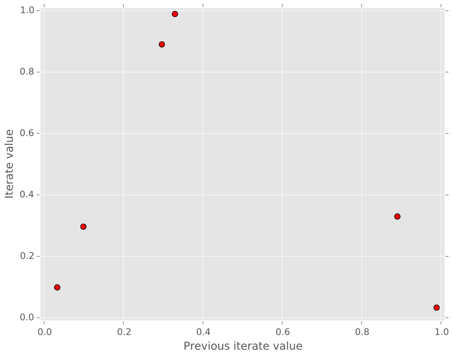

FIG. 4. Return Plot: Initial Condition 1/91

The next fraction in the series, with k = 4, is $1/(3^4 + 1)$ = 1/82. Iteration of this starting point with the inverted V function will end in an orbit of period 4. And so on.

But it gets better, and easier. Let's examine again the base-3 representations of the fractions 1/4, 1/10, 1/28, 1/82, etc. They are: 1/4 = 0.(02); 1/10 = 0.(0022); 1/28 = 0.(000222); and 1/82 = 0.(00002222). See the pattern developing? The fraction 1/4 orbits to a fixed point (ie, period 1), and it's ternary representation has repeating digits of (02). The fraction 1/10 orbits to a period 2, and it's ternary representation has repeating digits of (0022). The fraction 1/28 orbits to a period 3, and it's ternary representation has repeating digits of (000222). And, finally, the fraction 1/82 orbits to a period 4, and it's ternary representation has repeating digits of (00002222).

See the pattern? What do you think the ternary representation 0.(0000022222) = 1/244 will do? Sure, it will orbit to a period 5.

The ternary representations for 1/4, 1/10, 1/28, and 1/82 that we just considered can be expanded easily to any number of fractions that will have periodic orbits. I'll just show one example. Let's consider 0.(0022) = 1/10 that has a period 2 orbit. If you multiply the ternary number by 0.1 or 0.01 or 0.001 etc, you get new fractions that will also have period 2 orbits. These new fractions would be 0.0(0022) = 1/30, 0.00(0022) = 1/90, and 0.000(0022) = 1/270.

Practice Exercise

The Cantor set has numerous patterns. Here's an example for you to ponder. For a solution, see Appendix A on page 6.

1. The fraction 1/13 = 0.(002) orbits to a period 3. Based on this, find another fraction that also has period 3. What is that fraction in base-10?

Chaotic Orbits on the Cantor Set

The discussion above only considered rational numbers, either those that terminate with finite length (i.e., end points) or those that repeat indefinitely. But, what about irrational numbers? According to the number theoretic definition, an irrational number between 0 and 1 that is expressed in base 3 using only 0s or 2s should also be a valid member of the Cantor set.

Irrational numbers do not terminate, nor do they ever repeat. Common examples include π and $\sqrt{2}$. Both are just infinitely long strings of digits. Neither of these is a member of the Cantor set because their base 3 representation contains the digit 1. But an infinite number of other irrationals do exist that contain only 0 and 2 in base 3. An initial condition that is one of these irrational numbers should orbit under the inverted V function and remain on the Cantor set.

That does indeed happen, and in a most wonderful way. The orbits are chaotic, moving from point to point, never settling down as a fixed point or a periodic orbit, and never repeating a point they've already visited.

We'll see this happen shortly, but first a note about methodology. It turns out that long orbits generated with a function such as Mandelbrot's inverted V function can sometimes run afoul of precision problems. In such cases, the iterate values become unreliable. So it's helpful to manage the precision involved in iterations so that this problem does not happen.

With the inverted V function, there's a particularly easy way to manage precision of the iterates. It's a procedure called a binary shift automaton. I won't discuss the details here, but Joseph McCauley[6] provides a good explanation as applied to the inverted V function. I used McCauley's procedure to produce the two figures that follow.

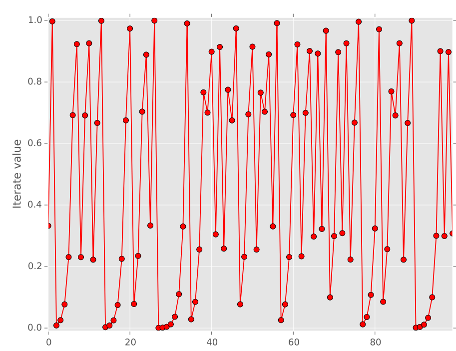

FIG. 5. Time Domain Plot: Irrational Initial Condition

Note in Figure 5 the chaotic orbit generated with an irrational initial condition. This graph plots 100 iterates without a noticeable pattern of any kind.

