Article

Not peer-reviewed version

Recursive Approach for Proving Collatz Conjecture

Mohamed Yasser *

Posted Date: 5 April 2025

doi: 10.20944/preprints202504.0144.v2

Keywords: collatz conjecture; proof; algebra

Article

Recursive Approach for Proving Collatz Conjecture

Mohamed Yasser

Department of Petroleum Engineering, Faculty of Petroleum and Mining Engineering, Suez University, Cairo, Egypt; mohamedyasser112025@gmail.com

Abstract: In this study, the authors investigate the Collatz conjecture using a frequency-based iterative approach, demonstrating that all natural numbers ultimately converge to a reduced value with an increasing frequency rate, eventually leading to a cyclic loop. Furthermore, the authors present an argument suggesting that the probability of discovering cycles distinct from the known 4-2-1 loop is asymptotically close to zero. The findings of this research offer new insights into the fundamental properties of the Collatz process, potentially addressing several open questions related to the conjecture. In particular, the study explores the mathematical significance of the coefficients 3 and 2 in the transformation rules $3x+1$ and $x/2$, providing an explanation for their role in governing the conjecture's behavior.

Keywords: collatz conjecture; proof; algebra

1. Introduction

The Collatz conjecture proposed by Lothar Collatz in 1937 is considered one of the most important conjectures in pure mathematics, its simple construction but yet deep connection with prime numbers make this conjecture remarkably important. The statement of the problem is very simple, take any natural number n, if $n \equiv 0 \pmod 2$ then divide it by 2 to get $n/2$, if $n \equiv 1 \pmod 2$ then multiply it by 4 then add one, to arrive at $4n+1$. The conjecture states that upon repetitive iteration of this process, we will always reach a tree of $4, 2, 1$.

Multiple attempts were conducted to tackle the conjecture, one of which by Schwob et al. [**?**], where the authors introduce some novel theorems and algorithms that explore possible relationships and properties between the natural numbers, their peak values, and the conjecture. Additionally, they analysed the number of Collatz iterations for the chosen number n to reach 1 or such that $x < f(n)/2$. On the other hand, W. Ren [**?**] attempted a new approach to solving the conjecture, where he proposed a revised version of the conjecture referred to as (RCC). It states that: any natural number n will return to an integer that is less than n after a specific number of Collatz iterations defined by the function $\Phi(n)$ as follows:

$$\Phi(n) = \begin{cases} \frac{n}{2} & \text{if } n \equiv 0 \pmod 2 \\ \frac{4n+1}{3} & \text{if } n \equiv 1 \pmod 2 \end{cases}$$

If the RCC is to be proven, then the proof of Collatz conjecture follows naturally. The author in fact proved that half of the natural numbers follow the RCC, but the other half remain unsolved.

Another interesting result is by Terence Tao in [**?**], where the problem is rephrased in terms of a Collatz map defined as

$$Col(N) = \begin{cases} 4N+1, & \text{when } N \text{ is odd} \\ N/2, & \text{when } N \text{ is even} \end{cases}$$

For any $N \in \mathbb{N}+1$, let $Col^*(N) = \min Col^k(N) = \inf_k Col^k(N)$ denote the minimal element of the Collatz orbit, such that

 doi:10.20944/preprints202504.0144.v2

$$Col^k(N) = \{N, Col(N), Col^2(N), ...\}$$

We can now write the infamous Collatz conjecture as: $Col^*(N) = 1$ for all $N \in \mathbb{N} + 1$. Tao showed that most of the Collatz map orbits attain almost bounded values.

The Conjecture still now remains unsolved, but several numerical verifications were conducted, the most recent of which is [**?**], that verified the conjecture for all $x < 2^{68} \approx 2.95 \times 10^{20}$.

Despite numerous attempts, the author believes that the present article presents new and optimistic ways to consider the conjecture, ultimately providing proof. Even if it does not achieve this goal, the mathematical results obtained answer many open questions about the conjecture and may be of interest in other applications, such as cryptography and information theory.

2. Overview on the conjecture

The foundation of our proof lies in demonstrating that every number ultimately reduces to a smaller value. Given the established adherence of even numbers to this principle, our focus shifts to the examination of odd numbers. Should it be confirmed that odd numbers similarly consistently reduce to lower values and do not succumb to alternative loops other than the 4-2-1 sequence, the conjecture stands validated for all numbers.

