Article

Not peer-reviewed version

Collatz Conjecture

Asset Durmagambetov * and Aniyar Durmagambetova

Posted Date: 5 February 2024

doi: 10.20944/preprints202401.0227.v11

Keywords: binary representation; Collatz conjecture

doi:10.20944/preprints202401.0227.v11

Article

Collatz Conjecture

Asset Durmagambetov, Aniyar Durmagambetova

[1] L.N. Gumilyov Eurasian National University; aset.durmagambet@gmail.com
[2] Nazarbayev University; durmagambetovaa4@gmail.com
* Correspondence: aset.durmagambet@gmail.com; Tel.: +77787286399

Abstract: This paper presents an analysis of the number of zeros in the binary representation of natural numbers. The primary method of analysis involves the use of the concept of the fractional part of a number, which naturally emerges in the determination of binary representation. This idea is grounded in the fundamental property of the Riemann zeta function, constructed using the fractional part of a number. Understanding that the ratio between the fractional and integer parts of a number, analogous to the Riemann zeta function, reflects the profound laws of numbers becomes the key insight of this work. The findings suggest a new perspective on the interrelation between elementary properties of numbers and more complex mathematical concepts, potentially opening new directions in number theory and analysis.

Keywords: binary representation; Collatz conjecture

---

1. Introduction

We will use the following well-known fact: if, for the members of the Collatz sequence, zeros predominate in their binary representation, then these members will lead to a decrease in the subsequent members according to the Collatz rule. A striking example is when the initial number in the Collatz sequence is equal to 2^n. Let's write the solution of the equation $n = 2^x$ in the form $x = \{x\} + [x]$ and note that the smaller x, the more zeros in the corresponding binary representation for n. Developing this idea, we come to the following steps.

- Analysis of the binary representation of simple cases of natural numbers.
- Creation of a process for decomposing an arbitrary natural number into powers of two.
- Analysis of the proximity of the process to binary decomposition at the completion of decomposition at each stage.
- Calculation of the number of zeros in the binary decomposition of a natural number.
- Estimation of the Collatz sequence members depending on the number of ones in the binary decomposition.

2. Results

This document reveals a comprehensive solution to the Collatz Conjecture, as first proposed in [1]. The Collatz Conjecture, a well-known unsolved problem in mathematics, questions whether iterative application of two basic arithmetic operations can invariably convert any positive integer into 1. It deals with integer sequences generated by the following rule: if a term is even, the subsequent term is half of it; if odd, the next term is the previous term tripled plus one. The conjecture posits that all such sequences culminate in 1, regardless of the initial positive integer. Named after mathematician Lothar Collatz, who introduced the concept in 1937, this conjecture is also known as the 3n + 1 problem, the Ulam conjecture, Kakutani's problem, the Thwaites conjecture, Hasse's algorithm, or the Syracuse problem. The sequence is often termed the hailstone sequence due to its fluctuating nature, resembling the movement of hailstones. Paul Erdős and Jeffrey Lagarias have commented on the complexity and mathematical depth of the Collatz Conjecture, highlighting its challenging nature. Consider an operation applied to any positive integer:

- Divide it by two if it's even.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2024 doi:10.20944/preprints202401.0227.v11

- Triple it and add one if it's odd.

This operation is mathematically defined as:

$$f(n) = \begin{cases} \frac{n}{2}, & \text{if } n \equiv 0 \mod 2, \\ 3n+1, & \text{if } n \equiv 1 \mod 2. \end{cases}$$

A sequence is formed by continuously applying this operation, starting with any positive integer, where each step's result becomes the next input. The Collatz Conjecture asserts that this sequence will always reach 1 Recent substantial advancements in addressing the Collatz problem have been documented in works [2]. Now let's move on to our research, which we will conduct according to the announced plan. For this, we will start with the following

Theorem 1. *Let*

$$\begin{aligned} M &\in \mathbb{N}, \\ [\alpha_j] - [\alpha_{j+1}] &= \delta_j > 0, \\ \epsilon_1 &< 0.65, \\ |F_j(x)| &< |x|, \\ \alpha_j &= [\alpha_j] + \epsilon_j, \\ \epsilon_j &< 1, \\ \sigma_j &= 1 - \epsilon_j. \end{aligned}$$

$$M = \sum_{i=1}^{j-1} 2^{[\alpha_i]} + 2^{\alpha_j}, M = \sum_{i=1}^{j} 2^{[\alpha_i]} + 2^{\alpha_{j+1}}, \tag{1}$$

