A Novel Approach to Collatz Conjecture Proof: Effect of Addition on Prime Factorization and Unique Numeric Potential Concept

Introduction

The Collatz Conjecture

The Collatz conjecture is an unproven conjecture in mathematics. It states that any initial input to the Collatz sequence will give an eventual output of one. The Collatz sequence is a mathematical sequence that takes an initial input and does one of two things to that input; if the input is even, divide by 2 to get the output, if the number is odd, multiply by 3 and add 1. Consider:

56 is even;
56/2=28;
Output = 28
21 is odd;
21*3+1=64;
Output = 64

The output will be used as the next input for the function to run again, until the output reaches 1. The function continues until the output is 1.

The Collatz sequence function can be expressed as:

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

Even numbers result in an input to divided by 2. This has the effect of shortening the stopping time. It is important to note, for every two as an input's prime factor, its Collatz sequence will approach 1 faster. Odd inputs always produce an even output. However, the amount of twos in the prime factorization of the next output cannot be predicted. This is why most research concerning Collatz Conjecture focuses on odd inputs.

Impact

The importance of proving (or disproving) the Collatz sequence is due to the mathematical fields it crosses. Research on the conjecture has been considered in number theory, ergodic theory and dynamical systems, computational theory, and probability theory. The problem seems very simple, utilizing nothing more than positive integers and elementary operations. However, it evolves into a complex dynamical system. John Conway questions if the problem can even be solved through computation in his paper "Two Undecidable Variants of Collatz's Problems".

Past research

Past work has considered both the theoretical and probable side of the Collatz Conjecture. This includes Conway's paper above. Other work focuses on parallels and simplifications of the problem. These include the "qx+1" problem (Steiner), and the "3x+d" problem (Belaga and Mignotte). These problems were solved and showed that statements similar to the Collatz conjecture could be proven.

Objectives

This research focused on the experimental side of the conjecture by looking at shared traits by numbers linked in the sequence.

There are three objectives to this research:
•Identify correlations between prime factorization of all inputs to the Collatz sequence that results in outputs with significantly shorter stopping times
•Determine the significance in patterning a Collatz numeral based on its Collatz potential
•Identify pathways to proving or disproving the Collatz conjecture based on the previous two objectives

Methodology

The methodology identifies the prime factors for any given even-numbered Collatz input, specifically for that set of numbers containing multiple 2's, (as the stopping times are functionally the shortest) and from which one is subtracted (i.e. to make odd-numbered) The total number of specific prime factors in the odd-numbered input is compared, as a ratio, to the total number of specific prime factors in all numbers. Then "potential" is developed based on the information gained from the experiment. This will help find new ways of examining the conjecture and potentially proving or disproving it.

Background

It is unclear why numbers containing prime factors of multiple 2s are important in the Collatz sequence. In the sequence, an even number is divided by two, making it less probable that the sequence will diverge to infinity. The resulting output is closer to the desired sequence output of one and farther from infinity.

After every odd sequence step, there must be an even step. However, if there is only one even step, the resulting final output is larger than the original odd input:

Let 3 be the input
c(3)=10
c(10)=5
5>3

If there is more than one even step, the resulting final output is smaller than the original odd input:

Let 9 be the input
c(9)=28
c(28)=14
c(14)=7
7<9

To produce more than one even step, there must be more than 1 two in the prime factorization, hence, the number must produce multiples of four as factors. The overarching question becomes, in the Collatz sequence, what kind of numbers, specifically odd numbers (note that even numbers producing these odd numbers have a very similar prime factorization), produce multiples of four?

The input sequence of an odd number is multiplied by three, having essentially an insignificant effect on the prime factorization of the new number. However, once one is added, the sequence results in a potentially significant effect on the prime factors of the new input. This research aims to find the effect of the prime factorization of a number, specifically after the addition of one, on the Collatz Conjecture, and further, identify the numeric relationships, specifically within the prime factorizations between the inputs and resulting outputs of the Collatz sequencing resulting in new methods of analyzing the Collatz Conjecture.

Hypothesis

Consider a ratio of the following:

•totaled individual prime factors of numbers which, when inputted, lead to shorter stopping times of the Collatz sequence, (f_s) to,
•the totaled individual prime factors of all numbers inputted, (f_t)

$$r_f = \frac{f_s}{f_t}$$

Also consider the following ratio:

•the amount of numbers that shorten the stopping time, (n_s) to,
•all numbers inputted, (n_t)

$$r_n = \frac{n_s}{n_t}$$

Based on the bulleted objectives above, the research hypothesizes the ratios will not be equal, ≠

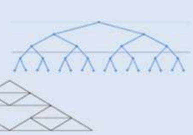


Figure 1-The resemblance of graph 3 to Sierpinski's triangle. On the right is a screenshot of graph 3 with lines drawn on top of it. On the bottom is Sierpinski's triangle. There is a clear resemblance, more information is in the results section.