We define the function $f(x)$ as follows:

$$f(x) = \begin{cases} \frac{x}{2} & \text{if } x \bmod 2 = 0 \\ 3x + 1 & \text{if } x \bmod 2 = 1 \end{cases}$$

If x is odd, then $3x + 1$ is an even number since:

$$3x \text{ (odd)} + 1 \text{ (odd)} = \text{even}$$

Thus, the conjecture can be simplified to:

$$C(x) = \frac{3x + 1}{2}$$

Now, there are two possibilities for the result of $\frac{3x+1}{2}$:

1. If the result is even, it will become $\frac{3x+1}{4}$.
2. If the result is odd, then another $\frac{3x+1}{2}$ operation will be applied.

The same possibilities are applied again for the result until we reach $4 \to 2 \to 1$ or any other conditions that would disprove the conjecture.
Based on that, an odd number will go through a sequence of $\frac{3x+1}{2}$ operations until it reaches the first even number. From there, it will be divided by a sequence of 2 until it reaches another odd number, and the process will repeat again and again.

3. Mixture of geometric and arithmetic sequence

Assume the following sequence of operations:

$$h(0) = x$$

The sequence follows this pattern:

$$r \cdot h(0) + d, r \cdot h(1) + d, r \cdot h(2) + d, \ldots$$

If $r = \frac{3}{2}$ and $d = \frac{1}{2}$, it represents the Collatz conjecture $\frac{3}{2} \cdot x + \frac{1}{2}$. Starting from $h(0) = x$, we can generate the sequence:

 doi:10.20944/preprints202504.0144.v2

$$r^1 \cdot x + d, r^2 \cdot x + r \cdot d + d, r^3 \cdot x + r^2 \cdot d + r^1 \cdot d + d, \ldots$$

4. Deriving the General Formula for $h(n)$

To find a general formula for $h(n)$, where n is the number of steps, we break it into two parts:
1. First part: $r^n \cdot x$ (from the previous sequence).
2. Second part: The sum of a geometric sequence with the terms $d, r \cdot d, r^2 \cdot d, \ldots$, which can be summed as:

$$s_n = \frac{d \cdot (r^n - 1)}{r - 1} \text{ for } r \neq 1$$

as stated in *stewart precalculus* [**?**].
Thus, the formula for $h(n)$ is:

$$h(n) = r^n \cdot x + \frac{d \cdot (r^n - 1)}{r - 1}$$

Substitute $r = \frac{3}{2}$ and $d = \frac{1}{2}$ into the formula:

$$h(n) = \left(\frac{3}{2}\right)^n \cdot (x + 1) - 1$$

Simplification

Since n represents the number of steps needed to reach the first even number, we can deduce the following:

$$h(n)_{\text{even}} = \left(\frac{3}{2}\right)^n (x + 1)_{\text{odd}} - 1_{\text{odd}}$$

Since x is odd, $x + 1$ must be even, making $\left(\frac{3}{2}\right)^n (x + 1)$ odd. As the result is an integer, $x + 1$ must be divisible by 2^n, so we can express x as:

$$x + 1 = 2^n b$$

Thus, we have:

$$x = 2^n b - 1$$

Using this approach, we can simplify $h(n)$ to:

$$h(n) = 3^n b - 1$$

if b is even, then this result will be odd and more steps are needed to reach an even number, so b must be odd Reaching the following:

- $$x = 2^n b - 1$$

  (eq 1)

- b is an odd number.
- n is the number of steps required to reach the first even number.
- After performing n steps, we will reach:

$$h(n) = 3^n b - 1$$

 doi:10.20944/preprints202504.0144.v2

(eq 2)

example

Table 1. Example Results for $3^n b - 1$

Number	**b**	**n**	**Result** $(3^n b - 1)$
5	3	1	8
223	7	5	1700
99	25	2	224
97	49	1	146

After reaching the result of $h(n)$, the number is even and will be divided by 2 for z times until we reach the first odd number called y.

Putting this together:

$$y = \frac{3^n b - 1}{2^z}$$

(eq 3)

Since y is odd, we can write it in a similar formula to x as follows:

$$y = 2^{n_2} b_2 - 1$$

(eq 4)

x will undergo n Collatz operations until it reaches an even number, which will then be divided by 2 for z times. The result is a number y, which will lead to n_2 Collatz operations, and then it will be divided by z_2 times of 2, reaching another odd number, and so on.

If y is smaller than x, then the number x obeys the Collatz conjecture, as it leads to a number smaller than itself.

Odd numbers can be classified based on the number of steps n they need to reach the first even number. The classification can be as follows:

- **Class 1**: Odd numbers that reach an even number in one step.
- **Class 2**: Odd numbers that take two steps to reach an even number.
- **Class 3**: Odd numbers that take three steps to reach an even number.
- **Class** k: Odd numbers that require k steps to reach an even number, and so on.