Then for $\delta_j = 1$

$$\sigma_j = 2^{-1}\sigma_{j+1}\left(1 - \frac{\sigma_{j+1}\ln 2}{2}\right) + F_j\left(\frac{\sigma_{j+1}^3}{12}\right), \tag{2}$$

and for $\delta_j > 1$

$$\sigma_j = 2^{-\delta_j}\sigma_{j+1} + 1 - \frac{2^{-\delta_j} - 2^{-2\delta_j+1}}{\ln 2} - 2^{-2\delta_j}\frac{\sigma_{j+1}^2 \ln 2}{4} + 2^{-2\delta_j}R_j\left(\frac{\ln^2 2\sigma_{j+1}^3}{8}\right). \tag{3}$$

Proof. Consider

$$\begin{aligned} M - M = 0 &= \sum_{i=1}^{j} 2^{[\alpha_i]} + 2^{\alpha_{j+1}} - \left[\sum_{i=1}^{j-1} 2^{[\alpha_i]} + 2^{\alpha_j}\right] \\ &= 2^{[\alpha_j]} + 2^{\alpha_{j+1}} - 2^{\alpha_j} \\ 2^{\alpha_j} &= 2^{[\alpha_j]} + 2^{\alpha_{j+1}} = 2^{[\alpha_j]} + 2^{[\alpha_{j+1}] - [\alpha_j] + [\alpha_j] + \epsilon_{j+1}}. \end{aligned}$$

Next, we move to functional relations between σ_j and σ_{j+1}:

$$\begin{aligned} 2^{\epsilon_j} &= 2^{-\delta_j + \epsilon_{j+1}} + 1 \\ \Rightarrow 2^{1-\sigma_j} &= 2^{-\delta_j + 1 - \sigma_{j+1}} + 1 \\ \Rightarrow \ln(2^{1-\sigma_j}) &= \ln 2 - \sigma_j \ln 2 = \ln(2^{-\delta_j + 1 - \sigma_{j+1}} + 1). \end{aligned}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2024 doi:10.20944/preprints202401.0227.v11

Calculating for $\delta_j = 1$, we get:

$$\ln(2^{-\delta_j+1-\sigma_{j+1}}+1)\Big|_{\delta_j=1} = \ln(2^{-\sigma_{j+1}}+1)$$
$$= \ln 2 + \ln\left(1 - \frac{\sigma_{j+1}\ln 2}{2} + \frac{\sigma_{j+1}^2 \ln^2 2}{4} + F_j\left(\frac{\sigma_{j+1}^3}{12}\right)\right).$$

Continuing calculations for $\delta_j > 1$, we get:

$$\ln(2^{-\delta_j+1-\sigma_{j+1}}+1) = \ln\left(1 + 2^{-\delta_j+1} - 2^{-\delta_j+1}\frac{\sigma_{j+1}\ln 2}{2} + 2^{-\delta_j+1}F_j\left(\sigma_{j+1}^2 + 2^{-\delta_j+1}\right)\right)$$
$$= 2^{-\delta_j} - 2^{-2\delta_j+1} - 2^{-\delta_j}\frac{\sigma_{j+1}\ln 2}{2} + 2^{-2\delta_j}F_j\left(\sigma_{j+1}^2\right).$$

Thus, we obtain the final formulas. □

Theorem 2. *Let*

$$M = 3^n = 2^{[\alpha]+\{\alpha\}} = \sum_{i=1}^{n^*} \gamma_i 2^i,$$
$$1 - \{\alpha\} > 0.55, \quad n^* = \left[n\frac{\ln(3)}{\ln(2)}\right], \tag{4}$$

then

$$\sum_{\gamma_i=0} 1 \geq \frac{n^*}{2}.$$

Proof. Let

$$3^n = 2^\alpha \Rightarrow \alpha = \frac{n}{\ln(3)/\ln(2)} \Rightarrow 3^n = 2^{[\alpha]+\{\alpha\}}.$$

Using Theorem 1, we create a sequence

$$\epsilon_i,\ m_i,\ \epsilon_1 = \{\alpha\},$$
$$2^{\epsilon_1} = \sum_{k=0}^{i-1} 2^{[\alpha_k]-\alpha_1} + 2^{\alpha_i-\alpha_1}.$$

Assuming that the binary decomposition process, according to formula (1), stops at the j-th step. From this it immediately follows that the other terms of the decomposition are zeros and we immediately reach the truth of the statement of the Theorem. Therefore, we will consider the case when the generation of the decomposition according to formula (1) does not stop, and j reaches n. This means that all $\sigma_j > 0, j < n$

