Discerning New Patterns in the Generalized Collatz Function Utilizing Big Data Analysis

Conference Paper · January 2017

1 author:

Shyam Sai
Illinois Mathematics and Science Academy

Discerning New Patterns in the Generalized Collatz Function Utilizing Big Data Analysis

Shyam Sai

Illinois Mathematics and Science Academy

under the direction of

Dr. James Choi

Sabio Research

as submitted to the Siemens Competition in Math, Science, and Technology

as presented at the Joint Mathematics Meetings

January 4-7, 2017

Atlanta, Georgia

future presentation at the American Association for the Advancement of Science Annual Meeting

February 18-19, 2017

Boston, Massachusetts

Abstract

The Collatz Conjecture has eluded mathematicians for decades. The small fact that for any initial number n, when recursively inputted to the function $f(n) = \begin{cases} \frac{n}{2}, & \text{if } n \text{ is even} \\ 3n+1, & \text{if } n \text{ is odd} \end{cases}$, will converge to 1 has been a point of debate in the field of math for almost 80 years. However, the Generalized Collatz Function, a variable form of the Collatz, is a relatively new function that has not been studied by many. Posited by Zhongfu and Shiming, and in a different form by Bruschi, the form for the Generalized Collatz Function used in this research was $f(n) = \begin{cases} \frac{n}{d}, & \text{if } n \equiv 0 (mod\ d) \\ mn+1, & \text{if } n \neq 0 (mod\ d) \end{cases}$, where d is a list of consecutive positive primes until the d^{th} prime. We use Mathematica to run the generalized Collatz sequence 1 million times and use big data scanning tools to create visualizations with high-probability to contain a pattern. We discern three new patterns from scanning this big data. One, that all numbers that are divisible by a prime number squared can be set to n, so that, with a $m = n - 1$, creates a hailstone sequence ending in {4, 2, 1}. Two, that all generalized Collatz sequences with $m = n - 1$ will converge to 1, for any divisor, in three iterations. Three, that all Mersenne numbers, when set to m, will, for any divisor and for any initial number, cycle. This project, with the help of 3D graphing technology, was able to find never before seen patterns in a thought-to-be unpredictable function, the Generalized Collatz Function.

Introduction

The Collatz Conjecture has eluded mathematicians since its foundation in 1937. Many approaches have been used to try to tackle this "Half or Three *N* Plus One problem". Paul Erdös said himself that "Mathematics may not be ready for such problems." [1] But this "impossible" problem can be addressed in one function:

$$f(n) = \begin{cases} \frac{n}{2}, & \text{if } n \text{ is even} \\ 3n + 1, & \text{if } n \text{ is odd} \end{cases}$$

The Collatz Conjecture states that all initial numbers (n), when recursively inputted to the Collatz function, will converge to 1. Mathematicians around the world have tried to prove this conjecture, but to no avail.

The history of Collatz Conjecture started with Lothar Collatz. Lothar Collatz, a mathematician born in the German Empire in 1910, posed the famous Collatz Conjecture. Collatz created the Collatz Conjecture as a novelty of sorts, and, dying in 1990 in Bulgaria, he is mostly remembered for that Conjecture. [2]

Though mathematicians from the Americas, Europe, and Asia have all been stumped by the Collatz, it, in its heart, is a simple function.

Take any natural numbern. If n is even, divide it by 2. If n is odd, multiply it by 3 and add 1. Iterate this process by applying the same procedure to the result. Repeat this process indefinitely. This is represented in function form below:

$$f(n) = \begin{cases} \frac{n}{2}, & \text{if } n \text{ is even} \\ 3n + 1, & \text{if } n \text{ is odd} \end{cases}$$

This process creates what mathematicians call "Hailstone Sequences". [3] Hailstone Sequences are simply the sequence of numbers produced by running a number through the Collatz function, in a recursive fashion. Take a couple examples.

```
In[85]:= Map[sfClassicCollatz, Range[2, 11]]

In[86]:= {{2, 1},
  {3, 10, 5, 16, 8, 4, 2, 1},
  {4, 2, 1},
  {5, 16, 8, 4, 2, 1},
  {6, 3, 10, 5, 16, 8, 4, 2, 1},
  {7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1},
  {8, 4, 2, 1},
  {9, 28, 14, 7, 22, 11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1},
  {10, 5, 16, 8, 4, 2, 1},
  {11, 34, 17, 52, 26, 13, 40, 20, 10, 5, 16, 8, 4, 2, 1}}
```

Figure 1- Example Hailstone Sequences

The simple observation that all numbers converge to 1 is one that puzzles mathematicians' minds the world over. The fact that all starting numbers result in a convergence to 1, is, in fact, the Collatz Conjecture. Yet many difficulties arise in this search for a solution.

One problem that arises is the fickleness of the length of the resulting sequence (also known as Hailstone sequences). The sequence converges to one at differing lengths, creating issues in algorithms involving the length of the sequence. For example, the Hailstone sequence for the initial number 41 has 106 numbers in it ranging from 1 to 9232, whereas the Hailstone sequence for 37 has 21 numbers. Size of numbers in hailstone sequences is a factor as well. Small numbers can fluctuate greatly, while larger numbers can fall right to 1.

When put together, the fickleness of the Collatz, in both size and length, creates a difficult situation for programmers and mathematicians looking to construct the Collatz. The search for a pattern in the Collatz Function has been fruitless.