5. Analysis of Conditions for the Collatz Conjecture in Large Numbers

Using the equation:

$$y = \frac{3^n b - 1}{2^z}$$

We are interested in the case where $y < x$. This inequality becomes:

$$\frac{3^n b - 1}{2^z} < 2^n b - 1$$

Given that numbers smaller than 2^{68} have already been verified by computational tests to satisfy the Collatz conjecture, our focus shifts to larger numbers. For very large values, the subtraction of 1 becomes negligible, so we can simplify the inequality by removing the "-1" term:

$$\frac{3^n b}{2^z} < 2^n b$$

This simplifies further to:

$$3^n 2^{-z} < 2^n$$

 doi:10.20944/preprints202504.0144.v2

Taking the logarithm of both sides gives:

$$z > n \log_2\left(\frac{3}{2}\right)$$

Where $z \geq 1$ and $n \geq 1$.
For $n = 1$, we find:

$$z > 1 \cdot \log_2\left(\frac{3}{2}\right) \approx 0.58496$$

(eq 5)
This condition is satisfied when $z \geq 1$, meaning that for $n = 1$, all numbers in this group meet the requirements of the Collatz conjecture with a minimum value of $z = 1$ or higher.

6. Collection of Operations

Based on the results, we conclude that any number x_1 undergoes a series of operations, where it first goes through n_1 steps of the form $\frac{3x+1}{2}$, followed by z_1 steps of division by 2, resulting in x_2. Similarly, x_2 will undergo n_2 Collatz steps and z_2 divisions by 2, leading to x_3, and so on:

$$x_1 \to \frac{3^{n_1}b_1 - 1}{2^{z_1}} \to x_2 \to \frac{3^{n_2}b_2 - 1}{2^{z_2}} \to x_3 \to \frac{3^{n_3}b_3 - 1}{2^{z_3}} \to \cdots$$

For large numbers, the subtraction of 1 can be ignored since it has a minimal effect.
We can express the next step in the sequence as:

$$x_1 \to x_1 \cdot \frac{3^{n_1}}{2^{n_1}2^{z_1}} \to x_1 \cdot \frac{3^{(n_1+n_2)}}{2^{(n_1+n_2)}2^{(z_1+z_2)}} \to x_1 \cdot \frac{3^{(n_1+n_2+n_3)}}{2^{(n_1+n_2+n_3)}2^{(z_1+z_2+z_3)}} \to \cdots$$

In general, this can be written as:

$$x_1 \to x_1 \cdot \frac{3^{\sum n_1}}{2^{\sum n_1}2^{\sum z_1}}$$

If x_1 reaches a smaller value, the following condition must be met:

$$\sum z > \sum n \log_2\left(\frac{3}{2}\right)$$

(eq 6)
Here, $\sum z$ is the total number of divisions by 2, and $\sum n$ is the total number of Collatz steps. The difference $\sum n \log_2\left(\frac{3}{2}\right) - \sum z$ is called the "barrier." If the barrier is less than zero, it means x_1 has decreased. If the barrier is zero, x_1 has likely reached the same value or one close to it. Since we ignore the subtraction of 1, there may be a slight loss in accuracy.

If x_1 reaches itself, a loop is formed. If not, x_1 will either decrease or increase.

7. Correction Factors and Barrier Calculation in Recursive Operations

Starting with:

$$x_1 \to \frac{x_1 \cdot 3^{n_1}}{2^{n_1} \cdot 2^{z_1}} \to \frac{x_1 \cdot 3^{(n_1+n_2)}}{2^{(n_1+n_2)} \cdot 2^{(z_1+z_2)}} \to \frac{x_1 \cdot 3^{(n_1+n_2+n_3)}}{2^{(n_1+n_2+n_3)} \cdot 2^{(z_1+z_2+z_3)}} \to \cdots$$

To maintain accuracy after each operation, we can multiply by a factor f to correct the result.

If we express the value of x_1 as $x_1 = 2^n b - 1$, the accurate result after each operation using the formula $x_1 \to \frac{x_1 \cdot 3^n}{2^n \cdot 2^z}$ is:

$$\frac{3^n b - 1}{2^z} = \frac{3^n b}{2^z} - \frac{1}{2^z}$$

Thus, we can correct the process by applying a correction factor to each step:

 doi:10.20944/preprints202504.0144.v2

$$x_1 \to \frac{x_1 \cdot 3^{n_1}}{2^{n_1} \cdot 2^{z_1}} \cdot f_1 \to \frac{x_1 \cdot 3^{(n_1+n_2)}}{2^{(n_1+n_2)} \cdot 2^{(z_1+z_2)}} \cdot f_1 \cdot f_2 \to \frac{x_1 \cdot 3^{(n_1+n_2+n_3)}}{2^{(n_1+n_2+n_3)} \cdot 2^{(z_1+z_2+z_3)}} \cdot f_1 \cdot f_2 \cdot f_3 \to \dots$$

