Proof of the Collatz Conjecture

Henok Tadesse
email: entkidmt@yahoo.com

21May 2023

Abstract

Take any positive integer N. If it is odd, multiply it by three and add one. If it is even, divide it by two. Repeatedly do the same operations to the results, forming a sequence. It is found that, whatever the initial number we choose, the sequence will eventually descend and reach number 1, where it enters a closed loop of 1- 4 - 2 - 1. This is known as the Collatz conjecture which states that the sequence always converges to 1. So far no proof has ever been found that this holds for every positive integer. In this paper, we present a more rigorous, deterministic proof of the Collatz conjecture, guided by heuristic and probabilistic methods. The probabilistic approach provides both structural insight into the iteration and an intuitive explanation for why the deterministic proof should hold.We have noted that the ratio between the number of odd operations and even operations continues to decrease as the sequence length increases, approaching zero for infinite sequence length. This leads to the only possibility that the sequence must eventually decouple from its staring value and enter a cycle, with the only possible cycle being the 1-4-2-1 cycle. We have obtained an equation for the final sequence of infinite length, which is the 1-4-2-1 closed loop:

$$C_n = \frac{\frac{3}{2^{Kn-1}} + 1}{2^{Kn}}$$

Introduction

The Collatz conjecture, originally proposed by Lothar Collatz in 1937[1], is a famous unsolved problem in number theory. It concerns the behavior of the iteration defined on positive integers by:

$$C(N) = \begin{cases} 3N + 1\,, & \text{if N is odd} \\ \dfrac{N}{2}, & \text{if N is even} \end{cases}$$

The conjecture asserts that for every positive integer N, repeated iterations of C eventually reaches the number 1. Despite extensive computational evidence and numerous partial results, a full proof has remained illusive[2,3].

Several surveys and studies have explored the problem's rich structure and its generalizations (see, e.g. Lagarias[2]). This paper presents a novel approach that studies the properties of an infinite Collatz sequence, particularly the ratio between the number of even operations and odd operations.

Collatz Sequence of Odd Terms

Definition 1

Odd Collatz Sequence is an alternative representation containing only the odd terms of the sequence.

A term in a Collatz sequence can be even or odd. We start from the fact that even terms always lead to odd terms because of successive divisions by 2. Therefore, for any initial odd number N, an alternative representation of Collatz sequence can be defined as:

$$N \longrightarrow \frac{3N+1}{2^{K1}} \longrightarrow \frac{3\left(\frac{3N+1}{2^{K1}}\right)+1}{2^{K2}} \longrightarrow \frac{3\left(\frac{3\left(\frac{3N+1}{2^{K1}}\right)+1}{2^{K2}}\right)+1}{2^{K3}}$$

$$\longrightarrow \frac{3\left(\frac{3\left(\frac{3\left(\frac{3N+1}{2^{K1}}\right)+1}{2^{K2}}\right)+1}{2^{K3}}\right)+1}{2^{K4}} \longrightarrow \frac{3\left(\frac{3\left(\frac{3\left(\frac{3\left(\frac{3N+1}{2^{K1}}\right)+1}{2^{K2}}\right)+1}{2^{K3}}\right)+1}{2^{K4}}\right)+1}{2^{K5}}$$

$$\longrightarrow \frac{3\left(\frac{3\left(\frac{3\left(\frac{3\left(\frac{3\left(\frac{3N+1}{2^{K1}}\right)+1}{2^{K2}}\right)+1}{2^{K3}}\right)+1}{2^{K4}}\right)+1}{2^{K5}}\right)+1}{2^{K6}} \longrightarrow \cdots$$

All terms of the above sequence are odd numbers because every time an even term occurs in the sequence it is successively divided by 2 until an odd term occurs. That is, when an even term occurs, we don't include it into the sequence but divide it by an integral power of 2 until we get an odd term, which is included into the sequence. Thus, with this we have created an alternative Collatz sequence as shown above with all terms odd.

By expanding each term we get the following:

$$\frac{3\left(\frac{3N+1}{2^{K1}}\right)+1}{2^{K2}} = \frac{3^2N+3+2^{K1}}{2^{K1+K2}}$$

$$\frac{3\left(\frac{3\left(\frac{3N+1}{2^{K1}}\right)+1}{2^{K2}}\right)+1}{2^{K3}} = \frac{3\left(\frac{3^2N+3+2^{K1}}{2^{K1+K2}}\right)+1}{2^{K3}} = \frac{3^3N+3^2+3*2^{K1}+2^{K1+K2}}{2^{K1+K2+K3}}$$

$$\frac{3\left(\frac{3\left(\frac{3\left(\frac{3N+1}{2^{K1}}\right)+1}{2^{K2}}\right)+1}{2^{K3}}\right)+1}{2^{K4}} = \frac{3\left(\frac{3^3N+3^2+3*2^{K1}+2^{K1+K2}}{2^{K1+K2+K3}}\right)+1}{2^{K4}}$$

$$= \frac{3^4N+3^3+3^2*2^{K1}+3*2^{K1+K2}+2^{K1+K2+K3}}{2^{K1+K2+K3+K4}}$$

$$\frac{3\left(\frac{3\left(\frac{3\left(\frac{3\left(\frac{3N+1}{2^{K1}}\right)+1}{2^{K2}}\right)+1}{2^{K3}}\right)+1}{2^{K4}}\right)+1}{2^{K5}} = \frac{3\left(\frac{3^4N+3^3+3^2*2^{K1}+3*2^{K1+K2}+2^{K1+K2+K3}}{2^{K1+K2+K3+K4}}\right)+1}{2^{K5}}$$

$$= \frac{3^5N + 3^4 + 3^3 * 2^{K1} + 3^2 * 2^{K1+K2} + 3 * 2^{K1+K2+K3} + 2^{K1+K2+K3+K4}}{2^{K1+K2+K3+K4+K5}}$$

and so on.

From the above, if we take the $(3N+1)/2$ term to be the first (n=1) term, then the n^{th} odd Collatz term (C_n) will be:

$$C_n = \frac{3^nN + 3^{n-1} + 3^{n-2} * 2^{K1} + ... + 3^2 * 2^{K1+K2+...Kn-3} + 3 * 2^{K1+K2+...Kn-2} + 2^{K1+K2+K3+...+Kn-1}}{2^{K1+K2+K3+...+Kn}} \quad ... (1)$$

C_n can be written as:

$$C_n = \frac{3^nN + 3^{n-1}}{2^{K1+K2+K3+...+Kn}} + \frac{3^{n-2} * 2^{K1} + ... + 3^2 * 2^{K1+K2+...Kn-3} + 3 * 2^{K1+K2+...Kn-2} + 2^{K1+K2+K3+...+Kn-1}}{2^{K1+K2+K3+...+Kn}}$$

It should be noted that K_1 , K_2 , K_3 , . . . + Kn themselves are determined by N and n.

A fully rigorous direct approach to solving this equation appears intractable with current methods. We therefore adopt the following probabilistic and heuristic approaches.

The number of even Collatz operations compared to the number of odd operations

Intuitively, the number of even operations is significantly greater than the number of odd operations, and increasingly so at very large or infinitely large numbers. This follows from the fact that Collatz operation on an odd number always results in an even number, whereas an even operation may result in several successive even operations before an odd number occurs again, and this effect is even more pronounced at very large numbers. After every odd operation, the next term is always an even term and the probability that this even term has large powers of two as a factor (e.g. 2^{1000}) indefinitely increases as the length of the sequence approaches infinity. Although the rigor of such proof could be questioned (claiming that it is based on probability), one can arbitrarily increase this probability, making it approach 1 in the limit, that is by assuming infinite length of the sequence.

Assumption 1

The ratio of the number of odd operations to the number of even operations continuously decreases with increase in length of the sequence and approaches zero for infinite length of the sequence. That is:

$$\frac{n}{K1 + K2 + K3 + \ldots + Kn} \;\longrightarrow\; 0 \quad \text{at infinity}$$

Where n is the total number of Collatz odd operations and K1+K2+K3+ . . .+Kn is the total number of Collatz even operations in a sequence.

Assumption 2

This assumption concerns the terms in equation (1):

$$\frac{3^{n}}{2^{K1+K2+K3+\ .\ .\ .\ +Kn}} \quad \longrightarrow \quad 0 \;\; as \;\; n \;\; and \;\; K1 + K2 + K3 + \cdots + Kn \;\longrightarrow\; \infty$$

We postpone the proof of these assumptions. Later on we will show that Assumption 2 follows from Assumption 1. We can see that the first term is the coefficient of N. The convergence of these terms to zero at infinity (infinite length of the sequence) would reflect the fact that the number of even Collatz operations is significantly greater than the number of odd operations in a sequence, which intuitively means that the sequence always follows a general descent, eventually reaching the number 1.

Proof of the Collatz Conjecture

Theorem 1

All Collatz sequences converge to the trivial loop 1-4-2-1.

<u>Proof</u>

In this analysis, we employ a bold heuristic method in which less significant terms are systematically eliminated, and the resulting reduced equation is solved under the constraint that all Collatz iterates must be integers.