We conduct a more detailed analysis of the number of zeros and ones in our binary representation Introduce the following notations:
l- the number of zeros in the binary representation.
m- the number of ones in the binary representation.
n- the digit capacity of the binary decomposition and then
n=l+m.

$$\delta_j = 1, \alpha_j = 0, \beta_j = \left((1 - \frac{\ln 2\delta_{j+1}}{2})/2 + F_j\left(\frac{\sigma_{j+1}^2}{12}\right)\right)^{-1}$$

$$\delta_j > 1, \alpha_j = -2^{\delta_j}\left(1 - \frac{2^{-\delta_j} - 2^{-2\delta_j+1}}{\ln 2} + 2^{-\delta_j}R_j\left(\frac{\ln^2 2\sigma_{j+1}^3}{8} + \frac{2^{-2\delta_j+1}}{\ln 2}\right)\right), \beta_j = 2^{\delta_j}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2024 doi:10.20944/preprints202401.0227.v11

Solving the following system of equations

$$\sigma_{j+1} = \alpha_j + \beta_j \sigma_j$$

Conduct a series of transformations to understand the following steps.

$$\sigma_{n+1} = \alpha_n + \alpha_1 \prod_{k=0}^{n-2} \beta_{n-k} + \sum_{m=1}^{n-2} \beta_{n-m} \frac{\alpha_{n-m}}{\beta_{n-m}} \prod_{k=0}^{m-1} \beta_{n-k} + \sigma_1 \prod_{k=0}^{n-1} \beta_{n-k}$$

$$\sigma_{n+1} = \alpha_n + \beta_1 \frac{\alpha_1}{\beta_1} \prod_{k=0}^{n-2} \beta_{n-k} + \sum_{m=1}^{n-2} \beta_{n-m} \frac{\alpha_{n-m}}{\beta_{n-m}} \prod_{k=0}^{m-1} \beta_{n-k} + \sigma_1 \prod_{k=0}^{n-1} \beta_{n-k}$$

$$\sigma_{n+1} = \alpha_n + \frac{\alpha_1}{\beta_1} \prod_{k=0}^{n-1} \beta_{n-k} + \sum_{m=1}^{n-1} \frac{\alpha_{n-m}}{\beta_{n-m}} \prod_{k=0}^{m} \beta_{n-k} + \sigma_1 \prod_{k=0}^{n-1} \beta_{n-k} \tag{5}$$

Introduce the notations

$$\alpha_* = \inf_{0 \le i \le n} \frac{\alpha_i}{\beta_i}$$

$$\alpha^* = \sup_{0 \le i \le n} \frac{\alpha_i}{\beta_i}$$

$$A(m) = \sum_{k=1, \delta_j = 1}^{m} ln_2(\beta_j) + \sum_{k=1, \delta_j > 1}^{m} ln_2(\beta_j) = A_1(m) + A_2(m)$$

Note that δ_k, σ_k appear at points with coordinates $x(\delta_k), x(\sigma_k)$, $x(\delta_k) = x(\sigma_k)$ and by definition α_i

$$1 < \alpha_* < \alpha^* < 1.3$$

Thus, all possible variants with L-zeros will be determined by all possible sets of

$$(\delta_1, \delta_2 .... \delta_n)$$

With corresponding coordinates

$$(x(\delta_1), x(\delta_2) .... x(\delta_n))$$

$$m_* = \sum_{i=1, \delta_i > 1}^{n} \delta_i$$

Rewrite formula (5)

$$\frac{\sigma_{n+1}}{\prod_{k=0}^{n-1} \beta_{n-k}} = \frac{\alpha_n}{\prod_{k=0}^{n-1} \beta_{n-k}} + \alpha_1 \beta_1 + \sum_{m=1}^{n-1} \frac{\alpha_{n-m}}{\beta_{n-m}} \frac{1}{\prod_{k=m}^{n-1} \beta_{n-k}} + \sigma_1$$

$$\sigma_1 \ge \frac{\sigma_n}{2^{A(n)}} - \frac{\alpha_*}{2^{A(n)}} \sum_{i=1}^{n-1} 2^{A(i)}$$

To calculate the sum in the last inequality, we use the equations

$$2^k = 1 + \sum_{i=0}^{k-1} 2^i, \; 2^k + 2^l = 2^l \left(1 + \sum_{i=0}^{k-l-1} 2^i \right) = 2^l + \sum_{i=0}^{k-l-1} 2^{i+l} = 2^l + \sum_{i=l}^{k-1} 2^i$$