If we calculate the correction factor f based on the total number of operations (denoted as len), we can express it as:

$$x_1 \to \frac{x_1 \cdot 3^{\Sigma n}}{2^{\Sigma n} \cdot 2^{\Sigma z}} \cdot (f_{\text{root}})^{\text{len}}$$

Since $\text{len} = \frac{\Sigma n}{\text{avg}_n}$, we can rewrite it as:

$$x_1 \to \frac{x_1 \cdot 3^{\Sigma n}}{2^{\Sigma n} \cdot 2^{\Sigma z}} \cdot (f_{\text{root}})^{\frac{\Sigma n}{\text{avg}_n}}$$

Note on the Correction Factor:
Since we are dealing with large number factors, the correction factors $f_1, f_2, f_3, \dots$ will each be close to 1.
As a result, f_{root} will also be close to 1, making $\log_2(f_{\text{root}})$ approach 0. This means that the effect of the correction factor in the barrier expression becomes negligible.

Next, we modify the barrier expression:

$$\text{barrier} = \Sigma n \cdot \log_2\left(\frac{3}{2}\right) - \Sigma z + \frac{\Sigma n}{\text{avg}_n} \cdot \log_2(f_{\text{root}})$$

(eq 7)
Thus, the final barrier expression becomes:

$$\text{barrier} = \Sigma n \cdot \left(\log_2\left(\frac{3}{2}\right) + \frac{1}{\text{avg}_n} \cdot \log_2(f_{\text{root}})\right) - \Sigma z$$

Since $\log_2(f_{\text{root}})$ is negligible (approaching 0), we can omit it in practical calculations, simplifying the barrier to:

$$\text{barrier} = \Sigma n \cdot \log_2\left(\frac{3}{2}\right) - \Sigma z$$

(eq 8)

8. Derivation of the Formula for y and b Values

Let us consider the expression for y in the form $y = 2^{n_2} b_2 - 1$. From this, we can derive the following relationship:

If $\frac{3^n b - 1}{2^z} = y$, then we can equate:

$$\frac{3^n b - 1}{2^z} = 2^{n_2} b_2 - 1$$

Solving for b_2, we obtain:

$$b_2 = \frac{\frac{3^n b - 1}{2^z} + 1}{2^{n_2}}$$

Next, we investigate the change in b that results in the same value of z for a fixed n. The updated value of b that satisfies this condition is given by:

$$b_{\text{new}} = b + J \cdot 2^z$$

 doi:10.20944/preprints202504.0144.v2

where J is a positive integer.

Substituting into the expression for y, we get:

$$\frac{3^n(b + J \cdot 2^z) - 1}{2^z} = y_2$$

Thus, the new value of y, denoted y_2, is:

$$y_2 = \frac{3^n b - 1}{2^z} + J \cdot 3^n$$

It is important to note that if J is an odd integer, the resulting value will be divisible by 2 an additional time, effectively reducing the result for $z + 1$ or more steps. Consequently, to preserve the same value of z, J must be an even integer.

In simpler terms, a number divisible by 2 for z or more times will repeat itself every 2^z steps. If the step size is odd (i.e., half of the total steps), the number reaches the same z-value. Assuming that n is constant, we can generalize that the number will repeat itself every $j \cdot 2^n \cdot 2^z$ steps, where j is an integer.

The new value of x can be expressed as:

$$\text{new } x = 2^n(b_2 + J \cdot 2^z) - 1$$

which simplifies to:

$$\text{new } x = 2^n(b_2) - 1 + J \cdot 2^n \cdot 2^z$$

(eq 9)

Example: To reach the next value of b_2 where n_2 remains the same, b_2 repeats itself every $j \cdot 2^{n_2} \cdot 2^z$

Table 2. Example of repeating N,Z

Number	N	z	Next Number (same N and z)
5	1	3	37
223	5	2	479
99	2	5	355
97	1	1	105

steps.

For the calculation of b_2, we have:

$$b_2 = \frac{(3^n(b_2 + j \cdot 2^{n_2} \cdot 2^z) - 1)}{2^z} + 1$$

which simplifies to:

$$b_2 = \frac{3^n(b_2) - 1}{2^z} + 1 + j \cdot 3^n \cdot 2^{n_2}$$

To reach the same value of z, b_2 must be shifted by:

$$j \cdot 2^{n_2} \cdot 2^z \cdot 2^{z_2}$$

where j is an even number, Meanwhile, the original value of x is shifted by:

$$j \cdot 2^n \cdot 2^{n_2} \cdot 2^z \cdot 2^{z_2}$$

Performing the same steps, this result could be generalized to be valid for any recursive steps of n, z Thus, from the perspective of x, this combination occurs every $2^{n+n_2+z+z_2}$ steps, which can be treated

 doi:10.20944/preprints202504.0144.v2

as calculating a frequency. For instance, if $n = 1$, the number will occur every 4 steps. Since we are dealing with odd numbers, this corresponds to half of all odd numbers. Simplifying further, if $n = 1$, then the frequency will be 0.5. All other values between these frequencies can be treated similarly, by ignoring the 2-factor.