Using Assumption 2, we can see that the first terms of equation (1) diminish to zero for a sequence of infinite length.

$$C_n \;=\; \cancelto{0}{\frac{3^n N + 3^{n-1}}{2^{K1+K2+K3+\,.\,.\,.\,+Kn}}} \;+\; \frac{3^{n-2} * 2^{K1} + .\,.\,. \;+3^2 * 2^{K1+K2+\,.\,.\,.\,Kn-3} \;+3 * 2^{K1+K2+\,.\,.\,.\,Kn-2} + 2^{K1+K2+K3+\,.\,.\,.\,+Kn-1}}{2^{K1+K2+K3+\,.\,.\,.\,+Kn}}$$

$$C_n \;\longrightarrow\; \frac{3^{n-2} * 2^{K1} + 3^{n-3} * 2^{K1+K2} + .\,.\,. \;+3^2 * 2^{K1+K2+\,.\,.\,.\,Kn-3} \;+3 * 2^{K1+K2+\,.\,.\,.\,Kn-2} + 2^{K1+K2+K3+\,.\,.\,.\,+Kn-1}}{2^{K1+K2+K3+\,.\,.\,.\,+Kn}}$$

We now factor out the term 2^{K1} ,

$$C_n \;\longrightarrow\; \frac{2^{K1}\,(3^{n-2} + 3^{n-3} * 2^{K2} \;+ 3^{n-4} * 2^{K2+K3} + .\,.\,. \;+3^2 * 2^{K2+\,.\,.\,.\,Kn-3} \;+3 * 2^{K2+\,.\,.\,.\,Kn-2} + 2^{K2+K3+\,.\,.\,.\,+Kn-1}\,)}{2^{K1+K2+K3+\,.\,.\,.\,+Kn}}$$

Cancelling the 2^{K1} terms from both the numerator and the denominator,

$$C_n \;\longrightarrow\; \frac{(3^{n-2} + 3^{n-3} * 2^{K2} + 3^{n-4} * 2^{K2+K3} + ... + 3^2 * 2^{K2+\,...\,Kn-3} + 3 * 2^{K2+\,...\,Kn-2} + 2^{K2+K3+\,...\,+Kn-1})}{2^{K2+K3+\,...\,+Kn}}$$

This can be re-written as:

$$C_n \;\longrightarrow\; \frac{(3^{n-2})}{2^{K2+K3+\,...\,+Kn}} + \frac{(3^{n-3} * 2^{K2} + 3^{n-4} * 2^{K2+K3} + ... + 3^2 * 2^{K2+\,...\,Kn-3} + 3 * 2^{K2+\,...\,Kn-2} + 2^{K2+K3+\,...\,+Kn-1})}{2^{K2+K3+\,...\,+Kn}}$$

Again, as *n* and K2+K3+ … +Kn approach infinity, the first term diminishes to zero:

$$\frac{(3^{n-2})}{2^{K2+K3+\,...\,+Kn}} \;\longrightarrow\; 0 \qquad \left(Note\ that\ this\ follows\ from\ \frac{3^n}{2^{K1+K2+K3+\,...\,+Kn}} \;\longrightarrow\; 0 \right)$$

Therefore:

$$C_n \;\longrightarrow\; \cancelto{0}{\frac{(3^{n-2})}{2^{K2+K3+\,...\,+Kn}}} + \frac{(3^{n-3} * 2^{K2} + 3^{n-4} * 2^{K2+K3} + ... + 3^2 * 2^{K2+\,...\,Kn-3} + 3 * 2^{K2+\,...\,Kn-2} + 2^{K2+K3+\,...\,+Kn-1})}{2^{K2+K3+\,...\,+Kn}}$$

$$C_n \;\longrightarrow\; \frac{(3^{n-3} * 2^{K2} + 3^{n-4} * 2^{K2+K3} + ... + 3^2 * 2^{K2+\,...\,Kn-3} + 3 * 2^{K2+\,...\,Kn-2} + 2^{K2+K3+\,...\,+Kn-1})}{2^{K2+K3+\,...\,+Kn}}$$

Again, factoring out 2^{K2} ,

$$C_n \longrightarrow \frac{2^{K2}(3^{n-3} + 3^{n-4} * 2^{K3} + 3^{n-5} * 2^{K3+K4} + \ldots + 3^2 * 2^{K3+\ldots Kn-3} + 3 * 2^{K3+\ldots Kn-2} + 2^{K3+K4+\ldots+Kn-1})}{2^{K2+K3+\ldots+Kn}}$$

and cancelling the 2^{K2} terms from both the numerator and the denominator,

$$C_n \longrightarrow \frac{(3^{n-3} + 3^{n-4} * 2^{K3} + 3^{n-5} * 2^{K3+K4} + \ldots + 3^2 * 2^{K3+\ldots Kn-3} + 3 * 2^{K3+\ldots Kn-2} + 2^{K3+K4+\ldots+Kn-1})}{2^{K3+\ldots+Kn}}$$

Again, this can be re-written as:

$$C_n \longrightarrow \frac{3^{n-3}}{2^{K3+\ldots+Kn}} + \frac{(3^{n-4} * 2^{K3} + 3^{n-5} * 2^{K3+K4} + \ldots + 3^2 * 2^{K3+\ldots Kn-3} + 3 * 2^{K3+\ldots Kn-2} + 2^{K3+K4+\ldots+Kn-1})}{2^{K3+\ldots+Kn}}$$

Again applying our assumption about convergence to zero at infinity, that is as *n* and K2+K3+ … +Kn approach infinity, the first term diminishes to zero.

$$C_n \longrightarrow \cancelto{0}{\frac{3^{n-3}}{2^{K3+\ldots+Kn}}} + \frac{(3^{n-4} * 2^{K3} + 3^{n-5} * 2^{K3+K4} + \ldots + 3^2 * 2^{K3+\ldots Kn-3} + 3 * 2^{K3+\ldots Kn-2} + 2^{K3+K4+\ldots+Kn-1})}{2^{K3+\ldots+Kn}}$$

$$C_n \longrightarrow \frac{(3^{n-4} * 2^{K3} + 3^{n-5} * 2^{K3+K4} + \ldots + 3^2 * 2^{K3+\ldots Kn-3} + 3 * 2^{K3+\ldots Kn-2} + 2^{K3+K4+\ldots+Kn-1})}{2^{K3+\ldots+Kn}}$$

Next we factor out 2^{K3} and repeat the above procedure, then factor out 2^{K4} , and so on.

It can be shown that eventually we get:

$$C_n \longrightarrow \frac{3}{2^{Kn-1+Kn}} + \frac{2^{Kn-1}}{2^{Kn-1+Kn}}$$

$$C_n \longrightarrow \frac{3}{2^{Kn-1+Kn}} + \frac{1}{2^{Kn}}$$

$$C_n \longrightarrow \frac{1}{2^{Kn}} \left(\frac{3}{2^{Kn-1}} + 1 \right) = \frac{\frac{3}{2^{Kn-1}} + 1}{2^{Kn}}$$

Now, since C_n can only be an integer, then the term:

$$\left(\frac{3}{2^{Kn-1}} + 1 \right)$$

must be some factor of 2. The only possible value of 2^{Kn-1} is 1. That is:

$$2^{Kn-1} = 1$$

Therefore,

$$\left(\frac{3}{2^{Kn-1}} + 1 \right) = \frac{3}{1} + 1 = 4$$

Therefore:

$$C_n \longrightarrow \frac{\frac{3}{2^{Kn-1}} + 1}{2^{Kn}} = \frac{4}{2^{Kn}}$$

Again, since C_n can only be an integer, only three values are possible for C_n .

$$2^{Kn} = 1 \implies C_n = 4$$

$$2^{Kn} = 2 \implies C_n = 2$$

$$2^{Kn} = 4 \implies C_n = 1$$

Thus we have proved that (almost?) all Collatz sequences eventually converge to the 1-4-2-1 loop.

What we have done is we progressively eliminated much smaller (negligible) terms, by progressively assuming more and more sequence lengths, until only two terms remained. From the requirement that terms of the Collatz sequence must be integers, we arrived at the 1-4-2-1 closed loop. Note that theoretically we could have done the same at any stage in the elimination process but those equations would be much harder to solve. Instead of stopping when two terms remained, we can also continue the procedure until only one remains, as follows:

For a Collatz sequence of infinite length diverging to infinity, the first term is infinitely smaller than the second term:

$$\frac{3}{2^{Kn-1+Kn}} \ll \frac{1}{2^{Kn}}$$

This is because for a Collatz sequence on its way of diverging to infinity, large values of K_{n-1} and K_n definitely occur along its path, making the above inequality correct. Take $K_n = K_{n-1} = 10$, for example. We can see that the left hand side is much less than the right hand side. As already explained, as the sequence diverges to infinity, the probability that an even term has very large powers of two as a factor also increases accordingly.

Therefore, for a Collatz sequence diverging to infinity:

$$C_n \longrightarrow \cancelto{0}{\frac{3}{2^{Kn-1+Kn}}} + \frac{1}{2^{Kn}}$$

$$C_n \longrightarrow \frac{1}{2^{Kn}}$$

Thus, we are inevitably left with only one term. This is the only term that would remain and would not vanish if the Collatz sequence diverged to infinity. By contradiction, the Collatz sequence cannot diverge to infinity because this term is less than or equal to 1.

