Article

Not peer-reviewed version

Collatz Conjecture

Asset Durmagambetov * and Aniyar Durmagambetova

Posted Date: 11 January 2024

doi: 10.20944/preprints202401.0227.v4

Keywords: binary representation; Collatz conjecture

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2024 doi:10.20944/preprints202401.0227.v4

Article

Collatz Conjecture

Asset Durmagambetov * and Aniyar Durmagambetova

Faculty of Mathematics, L.N. Gumilyov Eurasian National University, Kazakhstan; 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:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2024 doi:10.20944/preprints202401.0227.v4

- Divide it by two if it's even.
- 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*

$$M \in \mathbb{N}, \ \ [\alpha_j] - [\alpha_{j+1}] = \delta_j > 0, \ \ \epsilon_1 < 1/2,$$

$$\alpha_j = [\alpha_j] + \epsilon_j, \epsilon_j < 1,$$

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

Then, for $\delta_j = 1$

$$\sigma_{j+1} \ln 2 = \frac{2\sigma_j \ln 2}{1 - \sigma_{j+1} \ln 2} + o(\sigma_{j+1}^2/4)$$

and for $\delta_j > 1$

$$\sigma_{j+1} \ln 2 = -2^{\delta_j - 1} \frac{\ln 2 - 2^{-\delta_j + 1}}{1 - \sigma_{j+1} \ln 2/2} + 2^{\delta_j - 1} \sigma_j \ln 2 \frac{1}{1 - \sigma_{j+1} \ln 2/2} + o({\sigma_{j+1}^2}^2 + 2^{-\delta_j + 1})$$

Proof.

$$0 = M - M = \sum_{i=1}^{j} 2^{[\alpha_i]} + 2^{\alpha_{j+1}} - \left[\sum_{i=1}^{j-1} 2^{[\alpha_i]} + 2^{\alpha_j} \right] = 2^{[\alpha_j]} + 2^{\alpha_{j+1}} - 2^{\alpha_j}$$

$$2^{\alpha_j} = 2^{[\alpha_j]} + 2^{\alpha_{j+1}}, \ \ 2^{[\alpha_j] + \epsilon_j} = 2^{[\alpha_j]} + 2^{\alpha_{j+1}} = 2^{[\alpha_j]} + 2^{[\alpha_{j+1}] - [\alpha_j] + [\alpha_j] + \epsilon_{j+1}}$$

$$2^{\epsilon_j} = 2^{-\delta_j + \epsilon_{j+1}} + 1 \Rightarrow 2^{1 - \sigma_j} = 2^{-\delta_j + 1 - \sigma_{j+1}} + 1 \Rightarrow$$

$$\ln(2^{1-\sigma_j}) = \ln 2 - \sigma_j \ln 2 = \ln(2^{-\delta_j + 1 - \sigma_{j+1}} + 1)$$

Calculating for $\delta_j = 1$

$$\ln(2^{-\delta_j + 1 - \sigma_{j+1}} + 1)|_{\delta_j = 1} = \ln(2^{-\sigma_{j+1}} + 1) = \ln(1 - \sigma_{j+1} \ln 2 + \sigma_{j+1}^2 \ln^2/2) + 1 + o(\sigma_{j+1}^2/4)$$

$$\ln(2 - \sigma_{j+1} \ln 2 + \sigma_{j+1}^2 \ln 2^2/2) = \ln 2 + \ln(1 - \sigma_{j+1} \ln 2/2 + \sigma_{j+1}^2 \ln^2/4 + o(\sigma_{j+1}^2/4))$$

$$\ln(2^{-\sigma_{j+1}} + 1) = \ln 2 - \ln 2\sigma_{j+1}/2 + \ln^2 2\sigma_{j+1}^2/2 + o(\sigma_{j+1}^2/4)$$

$$\ln 2 - \sigma_j \ln 2 = \ln 2 - \ln 2\sigma_{j+1}/2 + \ln^2 2\sigma_{j+1}^2/4 + o(\sigma_{j+1}^2/4)$$

$$\sigma_{j+1} \ln 2 = \frac{2\sigma_j \ln 2}{1 - \sigma_{j+1} \ln 2/2} + o(\sigma_{j+1}^2/4)$$

 doi:10.20944/preprints202401.0227.v4

Continuing the calculation for $\delta_j > 1$, we get

$$\ln(2^{-\delta_j+1-\sigma_{j+1}}+1) = \ln(2^{-\delta_j+1}2^{-\sigma_{j+1}}+1) =$$

$$\ln(1+2^{-\delta_j+1}+2^{-\delta_j+1}[-\sigma_{j+1}\ln 2+\sigma_{j+1}^2\ln 2^2/2]+o(\sigma_{j+1}^2/4+2^{-\delta_j+1})) =$$

$$2^{-\delta_j+1}+2^{-\delta_j+1}[-\sigma_{j+1}\ln 2+\sigma_{j+1}^2\ln 2^2/2]+o({\sigma_{j+1}^2}^2+2^{-\delta_j+1}) \Rightarrow$$

$$\ln 2-\sigma_j\ln 2 = 2^{-\delta_j+1}+2^{-\delta_j+1}[-\sigma_{j+1}\ln 2+\sigma_{j+1}^2\ln 2^2/2]+o({\sigma_{j+1}^2}^2+2^{-\delta_j+1})$$

$$\sigma_{j+1}\ln 2 = -2^{\delta_j-1}\frac{\ln 2-2^{-\delta_j+1}}{1-\sigma_{j+1}\ln 2/2}+2^{\delta_j-1}\sigma_j\ln 2\frac{1}{1-\sigma_{j+1}\ln 2/2}+o({\sigma_{j+1}^2}^2+2^{-\delta_j+1})$$