Methods

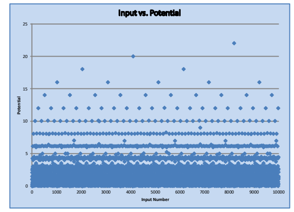
Investigate the prime factorizations of certain numbers. A program will be written in python (see buddy board) to find the prime factorization of over a million numbers. This program will also export the prime factorizations to a text file. Because of the syntax of the program, the text file will have to be edited to make the delimiters consistent. Delimiters are what separate the columns in a text file and will come into play later. To make this edit, a simple find and replace will suffice. Then the text file will be imported into an excel spreadsheet. Here is where delimiters are going to be important. I will tell excel that the characters used as delimiters should not be included in the data and define where the column dividers go. Then in the spreadsheet, numbers with more than one two in the prime factorization will be identified. Then one will be subtracted from these numbers and the resulting numbers will be put in a group, "desired". The prime factors will be tallied for both the "desired" numbers and all of the numbers. Then I will look for similarities and differences in the tallies.

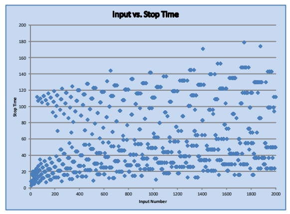
Compare trends and patterning of prime factorizations of 3x term in odd-numbered Collatz input by subtracting 1 from previous even-numbered output. Create a ratio between the 3x terms and all of the inputs for each specific prime factor. Then compare this ratio to the ratio of 3x terms to all inputs. If the ratios are different, then the 3x terms had a significant effect on the prime factorization.

Devise a method of calculating "Collatz Potential". Potential will aim to represent the ability of an input to get to one in the Collatz sequence. Then "potential" will be plotted against stopping time of an input, and the input itself. Stopping time will also be plotted against the input. Then these graphs will be analyzed for conclusions.

Compare and analyze Collatz Potential against stop time and the Collatz input. Make graphs comparing potential against input, and potential against stop time. Also graph stop time against input. Then determine and classify the shape of the graph. Determine the cause and meaning of this shape.

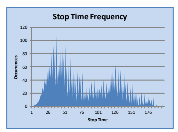
Graph 3-Input vs. Potential-The relationship between input and potential. This graph forms the shape of Sierpinski's triangle. This is shown in figure 1. The triangle is formed from potential's use of 2s in prime factorization. The exact rows occur because of the tendency of numbers with higher potentials to have a similar potential with other numbers.





Graph 5-The relationship between potential and stopping time. The large shape of the graph is an exponential curve. However, there is a "horn" sticking out. The horn seems to be centered around 128. This is a common stopping time for which a number with a high potential is in. This raises the potential for all the numbers that hit that number in the Collatz sequence., usually the ones around it. The vertical bars can once again be explained by the tendency for higher numbers to have a similar potential.

Graph 6-The amount of times a certain number is a stopping time. There are 2 major peaks, at 39 and 122. These are right above the radiating points on graph 4.



Proofs

let x = the first node

$$\begin{cases} \text{Branch is 1 long} & \text{If } x \bmod 3 = 2 \\ \text{Branch is 2 long} & \text{If } x \bmod 3 = 1 \\ \text{Branch is infinitely long} & \text{If } x \bmod 3 = 0 \end{cases} \quad [1]$$

Proof 1:

let x mod 3 = 2
let m, n, and p be positive integers
x can equal 3n + 2 because 3n mod 3 will always be zero because for any integer n, it will be multiplied by three before the mod 3 function, 2 mod 3 = 2. So, (3n + 2) mod 3 = 2 because mods can be added.
The next step in the inverse Collatz sequence is to multiply by two because x ≠ 3m + 1 because 3m + 1 mod 3 = 1 ≠ x mod 3 = 2.
after this step, x = 6n + 4 because 2(3n + 2) = 6n + 4.
x can equal 3(2n + 1) + 1 because 6n = 3(2n) and 3 can be subtracted from the four and substituted into the parenthesis giving 3(2n + 1) + 1
x can equal 3p + 1 because 2n + 1 is still a positive integer and can be replaced by p
since x can equal 3p + 1, the next inverse Collatz sequence step could either be (x-1)/3 or x*2 therefore, the branch splits and there was only one even step in between.