Therefore, since a Collatz term can only be an integer, the only possible value will be:

$$C_n \longrightarrow \frac{1}{2^{Kn}} = \frac{1}{2^0} = 1$$

$K_n = 0$ because there is no even operation after the final odd number in the sequence, which is the number 1.

It should be noted that this whole proof rests on one assumption: the ratio of the total number of odd operations to the total number of even operations diminishes to zero for a sequence diverging to infinity because the number of even operations after each odd operation continues to increase to infinity *if the sequence diverged to infinity*. Here we see the beauty of mathematics. We began by assuming what we wanted to disprove, namely that the sequence could diverge to infinity, but ended up in the trivial 1-4-2-1 loop.

This analysis constitutes a fully deterministic proof. It should be emphasized that the use of probabilistic insights serves only as a guiding tool and does not render the proof itself probabilistic.

Why the heuristic method works?

At this point, there is no doubt that the reader is wondering what actually is going on. Why does this method work? Is it valid?
We will not present a rigorous proof here, but we can give an intuitive explanation.

Once an explicit expression for the n^{th} iterate (C_n) is obtained, the entire Collatz problem - including the mechanism of convergence to 1 and the question of whether any sequence diverges to infinity – is encoded in that equation. Any combination of variables that satisfies that equation constitutes a valid solution, independent of its interpretation or underlying meaning. However, solving this equation directly is intractable, as it involves too many variables. As a result, one is left with brute-force or trial-and-error methods, which are unsatisfactory since they offer little conceptual insight. This is precisely where new heuristic methods play a crucial role.

The method introduced in this paper consists of making suitable assumptions about the values of the variables, allowing most terms to be eliminated in a controlled manner and yielding a reduced equation. Thus, the remaining terms capture most of the contribution to the value of the n^{th} iterate. This reduced equation not only simplifies the problem substantially, but also helps determine whether a solution exists at all.

The underlying logic is as follows. Solving the full equation,which contains the complete value of the n^{th} iterate, yields a result that is fully reliable. However, such a solution is generally attainable only through extensive computational methods and is almost impossible to carry out analytically. Moreover, while computationally definitive, this approach provides little conceptual insight.

A more feasible strategy is therefore to reduce the equation to a smaller number of terms that capture most of its total value, for example, 99.9% . The resulting reduced equation is significantly easier to analyze, while still yielding a highly reliable result. In principle, one may retain additional terms to increase reliability, but this comes at the cost of increased analytical complexity.

If, instead, the equation is reduced to terms capturing only a smaller fraction of the total value, say 90%, the analysis becomes even simpler. However, in this case the reliability decreases: even if the reduced equation admits a solution, the likelihood that this solution corresponds to a true solution of the full equation diminishes as the captured value becomes insufficient.

A reduced equation that captures a sufficiently large portion of the total value not only yields reliable solutions, but also faithfully reflects the solveability of the full equation. In particular, such a reduced equation admits a solution whenever the full equation admits one, and this solution is highly reliable. Conversely, if the reduced equation has no solution, one may conclude with certainty that the full equation also has no solution.

By contrast, a reduced equation that captures only a smaller fraction of the value of the n^{th} iterate may fail to admit a solution altogether. Even when such a reduced equation does admit a solution, that solution may be unreliable and need not correspond to a solution of the full equation. In this case, the absence or presence of solutions in the reduced equation doesnot faithfully reflect the solveability of the full equation.

Does a diverging Collatz sequence exist?

Although we have presented a proof for why (almost) all Collatz sequences converge to the number 1, a proof of the Collatz conjecture will not be complete without a proof that no sequence exists that diverges to infinity. In this paper, we present a slightly different analysis for proof of convergence to the number 1 and a rigorous proof that no diverging Collatz sequence exists, by applying heuristic methods.

Re-writing equation(1):

$$C_n = \frac{3^n N + 3^{n-1} + 3^{n-2} * 2^{K1} + \ldots +3^2 * 2^{K1+K2+\ldots Kn-3} + 3 * 2^{K1+K2+\ldots Kn-2} + 2^{K1+K2+K3+\ldots+Kn-1}}{2^{K1+K2+K3+\ldots+Kn}} \quad \ldots \quad (1)$$

Multiplying both the numerator and the denominator by 2^{Kn} :

$$C_n = \frac{3^n N + 3^{n-1} + 3^{n-2} * 2^{K1} + \ldots +3^2 * 2^{K1+K2+\ldots Kn-3} + 3 * 2^{K1+K2+\ldots Kn-2} + 2^{K1+K2+K3+\ldots+Kn-1}}{2^{K1+K2+K3+\ldots+Kn}} * \frac{2^{Kn}}{2^{Kn}}$$

$$C_n = \frac{3^n N + 3^{n-1} + 3^{n-2} * 2^{K1} + \ldots + 3^2 * 2^{K1+K2+\ldots Kn-3} + 3 * 2^{K1+K2+\ldots Kn-2}}{2^{K1+K2+K3+\ldots+Kn}} * \frac{2^{Kn}}{2^{Kn}} + \frac{2^{K1+K2+K3+\ldots+Kn}}{2^{Kn}\left(2^{K1+K2+K3+\ldots+Kn}\right)}$$

$$C_n = \frac{3^n N + 3^{n-1} + 3^{n-2} * 2^{K1} + \ldots + 3^2 * 2^{K1+K2+\ldots Kn-3} + 3 * 2^{K1+K2+\ldots Kn-2}}{2^{K1+K2+K3+\ldots+Kn}} * \frac{2^{Kn}}{2^{Kn}} + \frac{1}{2^{Kn}}$$

$$C_n = \frac{3^n N + 3^{n-1} + 3^{n-2} * 2^{K1} + \ldots + 3^2 * 2^{K1+K2+\ldots Kn-3}}{2^{K1+K2+K3+\ldots+Kn}} * \frac{2^{Kn}}{2^{Kn}} * \frac{2^{Kn-1}}{2^{Kn-1}} + \frac{3}{2^{Kn+Kn-1}} + \frac{1}{2^{Kn}}$$

$$C_n = \frac{3^n N + 3^{n-1} + 3^{n-2} * 2^{K1} + \ldots + 3^2 * 2^{K1+K2+\ldots Kn-3}}{2^{K1+K2+K3+\ldots+Kn}} * \frac{2^{Kn}}{2^{Kn}} * \frac{2^{Kn-1}}{2^{Kn-1}} * \frac{2^{Kn-2}}{2^{Kn-2}} + \frac{3}{2^{Kn+Kn-1}} + \frac{1}{2^{Kn}}$$

$$C_n = \frac{3^n N + 3^{n-1} + 3^{n-2} * 2^{K1} + \ldots + 3^2 * 2^{K1+K2+\ldots Kn-3}}{2^{K1+K2+K3+\ldots+Kn}} * \frac{2^{Kn}}{2^{Kn}} * \frac{2^{Kn-1}}{2^{Kn-1}} * \frac{2^{Kn-2}}{2^{Kn-2}} + \frac{3^2}{2^{Kn+Kn-1+Kn-2}} + \frac{3}{2^{Kn+Kn-1}} + \frac{1}{2^{Kn}}$$

Repeating the above procedure we finally get:

$$C_n = \frac{3^n N}{2^{Kn+Kn-1+Kn-2+\cdots+k1}} + \frac{3^{n-1}}{2^{Kn+Kn-1+Kn-2+\cdots+k2}} + \frac{3^{n-2}}{2^{Kn+Kn-1+Kn-2+\cdots+k3}} + \ldots + \frac{3^2}{2^{Kn+Kn-1+Kn-2}} + \frac{3}{2^{Kn+Kn-1}} + \frac{1}{2^{Kn}}$$

Perhaps this equation gives better insight.

Let us compare the last two terms, by taking their ratio:

$$\frac{\left(\frac{3}{2^{Kn+Kn-1}}\right)}{\left(\frac{1}{2^{Kn}}\right)} = 3 * 2^{-Kn-1} = \frac{3}{2^{Kn-1}}$$

As the value of the Collatz terms approaches infinity, the even number density increases, which means that 2^{Kn-1} can become so large that the following inequality holds:

$$\frac{3}{2^{Kn+Kn-1}} \ll \frac{1}{2^{Kn}}$$

Likewise,

$$\frac{3^2}{2^{Kn+Kn-1+Kn-2}} \ll \frac{3}{2^{Kn+Kn-1}} \quad , \text{ since } 2^{Kn-2} \text{ will also be large}$$

and so on.

Therefore all the terms except the last term , $1/\ 2^{Kn}$, can be eliminated.

$$C_n = \cancelto{0}{\frac{3^n N}{2^{Kn+Kn-1+Kn-2+\cdots+k1}}} + \cancelto{0}{\frac{3^{n-1}}{2^{Kn+Kn-1+Kn-2+\cdots+k2}}} + \cancelto{0}{\frac{3^{n-2}}{2^{Kn+Kn-1+Kn-2+\cdots+k3}}} + \ .\ .\ .+ \cancelto{0}{\frac{3^2}{2^{Kn+Kn-1+Kn-2}}} + \cancelto{0}{\frac{3}{2^{Kn+Kn-1}}} + \frac{1}{2^{Kn}}$$

$$\Longrightarrow \quad C_n = \frac{1}{2^{Kn}}$$

Since C_n is always an integer, the only possible value is if:

$$2^{Kn} = 1$$

Therefore,

$$C_n = \frac{1}{2^{Kn}} = \frac{1}{1} = 1$$

It should be noted that we are not saying that the inequality assumed above always holds at every step of the iteration. For example, K_{n-1} could be small, say 2, in which case the inequality assumed will not hold. The idea here is that we eliminate all those terms that can be eliminated and then solve the equation containing the remaining terms. Suppose that, after making valid eliminations, the last two terms remain. Then the equation will look like:

$$C_n = \frac{3}{2^{Kn+Kn-1}} + \frac{1}{2^{Kn}}$$

Then we need to solve the above equation, which is more complex, applying the fact that C_n is always an integer.