It is important to note that here k,l also have their coordinates $x(k), x(l)$ and all $i, l < i < k$, have coordinates $x(i)$ which are built on a uniform grid Thanks to these simple formulas and corresponding coordinates, we can calculate sums using integrals.

 doi:10.20944/preprints202401.0227.v11

$$I = \sum_{i=1}^{n} 2^{A(i)}$$

$$I(\gamma) = \int_0^{n-1} 2^{\gamma x} dx = I + R(n), \;\; \gamma = \frac{m_* + (\ln 2 + \epsilon) l}{n} >$$

where R(n) is the residual term which we can neglect for large n Where

$$L = n - l < m_*$$

- a given level of the number of zeros.

$$I(\gamma) = \frac{1 - 2^{A(n)}}{2\gamma \ln 2}$$

$$\alpha^* \frac{1 - 2^{-A(n)}}{2 \ln 2(1 + ln2 + \epsilon)} \leq \sigma_1, \;\; \frac{dI(\gamma)}{d\gamma} < 0 \Rightarrow$$

$$\alpha^* \frac{1 - 2^{-A(n)}}{2 \ln 2(1 + ln2 + \epsilon)} \leq \sigma_1, \;\; \frac{dI(\gamma)}{d\gamma} < 0 \Rightarrow$$

Note that the smaller γ the larger $I(\gamma)$ therefore to reach the given level L it is possible only with the corresponding σ_1 and to reach the level $L = n/2$ it is necessary to choose

$$0.55 = \frac{1.3}{2 \ln 2(1 + \ln 2 + \epsilon)} < \sigma_1$$

$$\Rightarrow L \geq n/2. \Rightarrow$$

The statement of the theorem is true. □

Proof. Let

$$3^n = 2^\alpha \Rightarrow \alpha = \frac{n}{\ln(3)/\ln(2)} \Rightarrow 3^n = 2^{[\alpha] + \{\alpha\}}.$$

Using Theorem 1, we create a sequence

$$\epsilon_i, \; m_i, \; \epsilon_1 = \{\alpha\},$$

$$2^{\epsilon_1} = \sum_{k=0}^{i-1} 2^{[\alpha_k] - \alpha_1} + 2^{\alpha_i - \alpha_1}.$$

Assume that the process of binary decomposition, according to formula (1), stopped at the j-th step. From which it immediately follows that the remaining terms of the decomposition are zeros and we immediately reach the truth of the Theorem's statement. Therefore, we will consider the case when the generation of decomposition according to formula (1) does not stop, and j reaches n. This means that all $\sigma_j > 0, j < n$

Let's conduct a more detailed analysis of the number of zeros and ones in our binary representation. Introduce the following notations:
l- number of zeros in the binary representation.
m- number of ones in the binary representation.
n- bit depth of the binary decomposition and then
n=l+m.

$$\delta_j = 1, \alpha_j = 0, \beta_j = \left((1 - \frac{\ln 2\delta_{j+1}}{2})/2 + F_j \left(\frac{\sigma_{j+1}^2}{12} \right)) \right)^{-1}$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2024 doi:10.20944/preprints202401.0227.v11

$$\delta_j > 1, \alpha_j = -2^{\delta_j}\left(1 - \frac{2^{-\delta_j} - 2^{-2\delta_j+1}}{\ln 2} + 2^{-\delta_j}R_j\left(\frac{\ln^2 2\sigma_{j+1}^3}{8} + \frac{2^{-2\delta_j+1}}{\ln 2}\right)\right), \beta_j = 2^{\delta_j}$$

Solving the following system of equations

$$\sigma_{j+1} = \alpha_j + \beta_j\sigma_j$$

we get

$$\sigma_{n+1} = \alpha_n + \sum_{m=1}^{n-1}\alpha_{n-m}\prod_{k=0}^{m-1}\beta_{n-k} + \sigma_1\prod_{k=0}^{n-1}\beta_{n-k}$$

Let's perform a series of transformations to understand the following steps.