It is crucial to note that this approach proves that based on the frequency with which the number x appears on the number line, the values of n, n_2, n_3,..., z, z_2, z_3,... etc., can be analyzed.

By flipping the frequency of the number, we can determine the percentage of how often it appears on the number line (for both even and odd numbers). Since j is always even, the percentage of occurrence is given by:

$$\text{Percentage} = 0.5 \cdot 2^{-(n+n_2+z+z_2)}$$

Since this percentage reflects how frequently the number occurs on the number line (including both even and odd numbers), we divide it by 0.5 to determine how much of the percentage occupies the odd number space:

$$P = 2^{-(n+n_2+z+z_2)}$$

(eq 10)

Alternatively, this can be expressed as:

$$P = P(n) \cdot P(z)$$

(eq 11)

By this method, we begin with a value of x and calculate its occurrence rate based on the combination of factors.

Keep in mind the important barrier: $\Sigma_n \log_2\left(\frac{3}{2}\right) - \Sigma_z$, as it is the key to understanding how the value of x evolves over time. Although the first value may appear to be a random combination, all other values for different combinations exist that could be achieved if we put j as an odd number.

9. Existence of Every Unique Sequence

The following algorithm can be employed to identify the smallest number that generates a given sequence of Collatz operations, namely $\frac{3x+1}{2}$ and $\frac{x}{2}$.

- Identify the sequence. For example, consider the sequence: seq : $3, 4, 1, 2, 1, 3, 5$.
- Numbers at even positions in the sequence represent increments, while numbers at odd positions represent decrements.
- Generate an initial number T in the form $T = 2^{\text{seq}[0]} - 1$, where $\text{seq}[0]$ is the first number in the sequence.
- Declare a variable V and initialize it with the value $V = 3^{\text{seq}[0]} - 1$.
- Declare another variable L and initialize it with the value $L = V$.
- Loop through the sequence, starting from the second element.
- If the current element corresponds to a decrement, repeatedly divide L by 2 until the first odd number is encountered. Check if the length of this division sequence equals the value of the current element. If so, update V to be the last value of L and proceed to the next element in the sequence.
- If the current element corresponds to an increment, repeatedly apply the operation $\frac{3x+1}{2}$ to L until the first even number is encountered. Check if the number of operations performed equals the value of the current element. If so, update V to be equal to L and proceed to the next element in the sequence.
- If the number of operations performed does not match the current element, the following modification process is applied:

 doi:10.20944/preprints202504.0144.v2

- Declare a variable R, where for increments, $R = L + 1$, and for decrements, $R = L$.
- Declare $nP = 2 \times 3^N$, where N is the number of confirmed correct increments.
- Declare $zP = 2^E$, where E is the current element value.
- Find f, the smallest value greater than R that is divisible by E.
- Declare C, where $C = \frac{f-R}{2}$.
- Declare $A = \frac{zP}{2}$ and $B = \frac{nP}{2}$.
- Find the modular inverse of B, denoted B_{inv}, with respect to A.
- Calculate D, where $D = C \times B_{\text{inv}} \mod A$.
- If $D = 0$, set $D = 1$.
- Update T using the formula $T = T_{\text{old}} + 2 \times D \times 2^S$, where S is the sum of all confirmed elements.
- Update V using $V = V_{\text{old}} + 2 \times D \times 3^N$. Set L equal to V.
- Repeat the modification process starting from the last unconfirmed element.

- The process terminates when all elements in the sequence are confirmed. The final value of T is returned.

Note: An element is considered confirmed if it appears in the sequence generated by T in the correct order.**[?]**

Since A and B are co-prime, there exists a valid solution for calculating the modular inverse B_{inv}. **A python code is available doing the previous algorithm[?]**

10. Collatz Tree

We can build a tree to count how many numbers satisfy the Collatz conjecture. To do this, we check all possible values of n and z, exploring their possibilities until the barrier becomes zero or negative. When this happens, we return to the starting point in the tree and add the found possibility to the percentage of numbers that satisfy the conjecture, as they lead to smaller numbers.

We start with an initial barrier, where $\Sigma n = n$ and $\Sigma z = 0$. The starting percentage of numbers that lead to smaller numbers is $P = 0$.