Likewise, if the last three terms remain after eliminating all other terms, we would need to solve the following equation, which is even more complex:

$$C_n = \frac{3^2}{2^{Kn+Kn-1+Kn-2}} + \frac{3}{2^{Kn+Kn-1}} + \frac{1}{2^{Kn}}$$

Likewise, if the last four terms cannot be eliminated, C_n will be:

$$C_n = \frac{3^3}{2^{Kn+Kn-1+Kn-3}} + \frac{3^2}{2^{Kn+Kn-1+Kn-2}} + \frac{3}{2^{Kn+Kn-1}} + \frac{1}{2^{Kn}}$$

which is even more complicated, and so on. Note that all the above four equations should give the same result, $C_n = 1$.

Thus we have proved that all Collatz sequences converge to 1, if one accepts that probabilistic laws govern the Collatz iteration globally. We have proved that no Collatz diverges to infinity by assuming infinite Collatz terms always converge to 1.

In other words, we used the following argument: how can a Collatz sequence with finite starting value diverge to infinity if Collatz sequences with infinite starting values always converge to 1, assuming probabilistic laws?

Can a Collatz sequence diverge to infinity?

Now we present a more rigorous proof that no Collatz sequence can diverge to infinity.

By proving that a Collatz sequence with infinite starting value N_1 cannot diverge to infinity, we prove that a Collatz sequence with finite starting value cannot diverge to infinity.

That is, if a Collatz sequence starting with a finite value diverged to infinity (N_1) , but N_1 converges and does not diverge to even greater infinity, then no Collatz sequence diverges to infinity (N_1).

Consider infinite Collatz operations (that is n approaches infinity) on some *infinite* odd starting number N_1 .

$$C_n = \frac{3^n N_1}{2^{Kn+Kn-1+Kn-2+\cdots+k1}} + \frac{3^{n-1}}{2^{Kn+Kn-1+Kn-2+\cdots+k2}} + \frac{3^{n-2}}{2^{Kn+Kn-1+Kn-2+\cdots+k3}} + \;.\;\;.\;\;.$$

$$.\;\;.\;\;. + \frac{3^4}{2^{Kn+Kn-1+Kn-2+Kn-3+Kn-4}} + \frac{3^3}{2^{Kn+Kn-1+Kn-2+Kn-3}} + \frac{3^2}{2^{Kn+Kn-1+Kn-2}} + \frac{3}{2^{Kn+Kn-1}} + \frac{1}{2^{Kn}}$$

A Collatz sequence diverging to infinity would require that the number of even operations ($K_n + K_{n-1} + K_{n-2} + \;.\;.\;.\; + K_3 + K_2 + K_1 = m$) isn't much greater than the number of odd operations (n). If the number of even operations was equal to the number of odd operations, the sequence would diverge. Also, if the number of even operations wasn't much greater than the number of odd operations, the sequence would diverge. This is because the coefficient of N_1 diverges.

Let

$$\frac{3^n}{2^{Kn+Kn-1+Kn-2+\cdots+k1}} = \frac{3^n}{2^m} = \frac{2^{\frac{log3}{log2}n}}{2^m} = 2^{\frac{log3}{log2}n-m}$$

The following is our reference when relating the number of even operations and the number of odd operations, that is, when we say the number of even operations is (or isn't) much greater than the number of odd operations.

$$\frac{log3}{log2}\, n - m \;=\; 0 \quad \Longrightarrow \quad m = \frac{log3}{log2}\, n \;\cong 1.585n$$

So, when we say the number of even operations (*m*) is (isn't) much greater than the number of odd operations (*n*), we mean that *m* is (isn't) much greater than 1.585*n* .

So to disprove a Collatz sequence diverging to infinity, we start by assuming that the number of even operations (*m*) isn't much greater than the number of odd operations (*n*) in the above equation and see what we will get. In order to reduce the above equation, we further assume that N_1 is infinite.

With infinite N_1 , the first term will be much greater than all the other terms combined, and therefore we can eliminate all other terms. We apply the same heuristic method we used before, that is , eliminating less significant terms and solving the reduced equation under the constraint that all Collatz iterates must be integers. This method has been employed because of its success in explaining known results; we will not try to prove its validity here. (An intuitive explanation was given in the preceding section)

Therefore,

$$C_n \;= \frac{3^n\, N_1}{2^{Kn+Kn-1+Kn-2+\cdots+k1}}$$

Since most of the value of the full equation is captured in this reduced equation, a solution to this equation will be a solution to the full equation, and a lack of solution to this equation reflects that the full equation also has no solution, with certainty.

The problem then is to prove that C_n cannot diverge to infinity, after infinite iterations. Now, for C_n to be an integer, the factor:

$$\frac{3^n}{2^{Kn+Kn-1+Kn-2+\cdots+k1}}$$

must also be an integer, since N_1 itself is an integer.

The only possible integer value of C_n is if $n = 0$ and K_n , K_{n-1} , K_{n-2} , ... , K_3 , K_2 , K_1 = 0 , making

$$C_n = \frac{3^n N_1}{2^{Kn+Kn-1+Kn-2+\cdots+k1}} = \frac{3^0 N_1}{2^{0+0+0+\cdots+0}} = \frac{1 * N_1}{1} = N_1$$

which is a contradiction. This is a contradiction because we started by assuming infinite Collatz iterations but ended in $n = 0$ (no Collatz iteration). We used a starting value N_1 and got the same value N_1 after 'infinite operations', which is a contradiction. This proves that no Collatz sequence diverges to infinity.

If the number of even operations ($K_1 + K_2 + K_3 + \ .... + K_n = m$) was not (significantly or much) greater than the number of odd operations n in a Collatz sequence, that is if m was not much greater than $1.585n$, then the the value of C_n would never be an integer and this proves that no sequence diverges to infinity. In other words, m must always be significantly greater than (for relatively small starting values) or much greater than (for very large or infinitely large starting values) $1.585n$, since all terms of the sequence must be integers. That is, the number of even operations can never be 'not significantly or much greater than the number of odd operations' in any Collatz sequence, including in a sequence that would diverge to infinity (if it existed). A diverging sequence is a contradiction because a sequence cannot diverge to infinity with the number of its even terms much greater than the number of its odd operations.

In any Collatz sequence, the number of even operations (m) is always significantly (or much) greater than 1.585 times the number of odd operations (n). In this case, the above reduced equation will not be valid. Since the total sum of the rest of the terms cannot be eliminated, these terms interact (add) with the first term to form an integer, implying a converging sequence (since m is much greater than n).

It should be noted that previously we assumed large values of K_n , K_{n-1} , K_{n-2} , ... , K_3 , K_2 , K_1 , that is, assumed that the number of even operations is much greater than the number of odd operations, and this led to the elimination of all terms, including the first term containing the starting value N, of the equation except the last term, that is $1/2^{Kn}$. This gave $C_n = 1$ and there was no contradiction.

And now, as we have just shown, we assumed that the number of even operations isn't much greater than the number of odd operations, as required for a sequence that would diverge to infinity. Additionally, in order to help reduce the equation, we assumed infinite starting value of N_1 . This led to the elimination of all terms except the first term, that is the term containing the starting value N_1 . These assumption, however, led to a contradiction, disproving any sequence diverging to infinity.

This should be the rigorous proof of the Collatz conjecture that mathematicians have been searching for about eight decades.

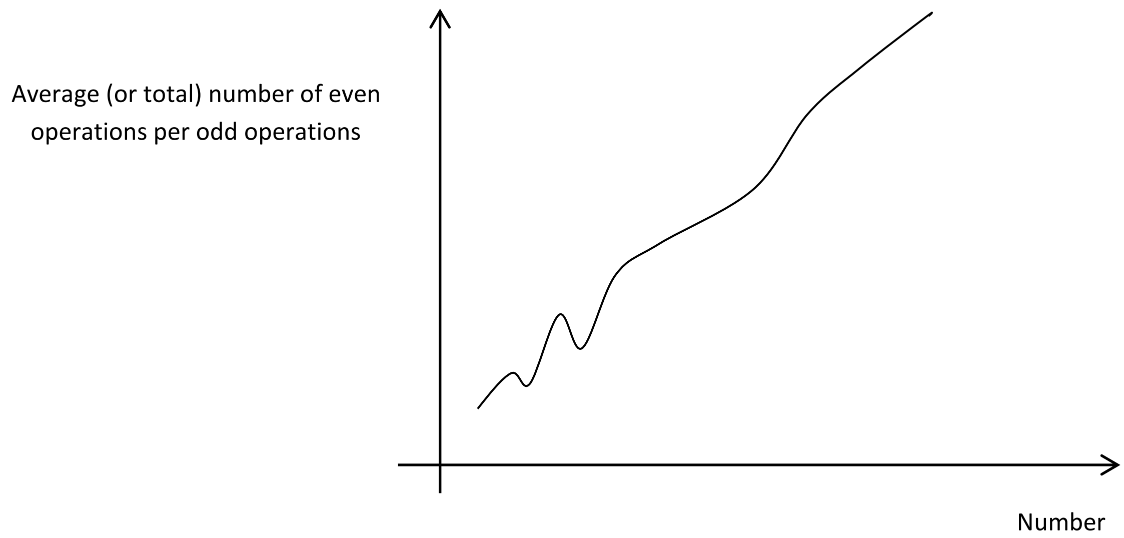
Probablistic proof of convergence to zero at infinity

In this section, we present a probabilistic proof of the Collatz conjecture as a complement to the deterministic proof presented above and as an intuitive explanation for why the deterministic proof should hold..