Another notable fact is that numbers in the form 2^n take $(n + 1)$ steps to reach one, as seen below.

Table 1 - Length for Hailstone Sequences with an initial number being a Power of 2

Initial Number	Power of 2	Length
2	1	2
4	2	3
8	3	4
16	4	5
32	5	6
64	6	7
128	7	8

This is because the power of two will always be divisible by 2 until it reaches 1.

Using this fact, it seems reasonable that all Collatz sequences end in the 4, 2, 1 sequence, a conclusion mathematicians have come to. When the sequence reaches 4, the 1 is eminent in 2 divisions, at which point, 1, an odd number, is multiplied by 3 and added 1 to reach 4, at which the cycle repeats. The 4, 2, 1, cycle is one that is synonymous with the Collatz.

Gerhard Opfer released a paper in May 2011 titled *An analytic approach to the Collatz 3n + 1 Problem.*[4] From the Hamburger Beiträge zur Angewandten Mathematik (Hamburg Post for Applied Mathematics) at the University of Hamburg, Opfer says in this paper that he has proven that the Collatz Conjecture is true. Opfer, who himself was a student of Lothar Collatz, claimed he had solved the decades long problem, sparking much internet debate. [5] Many doctorates in the field of mathematics were befuddled by this complicated proof.

In ideology, Opfer's proof follows a simple backwards approach. Opfer considers implementing the Collatz function, but backwards. From any number, we can seemingly go "backwards" to find which numbers could have created that number. If the backwards Collatz is represented by the function $bC(x)$, then we have $bC(2k) = k$. To reach k by dividing, we have to have the number $2k$ first. From k, we can go backwards to $2k$. In a similar sense, to get k from multiplying, we have $bC((k-1)/3) = ((3(k-1))/3) + 1) = k$. From k, we can also go backwards to $(k-1)/3$. Starting from 1, we can go backwards to 2; from 2 we can go backwards to 4; from 4 we can go to 8 or 1; from 8 we get 16; from 16 we get 32 or 5; and so on. In this way we can build up an infinite tree of numbers. [6]

Opfer combined this with holomorphic functions on the complex plane. Opfer is building off the work of Lothar Berg and Gunter Meinardus who connected the holomorphic functions with the unit open disc, the disc centered at $|z| < 1$. [7]

Opfer introduces two notations in his paper, building off Berg and Meinardus' work.

$\phi 0(z) :=$ 1 for all $z \in C$, $\phi 1(z) := z\, 1 - z$ for all z with $z \in D$,

$\Delta_2 := \{\phi 0, \phi 1\}$, where $\{\ldots\}$ symbolizes the linear hull over C.

Using the two linear functional equations Berg and Meinardus introduced in 1994 and 1995, below, Opfer says, by showing that all coefficients vanish except for those whose indices lie in a certain congruence class, that every natural number is contained in his tree of numbers.

1. The pair of functional equations can be solved by all functions $\phi \in \Delta 2$.

2. The Collatz conjecture is true if and only if there are no solutions of the pair of functional equations outside of $\Delta 2$.

Opfer goes on to define two linear operators, U and V.

$U[h](z) := h(z) + h(-z) - 2h(z^2)$

$V\,[h](z) := 2H[h](z) - T1[h](z) + T1[h](-z)$, where $T1[h](z) := h(z^3)$ and

$H[h](z) := \frac{1}{3z}\,(h(z^2) + \lambda h(\lambda z^2) + \lambda^2 h(\lambda^2 z^2))$ and $\lambda := \frac{1}{2}\,(-1 + \sqrt{3}\,i)$.

Opfer defines the *kernel* of the two linear operators as K.

$K := ker(U, V\,) := \{h \in H : U[h] = 0, V\,[h] = 0\}$

Opfer follows from Berg and Meinardus' two notation and goes on to prove using the holomorphic plane and his backward tree and annihilation graphs that $K = \Delta_2$. It then follows that the Collatz Conjecture is true, as per Berg and Meinardus' two linear functional equations. Opfer says on the last page of his paper that the Collatz Conjecture is true. [4]

Alas, Opfer's proof is blemished, one that internet bloggers found and also one Opfer published in his second revision. A statement on the 11th page of Opfer's paper was incomplete, and in Opfer's own words, "Thus, the statement 'that the Collatz conjecture is true' has to be withdrawn, at least temporarily." [8]

The Generalized Collatz however, uses a different approach than the Classic Collatz Conjecture. The Generalized Collatz Function is proposed and defined by Zang Zhongfu and Yang Shiming [9] as

$$f(x) = \begin{cases} \dfrac{x}{p_{i1}p_{i2}p_{i3}\cdots p_{ik}}, \text{where the } p_i \text{ are exactly the primes } \leq\ p_n \text{dividing the numerator,} \\ \qquad\qquad p_{n+1}x + 1,\ \ \text{if no prime } p_i\ \leq\ p_n \text{ divides x.} \end{cases}$$

Besides this paper by Zang Zhongfu and Yang Shiming, in which the Generalized Collatz isn't analyzed in much depth, there are a few papers following their work. John Lesieutre, under Zuoqin Wang, analyzes the individual trajectories of the hailstone sequences generated from the Generalized Collatz Function defined in Zhongfu and Shiming. Lesieutre proves that the generalized Collatz, for the first 10000 primes as a divisor, will converge to the constant $C \approx .520$. [10] Lesieutre also creates encoding matrices for specific multiples. With $p_{n+1} = 7$, Lesieutre derives the encoding matrix below. [10]

$$A_9(13) = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 2 & 0 & 0 & 0 & 0 & 1 \\ 3 & 2 & 1 & 3 & 2 & 2 & 4 & 3 & 1 \end{pmatrix}$$

However, Lesieutre is unable to find any patterns in the end form of the Generated Collatz, as he mainly focuses on the trajectories of the hailstone sequence. This gap in knowledge, in pattern in the end-results of the Generalized Collatz, is where we base our work. We build on the Generalized Collatz defined in Zhongfu and Shiming.