Proof 2:

let x mod 3 = 1
let m, n, and p be positive integers
x can equal 3n + 1 because 3n mod 3 will always be zero because for any integer n, it will be multiplied by three before the mod 3 function, 1 mod 3 = 1. So, (3n + 1) mod 3 = 1 because mods can be added.
After a split an even number is produced by multiplying by two, and an odd number is produced by subtracting one and dividing by 3. This is because in the regular Collatz sequence, only odd numbers are multiplied by three and added to one. The even number can never be 3n + 1 because after the split number is multiplied by two, it produces a 6m + 2 number because split numbers are always 3m + 1 and 2(3m + 1) = 6m + 2. Therefore because x = 3n + 1, x must be odd.
Because x is odd, it cannot be a branch number because all branch numbers must be even. They must produce an odd number from the (x-1)/3 operation because in the regular Collatz sequence, only odd numbers do the inverse operation of that one – 3x + 1.
The next step in the inverse Collatz sequence is to multiply by two by process of elimination
after this step, x = 6n + 2 because 2(3n + 1) = 6n + 2.
x mod 3 = 2 because (6n + 2) mod 3 = 2 because 6n mod 3 = 0 and 2 mod 3 = 2 and 0 + 2 = 2
The next step in the inverse Collatz sequence is to multiply by two because x ≠ 3m + 1 because 3m + 1 mod 3 = 1 ≠ x mod 3 = 2.
after this step, x = 12n + 4 because 2(6n + 2) = 12n + 4.
x can equal 3(4n + 1) + 1 because 12n = 3(4n) and 3 can subtracted from the four and substituted into the parenthesis giving 3(4n + 1) + 1
x can equal 3p + 1 because 2n + 1 is still a positive integer and can be replaced by p
since x can equal 3p + 1, the next inverse Collatz sequence step could either be (x-1)/3 or x*2 therefore, the branch splits and there were two even steps in between.

Proof 3:

let x mod 3 = 0
let m and n be positive integers.
x can be thought of as 3n because 3n mod 3 = 0
The next step in the inverse Collatz sequence is to multiply by two because x ≠ 3m + 1 because 3m + 1 mod 3 = 1 ≠ x mod 3 = 0.
after this step x = 6n
The next step in the inverse Collatz sequence is to multiply by two because x ≠ 3m + 1 because 3m + 1 mod 3 = 1 ≠ x mod 3 = 0.
after this it becomes obvious that the cycle will repeat because x mod 3 will remain zero and be infinitely multiplied by 2

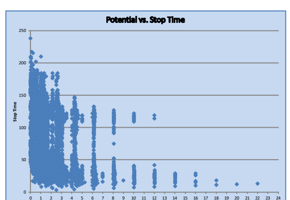
Results

•The pictorial representation of the relationship between potential and input is many horizontal levels in the shape of Sierpinski's triangle; a fractal.

•The relationship between potential and stop time is inverse, nonlinear, and has a large inconsistency. This inconsistency manifested as a protrusion from the expected curve of the graph (Graph 5), evidenced as a number with lower stopping time but an irregularly large potential. Consequently, the potential of the numbers around it also rose because of this number; reflected by a horn shape on the same graph.

•The functional relationship between input and stop time indicates curvature radiating from two points (Graph 4) and is attributable to the numerous square-root functions embedded in the data.

•The research utilized numbers between 2 and 1,048,572 which is the limit on an excel workbook. The summation of factor totals for numbers one less than multiples of four were 1/4 of their counterparts for all integers except for 2. 2 had none because all the numbers were odd.



Graph 4-The relationship between input and stop time. There are multiple square root functions radiating from 2 points on the y-axis. This is caused by long branches of even steps in the Collatz sequence. On the long branches, the input is cut in half every step, and the stop time goes down by 1. This is what causes the square root curves. As far as the research found, the radiating points are just common stopping times as seen on the frequency graph, graph 6.

Developing Collatz Potential

Early in this research, it became evident that a new mathematical function was necessary to better represent the patterns that were emerging, so the potential idea was developed.

Collatz potential is a function of the number of twos in an input's prime factorization and the potential of the next even input in the Collatz sequence. Collatz potential is equal to the number of twos in an input's prime factorization minus one plus half of the potential of the "next input" in the sequence. Potential was specifically designed for even numbers because they lead directly to the input approaching 1. This means that "next input" really means the next ***even*** input.