We start from the fact that the total number of even Collatz operations in a given Collatz sequence is significantly greater than the number of total odd Collatz operations, and this effect indefinitely increases for infinitely large numbers and infinitely long sequences. That is, if one starts with a very, very large number, the total number of even operations will be very large compared to the total number of odd operations, resulting in steep descent of the sequence. In fact, the steepness of the general descent indefinitely increases as the starting value approaches infinity. The steepness of the general descent will gradually decrease as the sequence descents towards finitely large numbers and the descent of the sequence becomes less and less pronounced at relatively small numbers, and more chaotic at smaller numbers, but eventually reaching 1 in all cases.

Why does the ratio of the total number of even operations to the number of odd operations, or the average number of even operations following each odd operation, would increase vastly if the sequence diverged to infinity? The reason for this is that at infinitely large numbers, unlike at smaller numbers, even integers can have very large powers of two as a factor, say 2^{10} , 2^{100} , 2^{1000} etc. Thus, since every Collatz operation on an odd number results in an even number, once the sequence lands on an even number it may go through tens or hundreds or thousands of successive even operations before ending in an odd number, that is before another odd operation occurs again, vastly dropping the sequence. There can never be successive odd operations, but there can always be successive even operations. Therefore, although every odd operation results in a term greater than the term before it, that operation always results in even operation(s), which may result in vast descent of the sequence, much more than its ascent by the single odd operation. In other words, at relatively small numbers, an even number can have 2, 2^2 , 2^3 , 2^4 , etc. as a factor and at very large numbers this could be all the way from 2, 2^2 , . . . to 2^{100} and still at immensely large numbers 2, 2^2 , . . . to 2^{10000} and so on , following the same trend indefinitely. Therefore, as the starting odd number increases, the average and total ratio of even to odd operations also increases accordingly. Thus, after just one odd operation, there could be thousands or billions of successive even operations before the next odd operation occurs, always resulting in a steep average descent of the sequence. As the length of the sequence increases towards infinity, the ratio between the total number of odd operations and the total number of even operations continuously tends towards zero. This results from the fact that as the starting value increases more and more even operations occur *per every odd operation*.

We can also view the problem as follows. To test whether a sequence diverges to infinity, we just consider a Collatz sequence with infinite or near infinite starting value. What we have found is that at infinitely large terms of the Collatz sequence, the sequence undergoes a steep, dramatic descent, which disproves a sequence diverging to infinity.



Note that the above graph is qualitative, meant only to explain the concept.

Proof of Assumption 2

From Assumption 1 :

$$\frac{n}{K1 + K2 + K3 + \ldots + Kn} \longrightarrow 0 \quad \text{at infinity}$$

Let

$$\frac{n}{K1 + K2 + K3 + \ldots + Kn} = r$$

$$\Longrightarrow \; n = r \; (K1 + K2 + K3 + \cdots + Kn)$$

Therefore,

$$\frac{3^n}{2^{K1+K2+K3+\ .\ .\ .\ +Kn}} = \frac{3^{r\ (K1+K2+K3+\cdots+Kn)}}{2^{K1+K2+K3+\ .\ .\ .\ +Kn}} = (\frac{3^r}{2})^{(K1+K2+K3+\cdots+Kn)}$$

Now since *r* approaches 0 at infinity (Assumption 1), *r* = 0 at infinity:

$$(\frac{3^r}{2})^{(K1+K2+K3+\cdots+Kn)} = (\frac{3^0}{2})^{(K1+K2+K3+\cdots+Kn)} = (\frac{1}{2})^{(K1+K2+K3+\cdots+Kn)} = 0\ ,\quad at\ infinity\ , that\ is\ for\ K1+K2+K3+\ .\ .\ .\ +Kn \to\ \infty$$

Proof of Assumption 1

Lemma 1

Let n_1 and n_2 be any two even positive integers, with $n_2 > n_1$. Let $n_1 = m_1 \times 2^{k1}$ and $n_2 = m_2 \times 2^{k2}$, where m_1 and m_2 are odd numbers. The probability that $k_2 > k_1$ is always greater than the probability that $k_1 > k_2$. In other words, the probability that a larger even number has a given power (k) of two is greater than the probability that a smaller even number has the same power of two.

Lemma 2 follows from the fact that the probability that an even number $n = m * 2^k$ having greater k increases as n increases. This follows from the fact that larger numbers can have higher factors of 2 than smaller numbers. For example, even numbers up to 100 can have 2^1 , 2^2 , 2^3 , 2^4 , 2^5 , 2^6 as factors, whereas even numbers up to 1000 can have 2^1 , 2^2 , 2^3 , 2^4 , 2^5 , 2^6 , 2^7 , 2^8 , 2^9 , as factors, and even numbers up to 1,000,000 can have 2^1 , 2^2 , 2^3 , 2^4 , 2^5 , 2^6 , 2^7 , 2^8 , 2^9 , 2^{10} , 2^{11} , 2^{12} , 2^{13} , 2^{14} , 2^{15} , 2^{16} , 2^{17} , 2^{18} , 2^{19} and even numbers up to 1,000,000,000 can have 2^1 , 2^2 , 2^3 , 2^4 , 2^5 , 2^6 , 2^7 , 2^8 , 2^9 , 2^{10} , 2^{11} , 2^{12} , 2^{13} , 2^{14} , 2^{15} , 2^{16} , 2^{17} , 2^{18} , 2^{19} , 2^{20} , 2^{21} , 2^{22} , 2^{23} , 2^{24}, 2^{25} , 2^{26} , 2^{27} , 2^{28} , 2^{29} and so on. We state this truth as follows: the *density* of even numbers increases as we go to higher and higher regions of the number space and this density increases indefinitely as we go to infinity.

This clearly proves that the probability that a bigger number has a given powers of 2 is greater than the probability that a smaller number has the same power of 2. It follows that the number of even operations after every odd operation increases with increase of the starting value, and increases towards infinity as the starting number goes to infinity.

The next graph illustrates this effect. We can see that the slope of the graph continues to increase as the number of Collatz operations increases. We can see that not only the local slope *dy/dx*, but also the global slope *y/x* , increases with the number of operations.

We can see that the ratio of the total number of even operations to the total number of odd operations is greater at point B than at point A on the curve, indicating a continuous increase of this ratio with increase in the number of Collatz operations (or sequence length), the ratio approaching infinity for infinite sequence length and, conversely, the ratio of odd to even operations approaching zero.

$$\frac{Number\ of\ even\ Collatz\ operations}{Number\ of\ odd\ Collatz\ operations} \simeq \infty \quad (for\ infinite\ length\ of\ sequence)$$

$$\Longrightarrow \frac{Number\ of\ odd\ Collatz\ operations}{Number\ of\ even\ Collatz\ operations} \simeq 0 \quad (for\ infinite\ length\ of\ sequence)$$

Now let us see how the number of odd operations and the number of even operations affects the sequence. Before that let us see one pattern I observed .

Any odd number N can be represented as:

$$N = N_1\, 2^m - 1 \qquad ,\ where\ N_1\ is\ also\ an\ odd\ number$$

A sequence with starting number N makes *m* consecutive (meaning every other Collatz term) odd operations before reaching an even number, which is the maximum. We see that an even number

$$N_1\, 2^m$$

also makes m even operations.

We know that the consecutive odd operations ascend the sequence. Let us see by how much:

From

$$N \longrightarrow \frac{3N+1}{2} = 1.5 * N + 0.5 \cong 1.5\, N \quad , for\ N \gg 0.5$$

Therefore,one odd operation , (3N+1)/2 , ascends the sequence by a factor of 1.5. One even operation (N/2) descends the sequence by a factor of 2.