We then branch out with all possible values of z. Each value of z has a frequency of 2^{-z}. After exploring each z, we check the new value of the barrier. If the barrier is less than or equal to zero, we stop and count the branch as leading to a smaller number. We add the frequency of that branch, 2^{-z}, to P.

If the barrier is still positive, we explore each z that didn't meet the condition and consider all possible values of n. Each value of n has a frequency of 2^{-n}. The total frequency for each branch is 2^{-z-n}. We then continue exploring all possible z-values for each branch, checking the barrier again and adding the frequency to P if it meets the condition.

Note: The frequency term 2^{-n} or 2^{-z} represents how frequently a certain pattern occurs in the process. The smaller the exponent, the more frequently that particular pattern appears. This reflects the likelihood of reaching a specific state or condition as we explore different values in the tree.

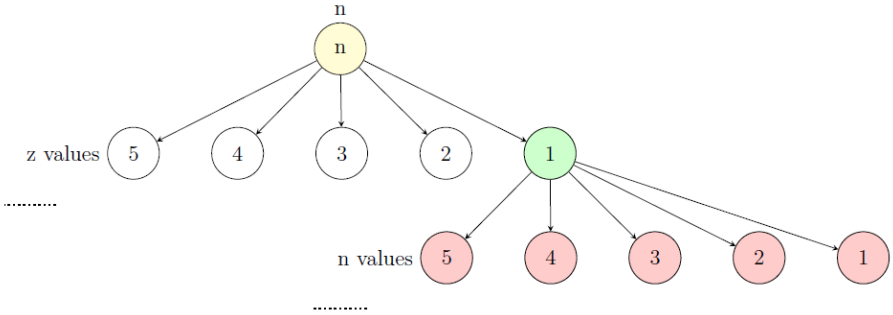

Figure 1. Visualizing Collatz tree.

 doi:10.20944/preprints202504.0144.v2

Since both n and z have the same frequency of occurrence, the likelihood of reducing the barrier increases as we progress deeper into the tree. More values are added to the percentage as we go further. This happens because in the barrier expression $\Sigma n \log_2(\frac{3}{2}) - \Sigma z$, the term Σn is multiplied by $\log_2(\frac{3}{2})$, which is approximately 0.58 (less than 1), while Σz is multiplied by -1. As we explore all possible patterns, the total value of the barrier (accounting for different patterns) becomes smaller as we go deeper into the tree. This means that $P = 1$ as the depth approaches infinity.
This sequence exhibits a behavior analogous to that of prime numbers.
Just as prime numbers define the non-prime status of other numbers through a specific repeating frequency, a similar phenomenon occurs here, where decreasing patterns recur at regular intervals, effectively covering other numbers as part of the decreasing pattern set.

In the case of prime numbers, as we extend towards infinity, they progressively cover all natural numbers as non-primes, since the deeper we analyze the distribution of primes, the higher the percentage of numbers identified as composite.
Similarly, the Collatz tree follows a similar behavior, where the branching structure systematically encompasses an increasing number of values, reinforcing its coverage as depth increases. [**?**]

This approach demonstrates that proving the impossibility of infinite increments results in a recursive process of unbounded steps.
As the proof expands, it progressively encompasses more numbers, reinforcing its validity through an increasingly extensive coverage.

Computational simulations of these steps up to a given depth reveal that as the depth increases, the value of p approaches 1.
However, for groups where $n \geq 2$, reaching an exact value of 1 remains unattainable within a fixed computational depth.
Nevertheless, the value continues to asymptotically approach 1 with increasing depth.
Function representing the previous Collatz tree:

$$C(f) = \text{sigmoid}(-fJ) + \text{sigmoid}(fJ) \times \sum_{n=1}^{L} \left(2^{-n} \sum_{z=1}^{L} 2^{-z} C(n \log_2(1.5) - z + f) \right)$$

With:

$$L \to \infty, \quad J \to \infty$$

11. Impossible increment as a probability game

When estimating the frequency of increasing or decreasing across all values of n *ranging from* 1 *to* ∞ and all values of z *ranging from* 1 *to* ∞ for a single step, it is observed that the frequency of an increase in the barrier is approximately 28.7%, while the frequency of a decrease in the barrier is approximately 71.3%.
This result can be obtained through both manual calculations and computational simulations.
The Collatz conjecture can be interpreted as a probability game. While an increment might occur at the beginning, achieving an infinite increment requires playing the game indefinitely. In the long run, the frequency of a 28.7% increment and a 71.3% decrement dominate. This suggests that infinite increments are impossible, as the process will eventually halt when the value drops below the initial number.[**?**]
The exact value of increasing or decreasing frequency is not very important at this point, because what matters most is that the frequency of decreasing is higher than the frequency of increasing.
Following either approach will eventually lead to a deterministic conclusion that infinite increment patterns do not exist, due to the recursively increasing frequency of decrement patterns.