□

Theorem 2. *Let*

$$M = 3^n = 2^{[\alpha]+\{\alpha\}} = \sum_{i=1}^{n^*}\gamma_i 2^i,$$

$$1-\{\alpha\} > \frac{1}{2}, \quad n^* = n\times\left[\frac{\ln(3)}{\ln(2)}\right], \tag{1}$$

then

$$\sum_{\gamma_i=0} 1 \geq \frac{n^*}{2}-5.$$

Proof. Suppose

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

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 $\delta_j = 1$, and in such cases, we have

$$\Rightarrow \sigma_{j+1}\ln 2 = \frac{2\sigma_j\ln 2}{1-\sigma_{j+1}\ln 2/2}+o\left(\ln 2\frac{\sigma_{j+1}^2}{4}\right)$$

$$2^{-1}\sigma_{j+1}\ln 2 = \frac{\sigma_j\ln 2}{1-\sigma_{j+1}\ln 2/2}+2^{-1}\cdot o\left(\ln 2\frac{\sigma_{j+1}^2}{4}\right)$$

$$2^{-1}\sigma_j\ln 2 = \frac{\sigma_{j-1}\ln 2}{1-\sigma_j\ln 2/2}+2^{-1}\cdot o\left(\ln 2\frac{\sigma_j^2}{4}\right)$$

Repeating j times, we get

$$2^{-j}\sigma_{j+1}\ln 2 = \frac{\sigma_1\ln 2}{\prod_{k=1}^{j}(1-\sigma_{k+1}\ln 2/2)}+\sum_{k=1}^{j}2^{-k}\cdot o\left(\ln 2\frac{\sigma_{k+1}^2}{4}\right)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2024 doi:10.20944/preprints202401.0227.v4

From Theorem 1 and the conditions of this theorem, we obtain

$$(1-\alpha)\ln 2 = \sigma_1 \ln 2 < o\left(\ln 2 \frac{\sigma_{k+1}^2}{4}\right)$$

From the smallness of α and the last estimate,

$$\Rightarrow \exists \delta_j > 1$$

Now let's consider the question of the number of $\delta_j > 1$ elements. For this, we consider the general case where there are $\delta_j > 1$ and $\delta_k = 1$. Let's introduce the following notations:

$$\text{for } \delta_j = 1 : \alpha_j = 0,\ \beta_j = \frac{1}{1-\sigma_{j+1}\ln 2/2}$$

$$\text{for } \delta_j > 1 : \alpha_j = -2^{\delta_j - 1}\frac{\ln 2 - 2^{-\delta_j+1}}{1-\sigma_{j+1}\ln 2/2},\ \beta_j = \frac{2^{\delta_j-1}}{1-\sigma_{j+1}\ln 2/2}$$

$$\ln 2\sigma_{j+1} = \alpha_j + \ln 2\beta_j\sigma_j$$

Solving the last equation for σ_j, we get

$$\sigma_{j+1}\ln 2 = \alpha_j + \sum_{m=1}^{j-1}\alpha_{j-m}\prod_{l=1}^{m}\beta_{j-l+1} + \ln 2\prod_{l=0}^{j}\beta_{j-l}\sigma_1$$

$$\frac{\sigma_{j+1}\ln 2}{\prod_{l=0}^{j-1}\beta_{j-l}} = \frac{\alpha_j}{\prod_{l=0}^{j-1}\beta_{j-l}} + \sum_{m=0}^{j-1}\frac{\alpha_{j-m}}{\prod_{l=m+1}^{j-1}\beta_{j-l}} + \ln 2\sigma_1$$

According to the condition $\alpha_j \leq 0$, we have

$$\frac{\sigma_{j+1}\ln 2}{\prod_{l=0}^{j-1}\beta_{j-l}} - \sum_{m=0}^{j-1}\frac{\alpha_{j-m}}{\prod_{l=m}^{j-1}\beta_{j-l}} \geq \sigma_1 \ln 2$$

Denoting by

$$T = \inf_{k<j}\{\beta_k\},$$

we obtain

$$\frac{\sigma_n \ln 2}{T^n} - \sum_{m=0}^{j-1}\frac{\alpha_{j-m}}{\prod_{l=m}^{j-1}\beta_{j-l}} \geq \sigma_1 \ln 2$$