$$\sigma_{n+1} = \alpha_n + \alpha_1\prod_{k=0}^{n-2}\beta_{n-k} + \sum_{m=1}^{n-2}\beta_{n-m}\prod_{k=0}^{m-1}\beta_{n-k}\frac{\alpha_{n-m}}{\beta_{n-m}} + \sigma_1\prod_{k=0}^{n-1}\beta_{n-k}$$

$$\sigma_{n+1} = \alpha_n + \beta_1\frac{\alpha_1}{\beta_1}\prod_{k=0}^{n-2}\beta_{n-k} + \sum_{m=1}^{n-2}\beta_{n-m}\prod_{k=0}^{m-1}\beta_{n-k}\frac{\alpha_{n-m}}{\beta_{n-m}} + \sigma_1\prod_{k=0}^{n-1}\beta_{n-k}$$

$$\sigma_{n+1} = \alpha_n + \sigma_1\frac{\alpha_1}{\beta_1}\prod_{k=0}^{n-1}\beta_{n-k} + \sum_{m=1}^{n-1}\frac{\alpha_{n-m}}{\beta_{n-m}}\prod_{k=0}^{m-1}\beta_{n-k} + \sigma_1\prod_{k=0}^{n-1}\beta_{n-k}$$

$$\sigma_{n+1} = \alpha_n + \sigma_1\frac{\alpha_1}{\beta_1}\prod_{k=0}^{n-1}\beta_{n-k} + \sum_{m=1}^{n-1}\frac{\alpha_{n-m}}{\beta_{n-m}}\prod_{k=0}^{m-1}\beta_{n-k} + \sigma_1\prod_{k=0}^{n-1}\beta_{n-k}$$

Introduce notations

$$\alpha_* = \inf_{0\le i\le n}\frac{\alpha_i}{\beta_i}$$

$$\alpha^* = \sup_{0\le i\le n}\frac{\alpha_i}{\beta_i}$$

$$A(m) = \sum_{k=1}^{m} ln_2(\beta_j) = \sum_{k=1,\delta_j=1}^{m} ln_2(\beta_j) + \sum_{k=1,\delta_j>1}^{m} ln_2(\beta_j)$$

$$A(m) = \sum_{k=1,\delta_j=1}^{m} ln_2(\beta_j) + \sum_{k=1,\delta_j>1}^{m}(\delta_j + 1) = A_1(m) + A_2(m)$$

$$m_*(t) = \inf_m\{A_2(m) - m \ge t\}$$

$$\nu_*(t) = \sum_{k=1}^{m_*(t)} ln_2(\beta_j) \ge m_*(t)(\ln 2 - \ln(1 - ln2\sigma_{j+1}/2)) \ge m_*(t)\ln 2$$

Note that δ_k, σ_k appear at points with coordinates $x(\delta_k), x(\sigma_k)$, $x(\delta_k) = x(\sigma_k)$ and by definition of α_i

$$1 < \alpha_* < \alpha^* < 1.3$$

Thus all possible variants with L-zeros will be determined by all possible sets of

$$(\delta_1, \delta_2 .... \delta_n)$$

With corresponding coordinates

$$(x(\delta_1), x(\delta_2) .... x(\delta_n))$$

Note that

$$k > m_* \Rightarrow \delta_k = 1,$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2024 doi:10.20944/preprints202401.0227.v11

$$\sigma_1 \geq \frac{\sigma_n}{2^{A(n)}} - \frac{\alpha_*}{2^{A(n)}} \sum_{i=1}^{n-1} 2^{A(i)}$$

To calculate the sum in the last inequality, we use the following formulas

$$2^k = 1 + \sum_{i=0}^{k-1} 2^i, \; 2^k + 2^l = 2^l \left(1 + \sum_{i=0}^{k-l-1} 2^i\right) = 2^l + \sum_{i=0}^{k-l-1} 2^{i+l} = 2^l + \sum_{i=l}^{k-1} 2^i$$

It is important to note that here k,l also have their coordinates $x(k), x(l)$ and all $i, l < i < k$, have coordinates $x(i)$ which are built on a uniform grid Thanks to these simple formulas and corresponding coordinates, we can calculate sums using integrals.

$$I = \sum_{i=1}^{m_*} 2^{A(i)} = \sum_{i=1}^{m_*-1} 2^i + R(n)$$

where R(n) is a residual term that we can neglect at large values

$$I(\gamma) = \int_0^{x(m*)-1} 2^{\gamma x} dx, \;\; \gamma = \frac{L + \nu_*}{x(m_*)} > \frac{L + m_*(L)\ln 2}{x(m_*(L))}, \;\; m_*(L) \geq n - L$$

Where L- given level of number of zeros.

$$I(\gamma) = \frac{1 - 2^{A(n)}}{2\gamma \ln 2}$$

$$\alpha^* \frac{1 - 2^{-A(n)}}{2\gamma \ln 2} \leq \sigma_1, \;\; \frac{dI(\gamma)}{d\gamma} < 0 \Rightarrow$$

$$\alpha^* \frac{1 - 2^{-A(n)}}{2\gamma \ln 2} \leq \sigma_1, \;\; \frac{dI(\gamma)}{d\gamma} < 0 \Rightarrow$$