Mario Bruschi defines the Generalized Collatz differently in his analysis of the function. [11] Bruschi defines the function as

$f(n,b,m) = \begin{cases} \frac{n}{b}, \text{ if } n \equiv 0(mod\ b) \\ (b^m+1)n + b^m - (n\ mod\ b^m), \text{ if } n \not\equiv 0(mod\ b) \end{cases}$, where n and m are positive integers, and b is a prime.

Bruschi simply introduces his Generalized Collatz function to the world in his paper, and does not go in to much in depth into analysis of the function. However, he does discover the cycles that the Generalized Collatz Sequences go into. If $= 3$, $n = 5$, and $m = 1$, Bruschi discovers that the hailstone sequence generated falls into an infinite recursive cycle. [11]

$$7, 30, 10, 42, 14, 57, 19, 78, 26, 105, 35, 141, 47, 189, 63, 21, 7$$

These cycles are also visible in our testing of the Generalized Collatz, and we also dubbed them cycles, as we built off the prior work of Bruschi. [11]

Hayden Messerman, Joey LeBeau, and Dominic Klyve also delve into their own version of the Generalized Collatz Function. Their testing occurred on the following Generalized Collatz Function. [11]

$$f(n) = \begin{cases} \dfrac{n}{m}, & \text{if } n \equiv 0 (mod\ m) \\ an + b, & \text{if } n \not\equiv 0 (mod\ m) \end{cases}$$

Messerman, LeBeau and Klyve introduce a new definition in the field of “cycle gravity”. They define the cycle gravity of a hailstone sequence as the proportion of integers that end up in a recursive cycle (similar to the ones Bruschi found), to the total length of the hailstone sequence. The motivation for this terminology was to better understand the recursive cycles in the Generalized Collatz Function. [12]

However, Messerman, LeBeau, and Klyve are unable to find any significant general patterns pertaining to the predicting the end result of a Generalized Collatz Function. They, instead, turn to focus on one set variant of the Generalized Collatz, seen below: [12]

$$f(n) = \begin{cases} \dfrac{n}{2}, & \text{if } n \equiv 0 (mod\ 2) \\ 3n + b, & \text{if } n \not\equiv 0 (mod\ 2) \end{cases}$$

After narrowing their search, Messerman, LeBeau and Kylve are able to find multiple corollaries pertaining to specific values in their analysis. First, if b is an odd multiple of 5, the hailstone sequence will contain at least two cycles of length 44 and two cycles of length 8. They also observe that if b is an odd multiple of 13, the hailstone sequence will have at least seven sequences of length 13. Thirdly, If b is an odd multiple of 29, the hailstone sequence will contain two cycles of length 106. Lastly, they derive that the hailstone sequence created will have precisely one cycle if and only if b is a power of 3. [12]

This discovery in patterns in one variant of the Generalized Collatz led us to aspire to discern multiple significant patterns in end result of the Generalized Collatz Function.

Thus, the purpose of this project is to discern novel patterns in the whole of the Generalized Collatz by building on the work of Zhongfu and Shiming, Bruschi, and Messerman, LeBeau and Kylve, quite possibly to assist the work of Opfer in creating a proof of the Collatz Conjecture.

Methods

First, the Classic Collatz Conjecture must be programmed and tested.

Program: $sfClassicCollatz[n]$.

INPUT: Positive Integer

OUTPUT: List containing initial value followed by next iteration, followed by the next, and the next, until the number 1 appears.

Step 1: Print initial number n to list.

Step 2: Check if n is Even.

Step 3: Use an IF statement to run $\frac{n}{2}$ if Step 1 returns yes, ELSE run $3n + 1$. Return the output from either of these calculations.

Step 4: Used the returned statement and assign it to n. Print n to list a.

Step 5: Repeat Steps 2, 3, and 4 as long as n>1.

Step 6: Return list a.

This algorithm for generating the Classic Collatz sequence worked without fail for all initial n tested until 1 million. However, the goal of this project was not to directly prove the Collatz Conjecture, as that feat hasn't been completed in 80 years. The aim of this project was to generalize the Collatz Conjecture, and to attempt to discern a pattern in multiple runs of the generalization and the hailstone sequences it generated.

Instead of using $3n + 1$ and $n/2$, a variable approach was taken. An algorithm changing the multiplier and the divisor in the Collatz Conjecture was created. We based the generalized Collatz function we tested off previous work in the field, including that of Zhongfu and Shiming and Bruschi. This algorithm used the following logic:

$f(n) = \begin{cases} \frac{n}{d}, & \text{if } n \equiv 0(mod\ d) \\ mn + 1, & \text{if } n \not\equiv 0(mod\ d) \end{cases}$, where d is a list of consecutive positive primes until the d^{th} prime.