Now let us compare *k* consecutive even operations and *n* consecutive odd operations.

$$\frac{2^k}{1.5^n}$$

Let us compare the effect for k = n =10 (we tak k = n for comparison of their effects):

$$(\frac{4}{3})^{10} = 17.7$$

For k = n = 20

$$(\frac{4}{3})^{20} = 315.3$$

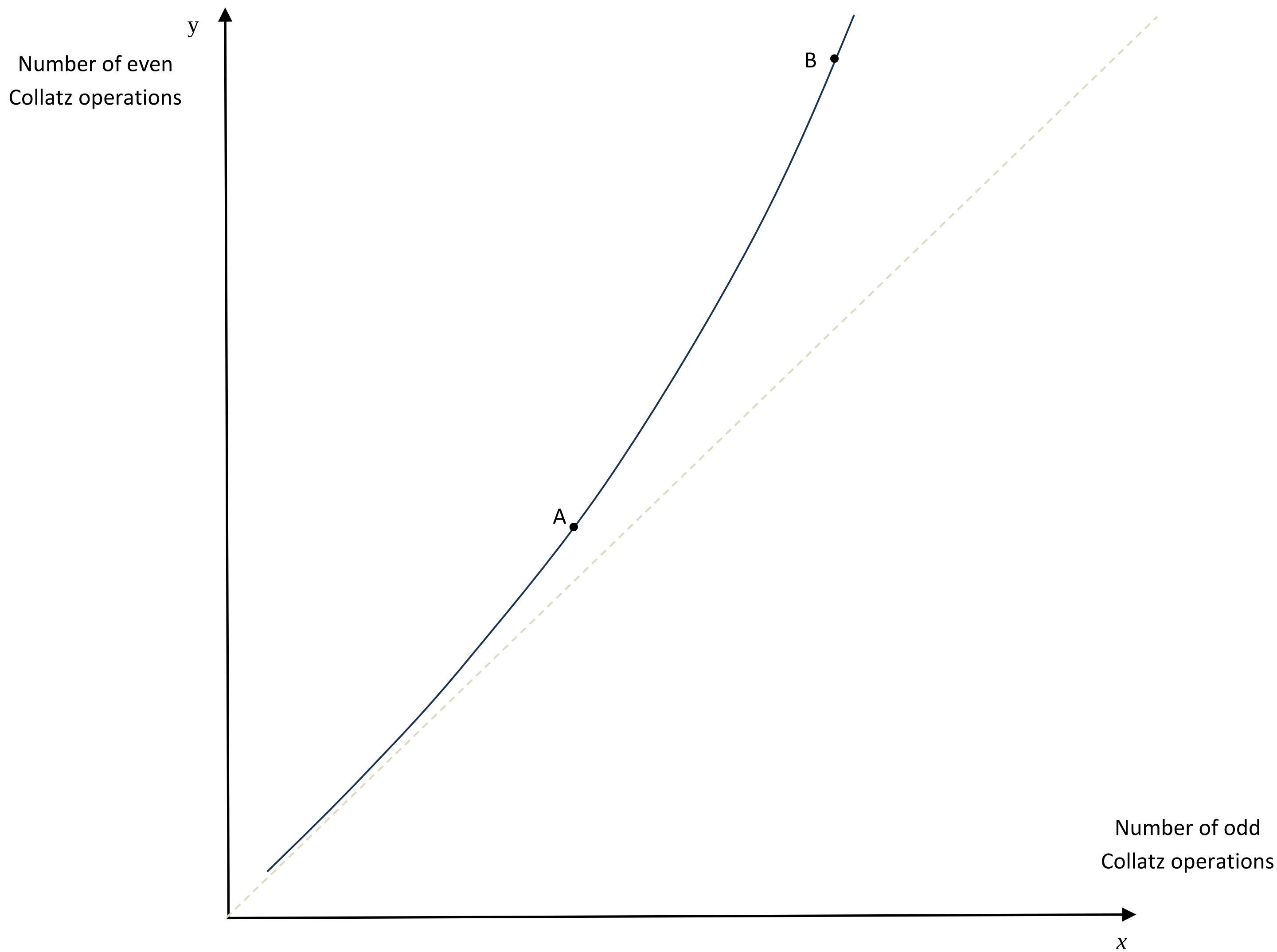
For k = n = 50

$$(\frac{4}{3})^{50} = 1765781$$

For k = n = 100

$$(\frac{4}{3})^{100} \cong 3117982410000$$

This means that100 consecutive odd operations followed by 100 consecutive even operations will descend the sequence by a factor of about three trillion. We can see that is effect is even more dramatic with increase of *n* and *k* to 300, 1000,1000000, etc. Therefore, we see that there is no mystery on why the Collatz sequence always converges if we accept that probabilistic laws globally govern the sequence.


y
Number of even
Collatz operations
B
A
Number of odd
Collatz operations
x

Now, there are two possibilities for the Collatz conjecture to be disproved.

1. The sequence indefinitely diverges to infinity.
2. The sequence may enter some closed loop (there might be more than one loop.

Since we have already proved that the Collatz sequence cannot diverge to infinity, next we consider the possibility of another closed loop other than the trivial loop.

Proof that no closed loops exist other than the 1-4-2-1 loop

Corollary 1

There are no closed loops of the Collatz sequence other than the trivial loop of 1-4-2-1.

We propose a property of a closed loop as follows: *any closed loop is decoupled from the sequences and the numbers in the sequences leading to it*. Since multiple Collatz sequences lead to the closed loop, the closed loop must be independent from any sequence (and its terms) leading to that closed loop. It necessarily follows that the coefficient of N vanishes at infinity, in the following equation:

$$C_n = \frac{3^n N + 3^{n-1}}{2^{K1+K2+K3+\ ...\ +Kn}} + \frac{3^{n-2} * 2^{K1} + \ ...\ +3^2 * 2^{K1+K2+\ ...\ Kn-3} +3 * 2^{K1+K2+\ ...\ Kn-2} + 2^{K1+K2+K3+\ ...\ +Kn-1}}{2^{K1+K2+K3+\ ...\ +Kn}}$$

Note that the coefficient of N here is:

$$\frac{3^n}{2^{K1+K2+K3+\ ...\ +Kn}}$$

which in turn brings us back to our previous assumption about the convergence of terms to zero at infinity, which in turn leads to the 1-4-2-1 loop! So we began with an assumption that another closed loop could exist somewhere in the darkness of infinity, which demanded the vanishing of the coefficient of N, which led back to the only known closed loop of 1-4-2-1. This should be a complete, rigorous mathematical proof (by induction) that no other closed loop exists.

In other words, assume that a closed loop exists at some immensely large number. Any closed loop requires that the coefficient of N must diminish to zero in the limit. Assume that the sequence was not decoupled from the starting number even at infinity, that is for infinite length of the sequence. By contradiction, there cannot be such a closed loop because C_n cannot be shown to converge to constant numbers (possible terms in the supposed closed loop), in the same way that we have proved that every Collatz sequence converges to the number 4, 2 or 1. In summary, any closed loop requires the coefficient of N to diminish to zero at infinity, but then this leads to the 1-4-2-1 loop. This proves that no other closed loop can exist.

Arbitrarily Close to One Probability Vs Absolute Certainty

Our analysis and arguments have shown that the probability of a Collatz sequence converging to the number 1 can be shown to be arbitrarily close to *one*. Conversely, the probability of a Collatz sequence diverging to infinity can be shown to be arbitrarily close to *zero*. But one can argue that, irrespective of all the arguments and analysis made so far, these arguments and analysis do not rule out at least one sequence diverging to infinity, out of infinitely many possible sequences.

$$\lim_{n\to\infty} P_D \;=\; 0 \quad , \;\; where\ P_D\ is\ the\ probability\ that\ a\ given\ sequence\ diverges$$

In this paper, we have proved that (at least) *almost all* Collatz sequences converge to the number 1. The question then is: could a Collatz sequence exist, at least one, that diverges to infinity? Can we say with absolute certainty that no Collatz sequence exists that diverges to infinity?

We raise the following fundamental questions:

1. If the analysis presented proves that *almost* all Collatz sequences converge, then what is the proof that *all* sequences converge?
2. Do we need a different, unrelated proof for the latter, unrelated to the number theories and Collatz sequence behaviors discussed in this paper?
3. Or, does such a proof even exist?
4. Or, could the ideas and analysis introduced in this paper lead to a complete proof of the conjecture?

We strongly feel that the analysis presented in this paper should also be *part of* the complete proof.

Postulate 1

If the convergence of sequences is rigorously proved based on probabilistic approach, then probabilistic approach must also prove or disprove the divergence of sequences with certainty.

Postulate 2

The ratio between the number of total Collatz operations to the number of even Collatz operations increases as we move to higher regions of the number space.

Why must this postulate be true? Because we have logical explanations and because we have been able to successfully explain why all known Collatz sequences converge.

Postulate 3

Postulate 2 applies to all Collatz sequences and all terms of Collatz sequences, including to sequences that diverge to infinity (if they exist).

Therefore, this principle is universal and applies to *all* Collatz sequences and their terms. To state that these principles could possibly not apply to sequences that diverge to infinity (if they exist) would contradict all the successful analysis we have done so far. If the principle about the ratio between the number of odd operations to the number of even operations explains why all known sequences converge, then the same principle should allow a sequence diverging to infinity, if any exists.

From the above, it follows that the probability that a sequence diverges to infinity is arbitrarily close to zero. As repeatedly explained, as the sequence continues to ascend to infinity, it would eventually enter a region and a border beyond which it cannot ascend anymore, because of high *density* of even numbers.

Definition 2

The density of even numbers in a region of the number space is qualitatively defined as a measure of the number of even numbers in that region having large powers of two as a factor.

Does a Collatz sequence exist, at least one sequence, that diverges to infinity?