Note that the smaller γ the larger $I(\gamma)$ therefore to reach a given level L only with corresponding σ_1 and to reach the level $L = n/2$ it is necessary to choose $\frac{1.3}{2(\ln 2 + \ln 2)} < \sigma_1$ and $x(m_*) = n/2 \Rightarrow L \geq n/2. \Rightarrow$ The statement of the theorem is true. □

Theorem 3. *Let*

$$a_n = \sum_{i=0}^{n} \gamma_i 2^i, \quad n > 1000, \quad \gamma_i \in \{0, 1\},$$

then

$$\exists j^* \in \{0, 1\}, \quad \textit{and} \quad a_{4n-j^*} < a_n.$$

Proof. Introduce operators defined as follows:

$$Pf = \frac{f}{2}, \quad Tf = 3f + 1, \quad Zf = 3f,$$

$$T_i \in \{P, T\}, \quad R_i \in \{Z, P\}.$$

Consider all possible scenarios of Collatz sequence behavior, which can be written in the following form:

$$a_{n+n} = T_1 T_2 \dots T_n a_n,$$

We need to estimate each $2n$-th term of the Collatz sequence based on the number of applied operators P, T, Z during n steps.

$$a_{n+n} = T_n T_{n-1} \dots T_1 a_n,$$

 doi:10.20944/preprints202401.0227.v11

Let a_n have m ones in its binary representation, then we count the number of applications of operator Z using the following formula:

$$m = \sum_{\substack{R_i=Z,\\ i\leq n}} 1,$$

and the number of applications of operator P using the following formula:

$$\sum_{\substack{R_i=P,\\ i\leq n}} 1 = m + n - m = n.$$

Since each application of Z is accompanied by operator P, and the number of applications of operator P corresponds to the number of zeros in a_n, which equals $n - m$. According to the rules of Collatz, after n steps we have:

$$a_{n+n} = \frac{3^m}{2^n} a_n + T_n T_{n-1} \dots T_1 1 = \frac{3^m}{2^n} a_n + B_n,$$

$$B_n \leq 2^{-n+m} \sum_{j=1}^{m} \frac{3^j}{2^j} a_n < 2^{-n+m} \cdot 3^m / 2^m \cdot a_n \leq 2^{-2n+1} \cdot 3^m \cdot a_n.$$

According to the last formula, we see that the growth of each term of the sequence depends on the number of ones in the binary representation. Next, we will show that a large number of ones at the $2n$-th step leads to an increase in the number of zeros at the $3n$-th step for binary representation according to the previous theorems, from which it follows that subsequent terms of the sequence decrease:

$$a_{2n} = 3^m a_n \cdot 2^{-n} + B_n = 3^m + 3^m (a_n - 2^n) + B_n,$$

Repeating the reasoning of Theorem 2, consider the equation

$$2^x = a_{2n} = 3^m a_n \cdot 2^{-n} + B_n = 3^m + 3^m (a_n - 2^n) \cdot 2^{-n} + B_n,$$

$$x \ln 2 = m \ln(3) + \ln\left(1 + (a_n - 2^n) \cdot 2^{-n} + B_n \cdot 3^{-m}\right),$$

From the last equation, to apply the results of theorem 2, we need $\sigma_1 > \frac{1}{2\ln 2}$. To satisfy the last inequality, consider $m_j = m - j$, $\theta = (a_n - 2^n) \cdot 2^{-n}$,

$$\{x\} = \min_{j<10} \left\{ \frac{(m-j)\ln(3)}{\ln(2)} + \frac{\ln(1+\theta)}{\ln 2} + F_j\left(\frac{1}{2^n \ln 2}\right) \right\},$$

Consider $p = (m-j)\frac{\ln 3}{\ln 2} = (2k+l)1.5849625007\dots$, $\epsilon = 1.5849625007\dots - 1.5$, we get

$$p = (2k+l)(1.5 + \epsilon + \frac{\ln(1+\theta)}{\ln 2}) = 3k + (2k+l)\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2},$$

$$\{p\} = \{1.5 \cdot l + (2k+l)\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\} = \{1.5 \cdot l + \{(2k+l)\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\}\},$$

Choosing l from even numbers less than 10, if inequalities $0 \leq \{(2k)\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\} \leq 0.5$, are true

$$\{p\} = \{2k\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\} = \{2k\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\},$$