The Generalized Collatz function we used was a compilation of previous work in the field, and what generalizations worked. We include the denominator used in Zhongfu and Shiming by using the list of prime numbers. We include the $\mathrm{n} \equiv 0(\mathrm{mod\ d})$ "If" statements found in Bruschi's work.

A novel approach was taken to graph these generalized Collatz sequences. This approach transpired by running the generalized Collatz algorithm for 100 initial numbers, from 1 to 100, saving the resulting hailstone sequence each time. This would be done for 100 multiples and 100 divisor lists, creating 10000 data sets. Each data set would contain 100 lists, the sequences generated from the 100 initial numbers tested. In summary, a program was needed to run the following logic:

$$f(n) = \begin{cases} \frac{n}{r}, & \text{if } n \equiv 0(mod\ d) \\ mn + 1, & \text{if } n \not\equiv 0(mod\ d) \end{cases} \quad for\ n \in [1,100], d \in [1,100], m \in [1,100]$$, where d is a list of consecutive positive primes until the d^{th} prime.

So, a program was written to do exactly that.

Program: $neoCollatz[n, d, m]$.

INPUT: Initial number (n), multiple (m), list of prime numbers until divisor (d).

OUTPUT: List containing generalized Collatz sequence for specific m and d.

Step 1: Check if n is divisible by any of the numbers in the list d. If so, divide out all of the divisors.

Step 2: Multiply m by n and add 1. Append this value to the final list. Assign this value to n.

Step 3: Repeat Steps 1 and 2 until one of the following conditions returns false: n is equal to a number already in the final list, or n is greater then 2^{64}. Return the final list once completed.

The two conditions to break the loop of division and multiplication were in the following logic: If n is a repeat of a number that has come before in the hailstone sequence, meaning the sequence has cycled and that the sequence should stop there. If n exceeds 2^{64}, it is exceeding the capacity of the 64-bit computer we used without going into Software. So, if n exceeded 2^{64}, the sequence was stopped and marked as diverging to infinity.

The program $neoCollatz$ was then run for n, m, d in the range 1 to 100. This program generated 1million data sequences. The data sequences were then analyzed. Sequences that converged to the Collatz sequence of {4, 2, 1} were apportioned, creating a 3D graph. This 3D graph contained 3D coordinates with 3 axes, one each for the initial number of n, multiplier$m$, and divisor$d$. The same was done for coordinate triplets that created hailstone sequences that converged to 1, and for coordinate triplets that created hailstone sequences that cycled.

Results

The algorithm used to run $neoCollatz$ for 100 values of n, m, and d generated 1 million sequences. The automated rapid scanning of big data in Mathematica resulted in a much smaller set of usable, pattern-discernable graphs, 3 of which were high-probability of containing a pattern. A visual review confirmed this.