We have analytically proved that the probability of a sequence diverging to infinity is zero (zero in the limit). But this raises an immediate objection: zero probability does not rule out at least one diverging sequence. One can argue that, divergence of one sequence out of infinitely many possible sequences does not contradict our result of zero probability and therefore,the analysis made so far in this paper has not ruled out at least one diverging sequence. Zero probability that a sequence diverges to infinity does not guarantee zero sum (total) of probabilitees of divergence of infinitely many Collatz sequences. That is, even if the probability of each sequence diverging to infinity is zero (almost zero), the sum of infinitely many zeros could give finite total probability. The only guarantee that not even a single diverging Collatz sequence exists is if the *sum* of the probabilities of of divergence of the infinitely many possible sequences itself is shown to be *zero*.

Let
P_1 = the probability of Collatz sequence 1 diverging to infinity = 0
P_2 = the probability of Collatz sequence 2 diverging to infinity = 0
P_3 = the probability of Collatz sequence 3 diverging to infinity = 0

$$\vdots$$

P_N = the probability of a collatz sequence N diverging to infinity = 0

Proving that not even one Collatz sequence exists that diverges to infinity requires that:

$$P_1 + P_1 + P_1 + \ . \ . \ \ . + P_N = 0 \ , \qquad where\ N\ is\ infinite$$

$$0 + 0 + 0 + \ . \ . \ \ . + 0 = 0$$

If

$$0 + 0 + 0 + \ . \ . \ \ . + 0 = 1$$

then one diverging sequence exists. If

$$0 + 0 + 0 + \ . \ . \ \ . + 0 = 2$$

then two diverging sequences exist and if:

$$0 + 0 + 0 + \ldots + 0 = 1000000$$

then one million diverging sequences exist, and so on.

The above is to illustrate that zero individual probabilities may not necessarily add up to zero total probability.

Now back to the problem: can we determine with certainty that no Collatz sequence exists that diverges to infinity ?
To prove this, we need to prove that the sum of the probabilities of divergence of each of the infinitely many Collatz sequences is zero. That is:

$$0 + 0 + 0 + \ldots + 0 = 0$$

Consider infinitely many sequences each with a zero probability of diverging to infinity, as already proved in the analysis. Assume that at least one diverging sequence exists. Such a sequence would have to indefinitely ascend to infinity, against all odds. Even number density, as we already defined, continues to increase as the sequence ascends. As the sequence ascends to higher and higher regions up the number space, and to infinity, the even number density also continues to increase indefinitely, and the probability of sustained ascent continues to diminish. Let us assume that a sequence diverges forever to infinity,irrespective of this effect.

Let the probability that a sequence ascends to 10^{100} be some value, say 10^{-10}. Then the probability of ascending the next 10^{500} would be even less, say 10^{-20} , and the probability of ascending the next 10^{1500} would be even less 10^{-40} , so on to infinity. Therefore, the probability of the sequence diverging to infinity will be the product of infinite, small, diminishing probabilities:

$$Probability\ of\ divergence\ of\ a\ sequence =$$

$$10^{-10} * 10^{-20} * 10^{-40} * \ldots * 10^{-1000} * \ldots * 10^{-1000000} * \ldots * 10^{-1000000000000000000} * \ldots * 10^{-\infty} = 0$$

To get the total probability of divergence, the probabilities of divergence of each of the Collatz sequences are added. Such zeros are so extremely small that the sum of infinite such zeros gives zero.

$$P_1 + P_1 + P_1 + \ldots + P_N = 0 , \quad where\ N\ is\ infinite$$

$$0 + 0 + 0 + \ldots + 0 = 0$$

This proves that no Collatz sequence exists that diverges to infinity. Note that the numbers used here are only conceptual. Later we will see a better estimate of these numbers.

Postulate 4

Not all infinities are equal, not all 0s are equal and not all 1s are equal.!

Fallacious and vague assumptions about infinities and zeros have caused stagnation in progress:

$$\infty * \infty \neq \infty \quad ; \quad \infty^2 \neq \infty \quad ; \quad \infty^\infty \neq \infty$$

$$0^2 \neq 0 \quad ; \quad 0^\infty \neq 0 \quad ; \quad 0^\infty \lll\lll\lll\lll\lll 0$$

Here we introduce new concept of infinities and zeros.

$$\infty \ , \quad \infty_1 \ , \quad \infty_2 \ , \quad \infty_3 \ , \quad etc.$$

$$0 \ , \quad 0_1 \ , \quad 0_2 \ , \quad 0_3 \ , \quad etc.$$

where ∞ and 0 are the current vaguely understood infinity and zero, and ∞_1 , ∞_2 , ∞_3 , … and 0_1 , 0_2 , 0_3 , … represent the newly introduced infinity and zero concept.

Therefore,

$$\infty * \infty = \infty_1 \neq \infty \quad ; \quad \infty^2 = \infty_1 \neq \infty \quad ; \quad \infty^\infty = \infty_1 \neq \infty$$

$$0^2 = 0_1 \neq 0 \quad ; \quad 0^\infty = 0_1 \neq 0 \quad ; \quad 0^\infty \lll\lll\lll\lll\lll 0$$

$$0 + 0 + 0 + \ldots + 0 = 0_1 \neq 0$$

We can equally define the new concept for 1:

$$1 * 1 * 1 * \ldots * 1 = 1_1 \neq 1$$

This means that even if the probabilility of each individual event is 1 (almost), the the probability of *all* events will not necessarily be 1.

Now let us apply this concept to the Collatz conjecture. I have presented a reasonably rigorous analysis and argument that the probability of a Collatz sequence diverging to infinity is zero in the limit. A counter-argument could be made that although the probability of divergence of each sequence is zero, the sum of the probabilities of divergence of the *infinite* possible Collatz sequences may not be zero, because $0*\infty$ is indeterminate, and therefore at least one diverging sequence has not been ruled out.
Let the probability of divergence of each sequence be P_1. Let the number of possible Collatz sequences be *Infinity 1*, ∞_1 .
Now, the sum of the probabilities of divergence of all sequences will be:

$$Sum\ of\ probabilities\ of\ diverence\ of\ infinite\ Collatz\ sequences = \ P_1 * \infty_1$$

Now, existence of one Collatz sequence diverging to infinity is excluded only if we can show that the above product is zero. That is:

$$P_1 * \infty_1 = 0$$

For the above product to be zero, P_1 must be a zero that is extremely so small that the product will be zero. Remember*: not all zeros are equal*!. (Conventionally, one would argue that zero times infinity is indeterminate).

As we have already discussed, P_1 is not an 'ordinary' zero ! P_1 is an infinite (let us call this *Infinity 2* , ∞_2) product of extremely small probabilities p (lower case) !

Let P_1 be:

$$P_1 = p * p * p * \ \dots\dots * p = p^{\infty_2} \ ,$$

where p (lower case) is the probability that a Collatz sequence ascends upto some value, or 'height' in the number space, say H_1.

We know that starting values of up to $2^{71} \approx 10^{21}$ have been checked and all sequences shown to converg to 1. Sequences that ascended as high as tens of millions times the starting number have been recorded. Even if we don't have information on the particular starting value that led to such sequence, we can safely assume the starting value for this particular sequence to be the highest checked value, that is 2^{71} , so the highest a sequence has ascended would be 10,000,000 times 2^{71} , which is about 10^{28} . We take this to be H1.

Therefore, for starting numbers upto 10^{21} , since at least one diverging sequence has not ascended beyond $H_1 = 10^{28}$, the probability that a sequence ascends beyond some value (H_1 in this case) is less than 1 part in 10^{21} , that is 10^{-21} , which we will take as p.

$$p = 10^{-21}$$

A sequence diverging to infinity would have to reach value H_1 , AND then reach H_2 , AND then reach H_3 , . . . and AND then H_{100} . . . AND then reach $H_{1000000000}$ ……………………….. AND then reach $H_{\infty 2}$. The ANDs convert to multiplication of probabilities. Although we have estimated H_1 based on computational data, its actual value is not important in our analysis, and therefore we don't need to estimate the other values also: H_2 , H_3 , H_4 , ……. . , and this does not affect our argument and analysis. We estimated the value of H_1 to help the reader understand what it is, and to avoid any confusions.

Therefore,

$$\textit{The probability that a sequence diverges to infinity } (P_1) =$$

$$\textit{the probability that it ascendes to some value } H_1 \;*\; \textit{the probability that it ascends from } H_1 \textit{ to some higher value } H_2 \;*$$

$$\textit{the probability that it ascends from } H_2 \textit{ to some higher value } H_3 \;*\; \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \;*$$

$$\textit{the probability that it ascends from } H_{\infty_2 - 1} \textit{ to some higher value } H_{\infty_2}$$

Assume that H_1 , H_2 , H_3 , … are such that the probability that a sequence reaches H_1 , the probability that the sequence ascends from H_1 to some higher value H_2 , the probability that the sequence ascends from H_2 to some higher value H_3 , ….. and so on, are all equal to p , for simplicity of argument.

Therefore,

$$P_1 = p * p * p * \ldots\ldots * p = p^{\infty_2} = 10^{-21} * 10^{-21} * 10^{-21} * \;.\;.\;.\; * 10^{-21} = (10^{-21})^{\infty_2}$$

Logically and intuitively, this means that a sequence can diverge to infinity only if it reaches some value AND if it reaches some next higher value AND if it reaches some higher next value AND if it reaches some next higher value and so on , this continuous indefinitely. In probability theory these ANDs mean multiplication of probabilities.