Choosing l from odd numbers less than 10, if inequalities $0.5 < \{2k\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\} < 1$, are true

$$\{p\} = \{2k\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\} = \{0.5 + (2k+l)\cdot\epsilon\},$$

 doi:10.20944/preprints202401.0227.v11

Using $\epsilon < 0.1$, also satisfy the condition $\sigma_1 = 1 - \{x\} > \frac{1}{2\ln 2}$.

$$m^* \text{ number of non-zero } \gamma_i,$$

According to theorem 2 we get

$$m^* \le n/2 + (n - j^*) \cdot \ln 3/\ln 2/2,$$

According to our application of Collatz rules, we have an element a_{4n-j^*}, and the order of its binary representation is

$$n_2 = n + (n - j^*) \cdot \ln 3/\ln 2/2,$$

After $3n - j^*$ steps of applying Collatz rules we have

$$a_{4n-j^*} = \frac{3^{m^*}}{2^{2n-j^*}} a_{2n} + T_{3n-j^*} T_{3n-1-j^*} \dots T_1 1 = \frac{3^{m^*}}{2^{2n}} a_{2n} + B_{3n},$$

$$a_{4n-j^*} = \frac{3^{m^*}}{2^{2n}} a_{2n} + T_{3n-j^*} T_{3n-j^*-1} \dots T_1 1 = \frac{3^{m^*}}{2^{2n}} \left(\frac{3^m}{2^{n-j^*}} a_n + B_n \right) + B_{3n-j^*},$$

$$a_{4n-j^*} = 3^{m^*+m} \cdot 2^{-3n-j^*} a_n + 3^{m^*} \cdot 2^{-2n-j^*} B_n + B_{3n-j^*},$$

$$a_{4n-j^*} \le q_1 \cdot a_n,$$

By definition of m^*, l^*, B_n we get

$$q_1 < 1,$$

Using $n > 1000$, it follows that $q_1 < 1 \Rightarrow a_{4n-j^*} < a_n$. □

Theorem 4. *Let*

$$a_n = \sum_{i=0}^{n} \gamma_i 2^i, \quad n > 1000, \quad \gamma_i \in \{0, 1\},$$

then

$$\exists j^* < 0.1n, \quad \text{and} \quad a_{4n-j^*} < a_n.$$

Proof. Let's introduce operators defined by the formulas

$$Pf = \frac{f}{2}, \quad Tf = 3f + 1, \quad Zf = 3f,$$

$$T_i \in \{P, T\}, \quad R_i \in \{Z, P\}.$$

Consider all possible scenarios of the behavior of the Collatz sequence, which can be written in the following form:

$$a_{n+n} = T_1 T_2 \dots T_n a_n,$$

It is necessary to calculate an estimate for each $2n$-th member of the Collatz sequence based on the number of P, T, Z operators applied during n steps.

$$a_{n+n} = T_n T_{n-1} \dots T_1 a_n,$$

Let a_n have m units in its binary representation, then calculate the number of applications of the Z operator by the following formula:

$$m = \sum_{\substack{R_i = Z, \\ i \le n}} 1,$$

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 February 2024 doi:10.20944/preprints202401.0227.v11

and calculate the number of applications of the P operator by the following formula:

$$\sum_{\substack{R_i=P,\\ i\leq n}} 1 = m + n - m = n.$$

Since each application of Z is accompanied by the P operator, and the number of applications of the P operator corresponds to the number of zeros in a_n, which is equal to $n-m$. According to the rules of Collatz after n steps, we have:

$$a_{n+n} = \frac{3^m}{2^n}a_n + T_n T_{n-1}\dots T_1 1 = \frac{3^m}{2^n}a_n + B_n,$$

$$B_n \leq 2^{-n+m}\sum_{j=1}^{m}\frac{3^j}{2^j}a_n < 2^{-n+m}\cdot 3^m/2^m\cdot a_n \leq 2^{-2n+1}\cdot 3^m\cdot a_n.$$

According to the last formula, we see that the growth of each member of the sequence depends on the number of units in the binary representation. Next, we will show that a large number of units on the $2n$-th step leads to an increase in the number of zeros in the $3n$-th step for the binary representation according to previous theorems, hence the reduction of subsequent members of the sequence:

$$a_{2n} = 3^m a_n\cdot 2^{-n} + B_n = 3^m + 3^m(a_n - 2^n) + B_n,$$

Repeating the reasoning of Theorem 2, consider the equation

$$2^x = a_{2n} = 3^m a_n\cdot 2^{-n} + B_n = 3^m + 3^m(a_n - 2^n)\cdot 2^{-n} + B_n,$$

$$x\ln 2 = m\ln(3) + \ln\left(1 + (a_n - 2^n)\cdot 2^{-n} + B_n\cdot 3^{-m}\right),$$