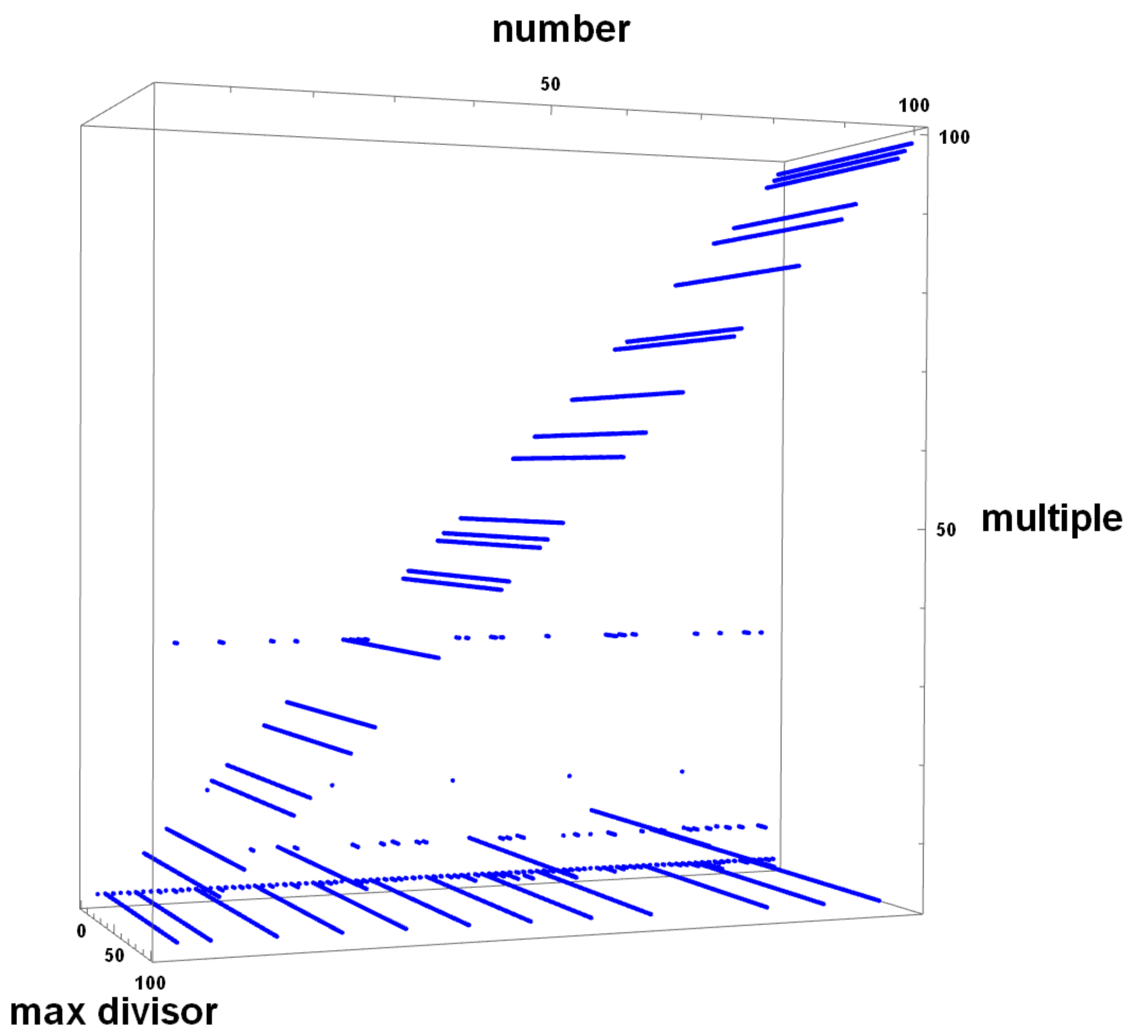

Figure 2- Graph containing all {n, d, m} coordinates that cycled into the {4, 2, 1} cycle. The coordinate triplets that created hailstone sequences that converged into the {4, 2, 1} Collatz sequence were plotted on a 3 axes plot with initial number, divisor, and multiple on each axis.

Figure 2 was created using the dataset created by running $neoCollatz$ 1 million times to create the 1 million sequences. The rapid data scanning tools were built to look for all hailstone sequences with last three numbers of {4, 2, 1}. That sequence was marked as cycling into the {4, 2, 1} cycle, and the initial number, multiple, and divisor triple that created that sequence was graphed in Figure 2.

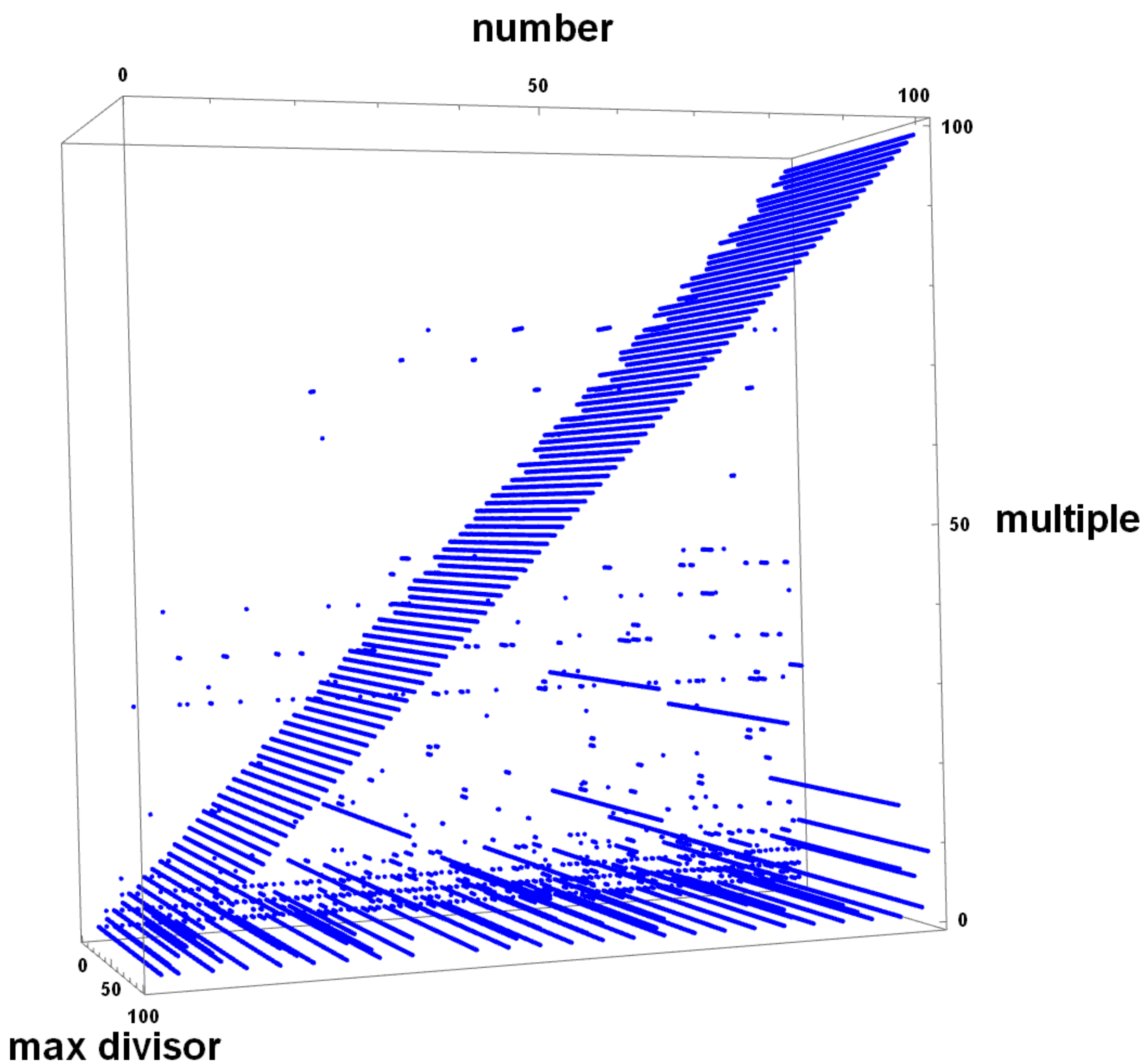

Figure 3 - Graph containing all {n, d, m} that converged to 1. The coordinate triplets that created a hailstone sequences that converged to 1 were plotted on a 3 axes plot with initial number, divisor, and multiple on each axis.