Potential equation form:

$$p_c = n_c - 1 + \frac{p_{(c+1)}}{2}$$

The number of twos in the prime factorization represents how quickly the input will approach 1. Every number will be even and have at least one 2, so one is subtracted to only count the twos that evens don't necessarily have in common. The second term of the potential calculation, $\frac{p_{(c+1)}}{2}$, incorporates the potentials of the next inputs. Inputs before those with high potentials generally also have high potentials.

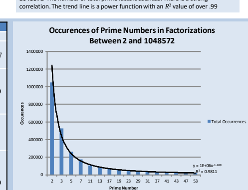
This is true for low potentials also. Potential loosely represents the stop time of a Collatz input. Stop time is the number of steps in the Collatz sequence it takes for an input to reach 1. Collatz potential is a transition step between input and stop time to find a solid relationship between 'potential and them both. Potential's final goal is breaking a way for a function able to calculate something correlating to the stop time of an input without calculating the whole Collatz sequence for that input.

If every number has a potential greater than zero (besides. 2 and one), the Collatz conjecture is proven because of the way the potential equation is designed. Even if potential can't be calculated for every number without going through its Collatz sequence, it could lead to a function that can do this.

Data Table 1-Overall, there was no correlation between all the numbers one less than numbers with more than one 2 in their prime factorization. The ratio of .25 at the bottom is showing that each of the tallies is one quarter of the total tally of the specific prime factor. This excludes 2 which was counted 0 times because all the numbers one less than numbers with more than one 2 in their prime factorization were odd.

Total Occurrence of Prime Factors	2	3	5	7	11	13	17	19	23	29	31	37	41	43	47	53
All Numbers between 2 and 1048572	1048556	524283	262143	174766	104864	87391	65550	58271	47684	37475	34982	29161	26253	25008	22841	20217
Ratio to Total Numbers (Three Decimal Places)	1.000	0.500	0.250	0.167	0.100	0.083	0.063	0.056	0.045	0.036	0.033	0.028	0.025	0.024	0.022	0.019
All Numbers One Less Than a Multiple of 4	0	131071	65533	43692	26214	21844	16383	14563	11916	9360	8738	7281	6553	6243	5699	5041
Ratio to Occurrence in all numbers (Three Decimal Places)	0.000	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.250	0.249

Graph 1-Occurrences of Prime Numbers in Factorizations between 2 and 1048572- The number of total prime factors counted. There is a strong correlation. The trend line is a power function with an R^2 value of over .99



Discussion

I developed the "potential" concept because I was inspired to create a new method of analyzing the Collatz Conjecture when I was doing my original experiment. It is meant to help identify underlying patterns in an input.

Although the Potential calculations were highly successful, my hypothesis was not supported. The stated ratios were equal. They both equaled ¼.

The Potential calculations of my experiment went very well and produced many interesting patterns. After calculating potential, some patterns emerged. Potential was compared with two other data points - the Collatz input, and the stop time for that input. When potential was compared to input, graph 3 occurred. It is interesting to note, the points are in the shape of Sierpinski's triangle, as shown in figure 1. Sierpinski's triangle is a common fractal created by repeatedly removing increasingly smaller triangles from one large one. The fact that there is such a huge link from the Collatz Conjecture to another field of mathematics means that there is another chance to prove it or to find a pattern. The potential of a given input is directly proportional to the number of 2s in the prime factorization of a number, i.e., n_c.

$$p_c = n_c - 1 + \frac{p_{(c+1)}}{2} \quad [1]$$

This "bubbles up" the numbers with more twos as prime factors to the top of the triangle. Numbers with one less two in the prime factorization will have approximately half as much potential, leading to perfect spacing in the x direction. Why are potentials similar in the y direction? As numbers get larger, their potential, which is directly proportional to the number of 2s in their prime factorization, renders the remaining terms with less significance. This means that numbers with the same amount of twos in their prime factorization will have more closely aligned potential.

As potential is plotted against stop time, another pattern emerges. Depicted in graph 5, it is nonetheless difficult to describe. The graph looks like a standard negative exponential curve, with a large horn situated in the middle. A gap is evident between the lower values and the horn. The gap spans from about 50 to 100 in the y direction. With lower value inputs in Collatz sequence, it is much easier to get to one simply because these values are closer to one. The potential is high for these numbers simply because ½ the potential of the next number term is a more significant term in lower values (because as n_c approaches 1, the term n_c-1 is not the significant term . When you get a little bit higher though, the number of twos in the prime factorization matters a lot more than the potential of the next number. This is why the potential drops here. The potential comes back up in a jut out further up. This part of the graph seems centered around the number 128 which has a high potential.