The corresponding return map appears in Figure 6. I added filling to emphasize how barren the Cantor set is. Yet remarkably, it contains an infinite number of end points, an even more infinite number of fractions whose base 3 representations repeat, and an even more infinite number of irrationals.

What we're seeing here is chaotic dynamics on a fractal, ie, the Cantor set. Michael Barnsley[7] provides much more discussion of such behavior.

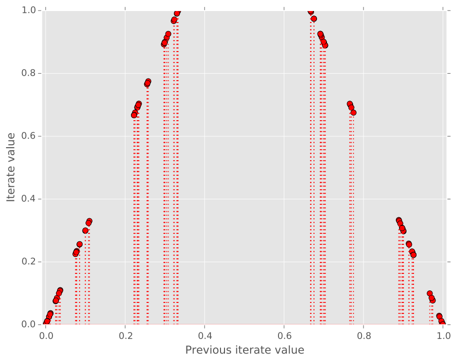

FIG. 6. Return Plot: Irrational Initial Condition

I find this figure starkly beautiful, but somewhat daunting too. What, if anything, does it mean for the natural world?

Concluding Comment

I don't know what it means. It very likely depends on whether the Cantor set, which is a mathematical object, has any counterpart in nature. Roger Penrose [8] seems dubious when he characterizes Cantor's theory of the infinite this way: "In my opinion, it is one of the most profoundly beautiful mathematical contributions in the whole of mathematical history. However, extraordinarily little of it seems to have relevance to the workings of the physical world as we know it."

Clearly the infinite and the infinitesimal still chafe in mathematics and theoretical physics. But I'll leave those musings for another article.

Appendix A: Answers to Practice Exercise

Here is the practice exercise that appeared on page 5.

1. The fraction 1/13 = 0.(002) orbits to a period 3. Based on this, find another fraction that also has period 3. What is that fraction in base-10?

Solution

This question contains an unasked prior question that is worth addressing first. How do we know that the base-3 repeating number 0.(002) is the base-10 fraction 1/13?

The answer involves a trick that is worth remembering. If x = 0.(002) = 0.002002..., then 1000x = 2.002002... and a simple subtraction will clear the repetend. But, remember we're using base-3 numbers here, so 1000 in base-3 is actually 27 in base-10.

$$\begin{aligned} x &= 0.002002\ldots \\ 1000x &= 2.002002\ldots \\ 222x &= 2 \end{aligned}$$

In base-10, this last line is $26x = 2$, or $x = 1/13$. Let's now address the actual question.

On page 5, I considered the base-10 fraction 1/10, which is 0.(0022) in base-3. This has a period 2 orbit. The key sentence is: "If you multiply the ternary number by 0.1 or 0.01 or 0.001 etc, you get new fractions that will also have period 2 orbits."

The concept applies equally well to orbits of any period. Our number is 0.(002). It's orbit has period 3, so 0.0(002), 0.00(002), 0.000(002), etc, will also have periods of 3. You can use the "trick" above to convert any of these base-3 numbers to base-10. Try it for 0.0(002). You'll see that the base-10 fraction is 1/39.

[1] Joseph Warren Dauben, *Georg Cantor: His Mathematics and Philosophy of the Infinite* (Princeton University Press, 1990).

[2] It's easy enough to calculate end points manually, but eventually I wrote a small Python program to identify end-points at any scale, as base-10 fractions and their ternary representation.

[3] Heinz-Otto Peitgen, Hartmut Jürgens, and Dietmar Saupe, *Chaos and Fractals: New Frontiers of Science* (Springer-Verlag, 2004).

[4] Benoit B. Mandelbrot, *The Fractal Geometry of Nature* (W. H. Freeman and Company, 1982).

[5] J. D. Hunter, "Matplotlib: A 2d graphics environment," Computing In Science & Engineering **9**, 90–95 (2007).

[6] Joseph McCauley, *Chaos, Dynamics, and Fractals: An Algorithmic Approach to Deterministic Chaos* (Cambridge University Press, 1994).

[7] Michael F. Barnsley, *Fractals Everywhere* (Dover Publications, Inc., 2012).

[8] Roger Penrose, *The Road to Reality: A Complete Guide to the Laws of the Universe* (Vintage Books, 2007).