We come to evaluating the most important term:

$$I_j = -\sum_{m=1}^{j-1}\frac{\alpha_{j-m}}{\prod_{l=m}^{j-1}\beta_{j-l}}$$

The sum is defined by non-zero terms α_{j-m}:

$$\xi_* \leq \xi_{j-m} = \frac{\ln 2 - 2^{-\delta_{j-m}+1}}{1-\sigma_{j+1-m}\ln 2/2} \leq \xi^*, \quad \alpha_{j-m} = -2^{\delta_{j-m}-1}\xi_{j-m}$$

$$\mu_* \leq \mu_j = \frac{1}{1-\sigma_{j+1}\ln 2/2} \geq \mu^*, \quad \beta_j = 2^{\delta_j - 1}\mu_j$$

 doi:10.20944/preprints202401.0227.v4

$$I_j = \sum_{m=1}^{j-1} \frac{2^{\delta_{j-m}-1}\xi_{j-m}}{\prod_{l=m}^{j-1} 2^{\delta_{j-l}-1}\mu_j}, \quad I_j = \sum_{m=1}^{j-1} \frac{\xi_{j-m}}{\prod_{l=m+1}^{j-1} 2^{\delta_{j-l}-1}\mu_j}$$

$$I_*(n) = \frac{\xi_*}{\mu^*}\sum_{m=1}^{n-1} 2^{A_m}, \quad A_m = \sum_{l=m+1}^{n-1}[-\delta_{j-l}+1], \quad S^*(n) = \frac{\xi^*}{\mu_*}\sum_{m=1}^{n-1} 2^{A_m},$$

$$S(n) = 2^{A_n}\sum_{m=1}^{n-1} 2^{A_m-A_n}, \quad B_n = A_m - A_n$$

Consider the cases $B_m = i \le \frac{n}{2}$ according to the definition B_n By definition, each Collatz sequence corresponds to a set A_m; therefore, by iterating over all the sets, we iterate over all corresponding variants of the Collatz sequence. Consider now all possible variants of reaching the maximum level $B_{m_*} = i$, due to the monotonicity of the sequence $B_{m_*+k} = i, k > 0$ and for each variant, calculate

$$S(n) = 2^{A_n}\sum_{m=1}^{m_*} 2^{B_m} + 2^{A_n}\sum_{m=m_*}^{n} 2^{B_m},$$

$$S(n) = 2^{-i}\sum_{m=1}^{m_*} 2^{m\frac{i}{m_*}} + 2^{i}\sum_{m=m_*}^{n} 2^{i},$$

$$J(n) = 2^{-i}\sum_{m=1}^{n} 2^{m\frac{i}{m_*}} \le S(n)$$

$$J(n) = n\frac{2^{-i}(2^i - 1 + O(1))}{i\ln 2}$$

$$\frac{n}{i\ln 2} \le \sigma_1, \quad \Rightarrow \text{the theorem statement}$$

□

Theorem 3. *Let*

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

Then

$$\exists j^* < 0.1n$$

and

$$a_{4n-j^*} < a_n.$$

Proof. Let's introduce operators defined by the formulas

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

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

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

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

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

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2024 doi:10.20944/preprints202401.0227.v4

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

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

and count the number of P operator applications by the formula:

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

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

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

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

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 in the binary representation according to previous theorems, from which it follows that subsequent terms of the sequence decrease:

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

Repeating the reasoning of Theorem 2, we consider the equation

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

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

From the last equation, to apply the results of Theorem 2, we need $\sigma_1 = 1 - \{x\} > 0.5$. To meet the last inequality, consider $m_j = m - j$,

$$\{x\} = \inf_{j < 0.1 \times n} \left\{ \frac{(m-j)\ln(3)}{\ln(2)} + \frac{\ln(1+\theta)}{\ln 2} + O\left(\frac{1}{2^n \ln 2}\right) \right\}.$$

Depending on the behavior of $\gamma_{n-1}, \gamma_{n-2}, \dots$ we can always choose a variant where the fractional part of x meets the conditions of Theorem 2. Denoting

m^* is the number of non-zero γ_i,
l^* is the number of zero γ_i,

according to Theorem 2 we get

$$m^* \leq (2n - j^*) \frac{\ln 3}{\ln 2} / 2 + 5,$$
$$l^* \geq (2n - j^*) \frac{\ln 3}{\ln 2} / 2 - 5.$$

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

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

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 11 January 2024 doi:10.20944/preprints202401.0227.v4

$$a_{4n-j^*} = 3^{m^*+m}2^{-3n-j^*}a_n + 3^{m^*}2^{-2n-j^*}B_n + B_{3n-j^*}.$$
$$B_{3n-j^*} = 3^{m^*}2^{-2n} - j^*B_n = 3^{m^*+m}2^{-2n-n-j^*}a_n.$$
$$a_{4n-j^*} \leq q_1 \times a_n.$$

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

$$q_1 < 1.$$

Using $n > 1000 \Rightarrow q_1 < 1 \Rightarrow a_{4n} < a_n$. □

Theorem 4. *Let*

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

then for a_n Collatz conjecture is true

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.