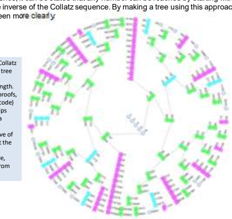
$$p_{128} = 10.03125$$

This causes the numbers around it to have a higher potential, but not as high as the potential of values at the lower end.

Stop time was plotted against input. This produced graph 4. The shape depicts many curves radiating from two points on the y-axis. This graph was hard to explain by itself, so a stop-time frequency graph was created. Graph 6 shows two major peaks at 39 and 122 stop time (stop time is how many steps a specific input in the Collatz sequence takes to output 1). These values were slightly above the two radiating points. From the frequency graph, it seemed like these two points were just popular. However, the curve patterns can be explained from related numbers in the Collatz sequence. Consider a number which can be divided by 2 in the Collatz sequence many times. The output (next input) has a stop time one less than the previous and the graphed value would be half of the previous. As a looped input/output this creates a square-root graph. These are the curve patterns on the graph.

Throughout this research, some patterns have become clear. First, it is easier to look at the Collatz Conjecture differently. Rather than stating that any number will get to one through the sequence, it can be stated that any number can be reached by starting with one and doing the inverse of the Collatz sequence. By making a tree using this approach, patterns can be seen more clearly

Figure 2-Collatz Tree. The tree visualizes branch length. (refer to proofs, for color code) It also helps visualize a different perspective of looking at the Collatz conjecture, starting from one.



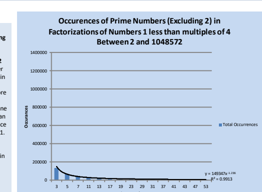
There are big main branches. This happens on big chains of even numbers, multiplying by two from the center. Also large branches that don't branch off at all are multiples of three because they can never evenly subtract one and divide by three. There are underlying patterns from the analyses of this research. Branches after the very first can be one of three set lengths before splitting. Branches can be 1, 2, or infinite even steps long, until a split. After looking for the cause of this, it was found. Prime factorization does play a part in it. Branches can be classified by the very first node on that branch after a split leading to that branch (see proofs).

The Potential research could also help prove the Conjecture. The patterns found could very well help find relevant values for inputs in the Collatz sequence without even applying the function on said inputs. If an input has a potential, then it must reach one because of the way Collatz Potential is defined. If a potential can be calculated for all numbers, then they all eventually get to one in the Collatz sequence.

There were also patterns found between stopping time and input. The graph of this can be used as a tree also. If the pattern can be isolated, the conjecture could be very easily proven or disproven. All of these patterns could help contribute to further Collatz research, and eventually a proof.

Potential was not the original goal of this experiment. At first I was focusing entirely on prime factorization. However, working with prime factorization inspired me to develop a new way of working with the conjecture. I had to totally reprogram myself to look at the conjecture in a different way as shown in figure 2.

Graph 2-Occurences of Prime Numbers (excluding 2) in Factorizations of Numbers 1 less than Multiples of 4 Between 2 and 1048572-The number of prime factors counted in numbers one less than numbers that contain more than one 2 in their prime factorization. The trend line is a power function with an R^2 value of over .99. Notice how similar it is to graph 1. If it is looked at closely, it would be seen that each value is a quarter of that in graph 1. The prime 2 is excluded because all the numbers counted were odd.



References

"Collatz Conjecture Analysis (But No Proof; Sorry)." *Collatz Conjecture Analysis (But No Proof; Sorry)*. N.p., n.d. Web. 06 Feb. 2017.

"Collatz Problem." *Collatz Problem -- from Wolfram MathWorld*. N.p., n.d. Web. 06 Feb. 2017.

Davies, Jason. "Collatz Graph: All Numbers Lead to One." *Collatz Graph: All Numbers Lead to One - Jason Davies*. N.p., n.d. Web. 06 Feb. 2017.

"Visualizing Pascal's Triangle Remainders." *The Math Less Traveled*. N.p., 20 Oct. 2012. Web. 06 Feb. 2017.

"Mathematics Stack Exchange." Mathematics Stack Exchange. N.p., n.d. Web. 24 Mar. 2017.

"Two Undecidable Variants of Collatz's Problems." Two Undecidable Variants of Collatz's Problems - ScienceDirect. N.p., n.d. Web. 24 Mar. 2017.

Hammett, Matthew. Collatz Conjecture: A Brief Overview. N.p., n.d. Web. 24 Mar. 2017.

Figure 1 credit oftenpaper.net, Figure 2 credit Jason Davies, All other figures and graphs are credit the presenter.