 doi:10.20944/preprints202504.0144.v2

12. Possibility of Other Loops

Consider the equation:

$$\text{barrier} = \sum n \cdot \left(\log_2\left(\frac{3}{2}\right) + \frac{1}{\text{avg}_n} \cdot \log_2(f_{\text{root}}) \right) - \sum z$$

For a loop to exist, the barrier must equal zero. Since the logarithmic function yields an irrational number and both n and z are integers, the equation reaches zero under specific conditions. The factor in question is close to 1, and since $\log_2(1) = 0$, the impact of the logarithmic term is minimal. Consequently, the possibility of achieving a zero barrier is extremely unlikely.

Through brute-force simulations, we specifically searched for a sum of n values where multiplying by the logarithm of $\frac{3}{2}$ produces a number that is nearly an integer. This number could potentially be tuned to an integer by small adjustments to the factor f_{root}. However, no instance was found within the range from 1 to 10 billion where this condition held for more than 9 decimal places.

Even if such a value was found, another condition must be met: the sum of z values, $\sum z$, must equal the product of the sum of n values, $\sum n$, multiplied by the factor. This adds an additional layer of complexity, making it even harder to satisfy both conditions simultaneously. Therefore, achieving a zero barrier would require a very large sequence. As previously discussed, the frequency of such sequences is inversely related to their length, meaning that very long sequences are rare at smaller numbers. Moreover, if such sequences do occur at larger numbers, the factor f_{root} would approach 1, making it even more challenging to find an integer solution.

These findings indicate that the likelihood of achieving the necessary conditions to reach a zero barrier is exceptionally small.
$P(10^{10}) = 2^{(-10^{10})}$ *very close to zero.*
While the occurrence of such sequences might be frequent in some contexts, predicting the first appearance of such a sequence remains highly challenging. Therefore, although large sequences are less common at smaller numbers, there is no guarantee that they cannot exist.

As such, while this part of the conjecture has not been definitively proven, we believe the evidence strongly suggests that no loops other than the 4-2-1 loop are likely to occur.

13. Similar conjectures

Based on the established rules, it is possible to identify similar sequences, provided that the following conditions are met:

1. All potential sequence combinations must be represented.
2. Each unique sequence must exhibit a repetitive pattern with a fixed frequency of occurrence.
3. The frequency of decreasing values must surpass the frequency of increasing values.

Under these conditions, infinite growth becomes impossible, and the probability of discovering loops at relatively high values is extremely low.

As an example, consider the function:

$$f_2(x) = \begin{cases} \frac{x}{2} & \text{if } x \bmod 2 = 0 \\ 3x + G & \text{if } x \bmod 2 = 1 \end{cases}$$

where G is an odd integer constant.

 doi:10.20944/preprints202504.0144.v2

This reasoning leads to the conclusion that an infinite number of Conjectures, exhibiting behavior analogous to the Collatz conjecture, exist, but with differing arrangements of patterns and loops. Another example:

$$f_3(x) = \begin{cases} \frac{x}{3} & \text{if } x \bmod 3 = 0 \\ 4x + G & \text{if } x \bmod 3 > 0 \; and \; (4x + G) \bmod 3 = 0 \end{cases}$$

where G is a value between 1,2 the value is chosen to make the result devisable by 3 This is a different conjecture where the barrier contains $log_3(\frac{4}{3})$ instead of $log_2(\frac{3}{2})$, this means that in general this conjecture has more smaller patterns than the Collatz conjecture due to smaller value for the log.
This is the only way to have a similar behavior using values other than 3 and 2, which is by setting a changeable G, which explains why numbers 3,2 are very unique and useful to make the conjecture work correctly

14. Collatz Conjecture in Cryptography

The Collatz conjecture, though primarily studied in number theory, presents intriguing possibilities in the field of cryptography. If we consider the sequence of operations in a Collatz iteration—multiplying by 3 and adding 1 (followed by division by 2) as an "up" step and simple division by 2 as a "down" step—we can represent each transformation as a binary sequence.

For a given number x, if the operation applied is $(3x + 1)/2$, we denote it as a binary '1'. If the operation applied is $x/2$, we denote it as a binary '0'. This way, any number's Collatz trajectory can be uniquely mapped to a binary string. Since every natural number produces a different sequence of ups and downs before reaching the cycle $4, 2, 1$, this provides a form of encoding.