Figure 3 was created using the same dataset used in Figure 2. The rapid data scanning tools were built to look for all hailstone sequences with last number of 1. That sequence was marked as converging to 1, and the initial number, multiple, and divisor triple that created that sequence was graphed in Figure 3.

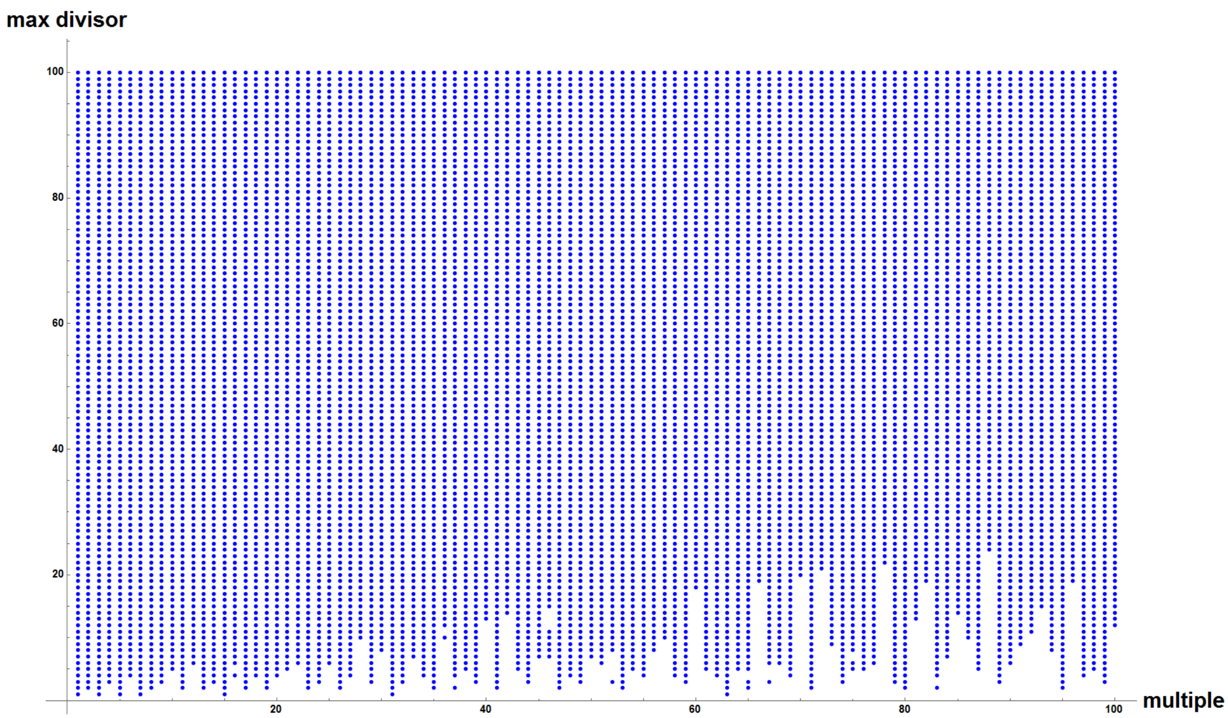

Figure 4 - Plot showing all m-d pairs that cycle for all 100 n. The coordinate triplets that created hailstone sequences that cycled were plotted on a 3 axes plot with initial number, divisor, and multiple on each axis. It was observed that the initial number had no effect on certain $m-d$ pairs. In other words, no matter the initial number$n$, these $m-d$ pairs cycled.

Figure 4 was created by running a 2D data scanner instead of a 3D data scanner. The big data scanner was instructed to find multiple-divisor pairs that, for all 100 initial numbers tested, cycled in a recursive sequence. These multiple-divisor pairs were then graphed, with multiple on the x-axis and divisor on the y-axis.

Discussion

First, instances where the classic $\{4, 2, 1\}$ Collatz cycle appeared were analyzed. 4388 instances of the {4, 2, 1} cycle was found in the 1 million sequences created. These 4388 coordinate triplets were plotted in Figure 2.

In Figure 2, a diagonal line brings itself out. A line defined by the equation multiple equal to the initial number minus 1 ($m = n - 1$), is seen. We found that for certain numbers on this line, all hailstone sequences generated, no matter the divisor, converged to the {4, 2, 1} cycle. This suggests these numbers follow a pattern. The n values for which all hailstone sequences when $m = n - 1$ converged to {4, 2, 1} were found, giving the following list.

$4, 9, 12, 18, 20, 25, 28, 36, 44, 45, 49, 50, 52, 60, 63, 68, 75, 76, 84, 90, 92, 98, 99, 100$

Finding a pattern in these digits was attempted multiple times. Square numbers in the form n^2 were tested first. We found $2^2 = 4, 3^2 = 9, 5^2 = 25, 6^2 = 36, 7^2 = 49, 10^2 = 100$ in the list of initial numbers for which all hailstone sequences converged to {4, 2, 1} if the multiple was one less than the initial number. We also found many multiples of 9. $9 * 1 = 9, 9 * 2 = 18, 9 * 4 = 36, 9 * 5 = 45, 9 * 5 = 45,\ 9 * 7 = 63, 9 * 10 = 90, 9 * 11 = 99$. Attempts to discern a pattern using basic exponential and multiplication rules bore no fruit.