Therefore, the probability of a Collatz sequence diverging to infinity will be:

$$P_1 = (10^{-21})^{\infty_2}$$

Now that we have obtained the probability of divergence of one Collatz sequence, we can compute the probability that no Collatz sequence diverges to infinity.

$$P_1 * \infty_1 = (10^{-21})^{\infty_2} * \infty_1$$

Now we have two separated infinities in this equation, not mixed up as in conventional understanding.
Now remember the usual counter-argument : even if the probability of divergence of each sequence is zero, the sum of infinite (∞_1) such zeros may not necessarily be zero, so a diverging sequence is not excluded.

However, the new insight being presented in this paper defines two different infinities and the counter-argument can be easily refuted because ∞_2 overwhelms the effect of ∞_1, because ∞_2 is an exponent, whereas ∞_1 is a multiplier, with vastly different effects. To illustrate this let :

$$\infty_1 = \infty_2 = 10^{1000}$$

Therefore,

$$(10^{-21})^{\infty_2} * \infty_1 = (10^{-21})^{10^{1000}} * 10^{1000} = 10^{-2100000000000000000000000000......000} * 10^{1000} \approx (10^{-21})^{10^{1000}} = 0$$

which is overwhelmingly zero. Although we have assumed the same value for both ∞_1 and ∞_2, the exponent effect of ∞_2 overwhelmed the multiplying effect of ∞_1. (Not all infinities are equal!). Therefore, this proves with certainty that no Collatz sequence exists that diverges to infinity. Perhaps any doubt beyond this would be philosophical, not mathematical.

∞_1 and ∞_2 can be assumed to be much larger or even infinity, yet another infinity, ∞_3 .

The key in this analysis is the idea that the probability of a Collatz sequence diverging to infinity is a *product* of infinite small probabilities. Therefore,

$$Sum\ of\ probabilities\ of\ diverence\ of\ infinite\ Collatz\ sequences = 0$$

Therefore,

$$Probability\ that\ no\ Collatz\ sequence\ diverges = 1$$

To get some perspective, we know that starting values up to 10^{21} have been checked. Imagine increasing this to $\infty 1$, where some very large value can be assigned to ∞_1 . What would be the probability of finding one diverging sequence? We have shown that the probability of a Collatz sequence diverging to infinity can be assumed to be of the form:

$$P_1 = \quad (10^{-21})^{\infty_2}$$

Therefore, the probability of finding one diverging sequence out of $\infty 1$ will be:

$$\infty_1 * \quad (10^{-21})^{\infty_2} = \infty_1 * 10^{-21\infty_2}$$

From the above equation, we can see that increasing $\infty 1$ increases the probability, whereas increasing $\infty 2$ exponentially decreases the probability. One would expect that increasing the sample size from 10^{21} to $\infty 1$ increases the probability of finding a diverging sequence. However, since $\infty 2$ is an exponent, while $\infty 1$is a multiplier, the effect of $\infty 2$ completely overwhelms that of $\infty 1$.

A foundational postulate completing the argument

We have shown, with reasonable rigor, that the probability that a sequence diverges is zero. The question is therefore no longer whether a diverging sequence might exist, but a more fundamental one: can probability be taken as the final word in mathematics? Even after establishing that the probability is zero, the mind does not cease its curiosity. Mathematics demands exactness and absolute certainty. This raises a deeper question: if probability has enabled a successful analysis, can probability also provide closure? We argue that it would be both natural and conceptually satisfying if probability itself delivered the final answer. The persistence of doubt may not signal a flaw, but rather the existence of a final, precise resolution.

We therefore propose the following foundational closure principle: if the probability that a sequence diverges is zero in the limit-that is , if the total probability assigned to all diverging sequences vanishes in the limit-then no diverging sequence exists.

Postulate 5 Zero-Probability Non-Existence

If an event is assigned probability zero in the limit under a fully specified and deterministic mathematical model, then the event does not exist.

$$Probability\ that\ a\ Collatz\ sequence\ diverges = \frac{number\ of\ diverging\ sequences}{total\ number\ of\ sequences} = \quad \frac{0}{\infty} = 0$$

We need to make a crucial distinction here. Not all zeros are equal!

$$\frac{0}{\infty} = 0 \quad but \quad \frac{1}{\infty} = 0_1 \neq 0$$

If the probability of an event is mathematically zero in the limit, then it does not exist. By ‘zero’ we mean exactly zero. In a sample space of ∞ Collatz sequences, there is exactly *zero* diverging sequence. With a probability of 1/∞ , one would find more and more diverging sequences if the sample space is increased. With a probability of 0/∞ , the number of diverging sequences found is exactly zero, regardless of the sample space. This is when probability meets certainty. A probability of zero (**0**) means absolute certainty that the event does not exist, and a probability of one (**1**) means absolute certainty that the event exists.

Note the distinction we have made:

$$\mathbf{0 \neq 0_1} \quad \boldsymbol{and} \quad \mathbf{1 \neq 1_1}$$

Conclusion

In this paper, we have been able to prove the Collatz sequence, by studying the property of the sequence of infinite length, thereby avoiding the chaos of the finite length sequences. We have presented a fully deterministic proof, guided by probabilistic and by a new heuristic method.

We noted that at infinite sequence lengths, the ratio of odd operations to even operations diminishes, approaching zero. This leads to the eventual decoupling of the sequence from the starting value, ending in the only possible path of all sequences: the 1-4-2-1 loop. We have also introduced a new insight about infinities and zeros, to close the gap that exists due to vague understanding regarding probabilities that depend on infinities and zeros.

We have shown that the probability that a sequence diverges is zero. The question, however, is no longer whether a diverging sequence might exist, but a deeper, philosophical one: can probability be taken as the final word in mathematics? Even after being told that the probability is zero, with reasonably rigorous analysis, the mind does not stop asking. To close this question, we have made a final postulate: *if the probability of an event is mathematically zero in the limit, then it doesn’t exist*.

An expected objection is that only deterministic approach, and not probabilistic approach, can be considered to be a rigorous proof of the Collatz conjecture. My counter-argument is that since probabilistic approach has successfully proved why almost all sequences converge, and converge to 1, then this must be because the Collatz sequences are globally governed by probabilistic laws. Therefore, the same probabilistic approach must apply to determine if any diverging sequences exist, particularly in the absence of progress with deterministic approaches.

Since the deterministic proof and the probabilistic proof compliment each other, objections raised against the probabilistic argument cannot undermine its validity.

Glory Be To Almighty God Jesus Christ and His Mother, Our Lady Saint Virgin Mary

Notes and References

1. L.Collatz, *On the behavior of sequences defined by simple arithmetic operations*, 1937

2. J.C, Lagarias, *The 3x+1 problem and its generalizations*, American Mathematical Monthly, vol.92, no.1,pp.3-23,1985

3. Jeffrey C.Lagarias (Ed.), *The ultimate challenge : The 3x+1 Problem*, AMS,2010

APPENDIX 1

Intuitive explanation of convergence

Intuitively, we can understand the Collatz sequence as resulting from the interaction of two ‘forces’: upward pushing force and downward pulling force. The upward pushing force is the odd operations and the downward pulling force is the even operations. As the sequence continues to ascend towards infinity, if such a sequence ever exists, the sequence will eventually enter a region in the number space beyond which it cannot ascend because the effect of the even operations can no more be overcome by the odd operations.

One can get more insights by studying other Collatz like sequences. For example:

$$C(N) = \begin{cases} 5N + 1\,, & \text{if N is odd} \\ \dfrac{N}{2}, & \text{if N is even} \end{cases}$$

Or, the even more explosive:

$$C(N) = \begin{cases} 3N^2 + 1\,, & \text{if N is odd} \\ \dfrac{N}{2}, & \text{if N is even} \end{cases}$$

Intuitively, it seems that, particularly the second sequence is divergent and if one were to do computations to study its behavior, one would conclude that this sequence always diverges to infinity. However, whether this sequence diverges forever or ever converges can only be determined by the kind of analysis we have already done. In this case, the ‘upward’ force is much stronger than the downward ‘force’. Whether the effect of increasing even number density as one ascends towards infinity would curb this sequence (as it does in the case of the Collatz sequence) can only be determined by analysis. Perhaps, the problem regarding this sequence could be to determine with certainty that the sequence always diverges to infinity, that is, determine with certainty that *all* sequences diverge to infinity. This problem would be quite the opposite of the Collatz sequence. Note that , unlike the Collatz sequence, no amount of computation can prove the divergence of even one sequence. Only convergence of a given sequence can be computationally proved and no computation exists that can prove the divergence of a sequence. Divergence can only be proved analytically. Alternatively, perhaps this sequence might eventually enter a region in the number space that is immensely large, that it cannot go beyond due to the even number density effect, and eventually enter a closed loop and hover there indefinitely. *If a sequence cannot diverge to infinity, it cannot progress indefinitely without eventually entering some closed loop.* Intuitively, I guess that this particular sequence diverges to infinity because its upward progress outpaces the downward ‘force’. However, only rigorous analysis by assuming infinite sequence length can tell what happens with this particular sequence.