From the last equation, in order to apply the results of theorem 2, we need $\sigma_1 = 1 - \{x\} > 0.5$. To fulfill the last inequality, consider $m_j = m - j, \theta = (a_n - 2^n)\cdot 2^{-n}$,

$$\{x\} = \min_{j\in\{0,1\}}\left\{\frac{(m-j)\ln(3)}{\ln(2)} + \frac{\ln(1+\theta)}{\ln 2} + F_j\left(\frac{1}{2^n\ln 2}\right)\right\},$$

Consider $p = (m-j)\frac{\ln 3}{\ln 2} = (2k+l)1.5849625007\dots, \epsilon = 1.5849625007\dots - 1.5$, we get

$$p = (2k+l)(1.5 + \epsilon + \frac{\ln(1+\theta)}{\ln 2}) = 3k + (2k+l)\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2},$$

$$\{p\} = \{1.5\cdot l + (2k+l)\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\} = \{1.5\cdot l + \{(2k+l)\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\}\},$$

Choosing $l = 0$, if the inequalities $0 \leq \{(2k)\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\} \leq 0.5$ are true,

$$\{p\} = \{2k\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\} = \{2k\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\},$$

Choosing $l = 1$, if the inequalities $0.5 < \{2k\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\} < 1$ are true,

$$\{p\} = \{2k\cdot\epsilon + \frac{\ln(1+\theta)}{\ln 2}\} = \{0.5 + (2k+l)\cdot\epsilon\},$$

Using $\epsilon < 0.1$, we also satisfy the condition $\sigma_1 = 1 - \{x\} > 0.51$.

$$m^* \text{ is the number of non-zero } \gamma_i,$$

According to theorem 2 we get

$$m^* \leq n/2 + (n - j^*) \cdot \ln 3/ \ln 2/2,$$

According to our application of the Collatz rules, we have the element a_{4n-j^*}, and the order of its binary representation is

$$n_2 = n + (n - j^*) \cdot \ln 3/ \ln 2/2,$$

After $3n - j^*$ steps of applying the Collatz rules, we have

$$a_{4n-j^*} = \frac{3^{m^*}}{2^{2n-j^*}} a_{2n} + T_{3n-j^*} T_{3n-1-j^*} \dots T_1 1 = \frac{3^{m^*}}{2^{2n}} a_{2n} + B_{3n},$$

$$a_{4n-j^*} = \frac{3^{m^*}}{2^{2n}} a_{2n} + T_{3n-j^*} T_{3n-j^*-1} \dots T_1 1 = \frac{3^{m^*}}{2^{2n}} \left(\frac{3^m}{2^{n-j^*}} a_n + B_n \right) + B_{3n-j^*},$$

$$a_{4n-j^*} = 3^{m^*+m} \cdot 2^{-3n-j^*} a_n + 3^{m^*} \cdot 2^{-2n-j^*} B_n + B_{3n-j^*},$$

$$a_{4n-j^*} \leq q_1 \cdot a_n,$$

By definition of m^*, l^*, B_n we get

$$q_1 < 1,$$

Using $n > 1000$, implies $q_1 < 1 \Rightarrow a_{4n-j^*} < a_n$. □

Theorem 5. *Let*

$$a_n = \sum_{i=0}^{n} \gamma_i 2^i, \quad n > 1000, \quad \gamma_i \in \{0, 1\},$$

then for a_n the Collatz conjecture is true.

Proof. The proof follows from Theorems 1-3. □

Proof. Proof follows from theorem 1-3

6. Conclusions

Our assertion proves that after $3n$ steps, a sequence with an initial binary length of n arrives at a number strictly smaller than the initial one, from which the solution to the Collatz conjecture follows. This is because by applying this process n times, we are guaranteed to arrive at 1.

References

1. O'Connor, J.J.; Robertson, E.F. (2006). "Lothar Collatz". St Andrews University School of Mathematics and Statistics, Scotland.
2. Tao, Terence (2022). "Almost all orbits of the Collatz map attain almost bounded values". Forum of Mathematics, Pi. 10: e12. arXiv:1909.03562. doi:10.1017/fmp.2022.8. ISSN 2050-5086.