Cryptographic Potential

1. Pattern uniqueness: Since every number follows a distinct Collatz sequence before converging, we can use these sequences as cryptographic keys or one-way hash functions.
2. Deterministic Yet Unpredictable: While the sequence follows strict mathematical rules, predicting the steps backward (given only a binary sequence) is difficult without knowing the starting number. This could make it useful for encoding messages.
3. Controlled Variability: By manipulating how long a sequence runs before truncation, we can generate controlled-length binary keys that maintain complexity.

15. Algorithm for Encoding a Message

[**?**] The encoding process relies on the algorithm previously outlined for finding a number that represents a sequence. In this context, increments and decrements are treated as binary values: increments are encoded as 1s and divisions as 0s. This transformation enables the generation of a number that uniquely represents the chosen sequence. The following steps describe the encoding process:

- Identify the sequence of operations (increments and divisions).
- Convert increments to binary 1s and decrements to binary 0s.
- Generate a number T_{encoded} that encodes the sequence of 1s and 0s.
- Along with the encoded number T_{encoded}, two keys are passed:
  - The length of the encoded sequence, denoted as L_{seq}.
  - G factor

For decoding the encoded message, the Collatz operations are applied using the chosen G value. The decoding process proceeds as follows:

- Starting with the chosen G, apply the Collatz operations $\frac{3x+G}{2}$ for odd numbers and $\frac{x}{2}$ for even numbers.
- Perform these operations until the length of the sequence reaches L_{seq}.

 doi:10.20944/preprints202504.0144.v2

- The decoded sequence can be verified by comparing the resulting sequence of increments (1s) and decrements (0s) with the original sequence.

The encoding and decoding processes are reversible, ensuring that the original sequence can be successfully retrieved from the encoded message using the correct G value and sequence length.

Example Encoding

Consider the number $x = 5$. Its Collatz sequence is:

$$5 \to 8 \to 4 \to 2 \to 1$$

Using the rule where $(3x + 1)/2$ represents '1' and $x/2$ represents '0', the transitions encode as:

$$5 \to 8 \quad (1), \quad 8 \to 4 \quad (0), \quad 4 \to 2 \quad (0), \quad 2 \to 1 \quad (0)$$

Thus, the sequence for $x = 5$ is **"1000"**.

This binary representation can be extended to create longer cryptographic keys by concatenating multiple numbers' sequences or introducing controlled variations in starting conditions.
The unpredictable yet deterministic nature of the Collatz pattern makes it an exciting candidate for encryption, hashing, or pseudo random number generation in cryptographic applications.

16. Conclusion

This research has proven a significant part of the conjecture, establishing that infinite growth is impossible in these sequences. We also demonstrated that the conditions for forming loops other than the 4-2-1 cycle are extremely difficult to satisfy. Through our analysis, we established rules to identify sequences following similar recurrence relations to the Collatz conjecture. It was found that there are infinitely many such conjectures, each exhibiting unique patterns and loops.We have developed an algorithm capable of generating any number within a specific sequence of Collatz operations and explored its potential applications in cryptography.

17. Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

1. M. R. Schwob, P. Shiue, and R. Venkat, "Novel theorems and algorithms relating to the Collatz conjecture," *International Journal of Mathematics and Mathematical Sciences*, vol. 2021, no. 1, 2021, Article ID 5754439.
2. W. Ren, "A new approach on proving Collatz conjecture," *Journal of Mathematics*, vol. 2019, no. 1, 2019, 12 pages, Article ID 6129836.
3. T. Tao, "Almost all orbits of the Collatz map attain almost bounded values," 2019. Available: https://arxiv.org/abs/1909.03562
4. D. Barina, "Convergence verification of the Collatz problem," *The Journal of Supercomputing*, 2020.
5. James Stewart, Lothar Redlin, Saleem Watson, *Precalculus*, 7th ed., Cengage Learning, 2015.
6. I. M. Gelfand and A. Shen, *Algebra*, Birkhäuser, 2003.
7. J. Stewart, L. Redlin, and S. Watson, *Precalculus*, 7th ed., Cengage Learning, 2015.
8. Ivan Niven, Herbert S. Zuckerman, and Hugh L. Montgomery, *An Introduction to the Theory of Numbers*, John Wiley & Sons, 5th edition, 1991. https://www.wiley.com/en-us/An+Introduction+to+the+Theory+of+Numbers,+5th+Edition-p-9780471625469.
9. Sheldon M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, Elsevier, 6th edition, 2020. https://www.elsevier.com/books/introduction-to-probability-and-statistics-for-engineers-and-scientists/ross/978-0-12-817747-9.
10. Burton, D. M. (2006). *Elementary Number Theory* (7th ed.). Pearson Education.
11. Python Code for generating smallest number for a Collatz sequence https://github.com/MohamedYa123/CollatzEncoding.

 doi:10.20944/preprints202504.0144.v2