However, we persevered and started to look and browse the On-Line Encyclopedia of Integer Sequences. After two days of searching, we were able to stumble upon an integer sequence that matched our sequence almost exactly. This Integer Sequence, A038109, is the integer sequence of numbers that are exactly divisible by the square of a prime. [13] The first 24 values of this integer sequence are listed below.

$$4, 9, 12, 18, 20, 25, 28, 36, 44, 45, 49, 50, 52, 60, 63, 68, 72, 75, 76, 84, 90, 92, 98, 99, 100$$

This integer sequence contains all 24 terms of the list of initial numbers we found converged to {4, 2, 1} if the multiple is exactly one less than the initial number, but the integer sequence found also contains 72. To combat this, another integer sequence was found exactly matching the 24 multiples found that converge to {4,2,1}.

$$4, 9, 12, 18, 20, 25, 28, 36, 44, 45, 49, 50, 52, 60, 63, 68, 75, 76, 84, 90, 92, 98, 99, 100$$

This integer sequence, A067259, was also found to correspond to the 24 initial numbers that generated a sequence convergent to {4,2,1} for $m = n - 1$. A067259 are all numbers that a cubefree but not squarefree. Thus, all numbers that are divisible by a prime squared and are cubefree, when set to n, will generate a sequence that converges to {4,2,1}, as long as $m = n - 1$.

The integer sequence we found is defined by being divisible exactly by the square of a prime, and is cubefree. Thus, the first 24 numbers that are divisible by the square of a prime number and are cubefree have all been found by our project to be multiples for the Generalized Collatz Sequence, that when the initial number is set to one less than it, will converge to the Classic {4, 2, 1} Collatz cycle. This novel finding means that for any multiple inputted into the Generalized Collatz Function as long as it is one less than the initial number which is divisible by the square of a prime and cubefree, it will converge to the Classic {4, 2, 1} sequence. In number form, this means:

If $m = n - 1$ in the Generalized Collatz Function, when $n = k * p^2$, where k is a constant and p is a prime number, and n is cubefree, then the hailstone sequence generated by those parameters will converge to the {4, 2, 1} cycle.

This newly discovered pattern suggests that there is a proof behind both the Generalized Collatz and the Classic Collatz Conjecture. We are currently working on the correlation on why a constant times a prime number squared, when cubefree, creates an initial number that converges to the {4, 2, 1}.

To further generalize, and possibly to find a larger pattern, all sequences that converged to 1 were found. 17817 of these were found out of the 1 million sequences. That was just about 4 times the Collatz sequence, meaning the Collatz cycle of {4, 2, 1} made up about 25% of all cycles converging to 1. To see any correlation, the $\{n, d, m\}$ trios that made the sequence converge to one were plotted in Figure 3.

As seen in Figure 3, the diagonal line seen before makes itself clear. With no holes or gaps, a solid line along the equation $m = n - 1$ is found. Along this line, for any d, the $\{n. m, d\}$ trio would converge to 1. From this observation, a prediction can be made. Any $\{n, m, d\}$ trio in the form$\{a - 1, a, b\}$, where a and b are any positive number, will converge to 1. In other words, any generalized Collatz sequence formed by a trio with equal multiple and initial number will converge to 1.

The reason for this pattern was analyzed by viewing the lengths of the calculations needed to converge to 1. In other words, the number of steps needed to get down to one from the initial number was calculated. Interestingly, the number of steps needed to get down to 1 for the majority of these hailstone sequences was 3. This means that in one division and one multiplication, in any order, all initial numbers with a multiple one less than it converged down to 1. In number form, the hailstone sequence with $n = a - 1, m = a,$ and $d = b,$ where a and b are positive integers, will converge to 1 in exactly 3 steps. This three step rule also worked for initial numbers that were prime. This observation suggests that even prime initial numbers are multiplied, divided, and manipulated in a way that will result in convergence to 1.

A plot containing all $m - d$ pairs that cycled for all n was found, shown in Figure 4.

Vertical lines are clearly seen in Figure 4, leading to the speculation that there are multiples, for any and all divisors, that will result in a cycle. By analyzing the vertical lines and finding those that went all the way from 1 to 100, the multiples that, for all divisors, had cycled were found. Those multiples were 1, 3, 5, 7, 15, 31, and 63. After closer observation of these numbers, with the exclusion of 5, a pattern was discovered. The numbers 1, 3, 7, 15, 31, and 63 form a pattern following the explicit equation: $2^k - 1$. All numbers in this sequence, as far as the data so as the data so far provides, will be a multiple that, for any divisor, will cycle. With this equation, another key prediction can be made: that 127 and all other multiples that are $2^k - 1$ will, for any divisor and any initial number, cycle.

We immediately saw the connection to Mersenne Numbers in this pattern. Mersenne numbers are numbers that take the form $2^k - 1$, exactly the form of multiple needed for the generated hailstone sequence to cycle. This means that if the multiple in a Generalized Collatz sequence is a Mersenne number, the hailstone sequence will cycle, no matter the divisor or initial number used. The reason for which all Mersenne numbers, when set to a multiple, creates a cycling Generalized Collatz function is still being investigated.

These three conclusions are incredibly new and unique in the field of the Collatz. Bruschi says in his 2008 paper that he had to resort to a massive computer investigation that was very time-expensive and proved nothing. [11] Bruschi was unable to find any patterns or conclusions of his own in his spar with the Generalized Collatz. However, the current report contradicts those claims. We have found three novel patterns and conclusions in our work, and we have used large computer investigations and big data to find them. These conclusions have never before been seen, by even the strongest analyses. Messerman, LeBeau and Klyve say in their 2012 paper that they were unable to find any patterns in the end results of the Generalized Collatz, a feat this paper was able to accomplish.

Conclusion and Future Work

This project, designed to create a visual representation of generalized hailstone sequences, and to use those visuals to help find patterns in those sequences, did exactly that. By combining 3D graphing technology with strong analysis tools, three patterns were observed.

One, that all numbers that are divisible by a prime number squared are also initial numbers, that with a multiple that is one less than it and the initial number is cubefree, creates a hailstone sequence ending in {4, 2, 1}. The reason for why this happens is still being investigated, with the On-line Encyclopedia of Integer Sequences and possible induction for a proof. We plan to apply basic algebra and further data analysis to determine, mathematically, as to why this pattern occurs. We are also striving to connect this pattern back to the Classic Collatz Conjecture to help build a proof for it.

Two, that all generalized Collatz sequences with a multiple equal to the initial number minus one will converge to 1, for any divisor, in 3 steps. This pattern is being investigated further to see why this pattern occurs, especially with only one multiplication and one division involved. Further data analysis into the three step rule we found will be done to understand and prove as to why this pattern takes place, and to connect it back to a proof of the Classic Collatz.

Three, that all Mersenne numbers, when set to a multiple, will, for any divisor and for any initial number, cycle. This pattern is being investigated by looking into the properties of Mersenne numbers and Mersenne primes and how they might relate to the Generalized Collatz. Algebra and further data analysis will be used to connect Mersenne numbers to the Generalized Collatz and to connect them back to the Classic Collatz Conjecture.

Messerman, LeBeau and Klyve say in their 2012 paper that if patterns are able to be found in the Generalized Collatz Function, it is highly likely that the original Classic Collatz Conjecture is not special, but rather that it's behavior follows from more general principals. [12] This report has been able to find multiple patterns in the Generalized Collatz Function, and future work will be done to analyze the behavior of the Classic Collatz Function, possibly leading up to a proof.

This project, with the help of 3D graphing technology, was able to find never before seen patterns in a thought-to-be unpredictable function, the Generalized Collatz.

Literature Cited

[1] Wikipedia. (2016, January 21). *Collatz Conjecture*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Collatz_conjecture

[2] University of St Andrews, Scotland. (2006, November). *Lothar Collatz*. Retrieved from School of Mathematics and Statistics , University of St. Andrews: http://www-groups.dcs.st-and.ac.uk/history/Biographies/Collatz.html

[3] Plus Math Magazine. (2014, December 9). *Mathematical Mysteries: Hailstone Sequences*. Retrieved from Plus Magazine: https://plus.maths.org/content/mathematical-mysteries-hailstone-sequences

[4] Opfer, G. (2011, May). An analytic approach to the Colaltz 3n+1 Problem. *Hamburger Beitrage zur Angewandten Mathematik*.

[5] North Dakota State University. (2016, January 27). *Gerhard Opfer*. Retrieved from Mathematics Genealogy Project: http://genealogy.math.ndsu.nodak.edu/id.php?id=27958

[6] Yorgey, B. (2011, June 4). *The Collatz Conjecture is safe (for now)*. Retrieved from The Math Less Traveled: http://mathlesstraveled.com/2011/06/04/the-collatz-conjecture-is-safe-for-now/

[7] Berg, L., & Meinardus, G. (1995). The 3n+1 Collatz Problem and Functional Equations.

[8] Cope, D. (2011, June 18). *What was the problem with Gerhard Opfer's 2011 proof of the Collatz Conjecture?* Retrieved from Quora: https://www.quora.com/What-was-the-problem-with-Gerhard-Opfers-2011-proof-of-the-Collatz-Conjecture

[9] Zhang Zhongfu and Yang Shiming. Ying She Shu Lie Wen Ti. Shu Xue Chuan Bo 22 (1998), no. 2, 76–88.

[10] John Lesieutre. On a Generalization of the Collatz Conjecture, Research Science Institute Massachusetts Institute of Technology (2007), 4-13. Retrieved from: http://web.mit.edu/rsi/www/pdfs/papers/2004/2004-lesjohn.pdf

[11] Mario Bruschi, A generalization of the Collatz problem and conjecture, **arXiv:0810.5169** v1 29 Oct 2008. Retrieved from: https://arxiv.org/pdf/0810.5169v1.pdf

[12] Messerman, Hayden R; LeBeau, Joey; and Klyve, Dominic (2012) Generalized Collatz Functions: Cycle Length and Statistics, *International Journal of Undergraduate Research and Creative Activities:* Vol. 4, Article 2. DOI: http://dx.doi.org/10.7710/2168-0620.1002

[13] N. J. A. Sloane, editor, *The On-Line Encyclopedia of Integer Sequences*, published electronically at https://oeis.org, 22 July 2012.