

Article

Not peer-reviewed version

---

# Collatz Conjecture

---

[Asset Durmagambetov](#)\*

Posted Date: 8 January 2024

doi: 10.20944/preprints202401.0227.v2

Keywords: binary representation; Collatz conjecture



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Collatz Conjecture

Asset Durmagambetov <sup>1,\*</sup>, Aniyar Durmagambetova <sup>2</sup>

<sup>1</sup> aset.durmagambet@gmail.com

<sup>2</sup> 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.
- Triple it and add one if it's odd.

This operation is mathematically defined as:

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

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*

$$\begin{aligned} x \in N, [\alpha_j] - [\alpha_{j+1}] &= \delta_j > 0, \epsilon_1 < 1/2, \\ \alpha_j &= [\alpha_j] + \epsilon_j, \epsilon_j < 1, \\ x &= \sum_{i=1}^{j-1} 2^{[\alpha_i]} + 2^{\alpha_j}, \quad x = \sum_{i=1}^j 2^{[\alpha_i]} + 2^{\alpha_{j+1}}, \quad \sigma_j = 1 - \epsilon_j \end{aligned}$$

Then

as  $\delta_j = 1$

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

as  $\delta_j > 1$

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

**Proof.**

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

Computing as  $\delta_j = 1$

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

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

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

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

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

Repeating computing as  $\delta_j > 1$  we get

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

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

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

□

**Theorem 2.** Let

$$x \in \mathbb{N}, \quad \alpha_j = [\alpha_j] \quad x = \sum_{i=1}^{j-1} 2^{[\alpha_i]} + 2^{\alpha_j}$$

Then the number of zeros in the binary representation  $C_z$  is calculated by the following formula

$$C_z = \sum_{i=1}^{j-1} [\delta_i - 1] + \alpha_j - 1$$

**Proof.**

$$C_z = \sum_{i=1}^{j-1} [\alpha_i - \alpha_{i+1} - 1] + \alpha_j - 1$$

By definition  $\delta_i$ 

$$C_z = \sum_{i=1}^{j-1} [\delta_i - 1] + \alpha_j - 1$$

□

Let's introduce  $\mu_k, \nu_k$  for  $x = \sum_{i=0}^n \gamma_i 2^i$  by following rules

$$\gamma_k + \gamma_{k+1} = 1, \quad \gamma_{k+\mu_k} + \gamma_{k+\mu_k+1} = 1, \quad \prod_{i=k+1}^{i=\mu_k} \gamma_i = 1;$$

$$\gamma_j + \gamma_{j+1} = 1, \quad \gamma_{j+\nu_j} + \gamma_{j+\nu_j+1} = 1, \quad \nu_j = \sum_{i=j+1}^{i=\nu_j} (1 - \gamma_i)$$

another words

 $\mu_k$ , is count of ones starting at point k with no zeros in between until the first zero or until the end of the sequence $\nu_j$  is count of zeros starting at point j with no ones in between until the first zero or until the end of the sequence**Theorem 3.** Let

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

$$\{\alpha\} > \ln 2, \quad n^* = n * [\ln(3) / \ln(2)] \quad (1)$$

then

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

**Proof.**

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

Using Theorem 1, we create a sequence

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

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

Suppose

$$\sum_{\gamma_i=0} 1 = 0$$

then by Theorem 1

$$\begin{aligned} \Rightarrow \sigma_{j+1} \ln 2 &= \frac{2\sigma_j \ln 2}{1 - \sigma_{j+1} \ln 2} + o(\ln 2\sigma_{j+1}^2/4) \Rightarrow \\ 2^{-1}\sigma_{j+1} \ln 2 &= \frac{\sigma_j \ln 2}{1 - \sigma_{j+1} \ln 2} + 2^{-1} * o(\ln 2\sigma_{j+1}^2/4) \end{aligned}$$

After repeating j times we get

$$2^{-j}\sigma_{j+1} \ln 2 = \frac{\sigma_1 \ln 2}{\prod_1^j (1 - \sigma_{k+1} \ln 2/2)} + \sum_1^j 2^{-k} * o(\ln 2\sigma_{k+1}^2/4)$$

By Theorems (1-2) and condition of the current Theorem proceed

$$\ln 2/2 < \sigma_1 \ln 2 < o(\ln 2\sigma_{k+1}^2/4)$$

immediately

$$\Rightarrow \sum_{\gamma_i=0} 1 > 0$$

Let's introduce

$$\text{as } \delta_k = 1 : \alpha_k = 0, \beta_k = \frac{1}{1 - \sigma_{j+1} \ln 2}$$

$$\text{as } \delta_k > 1 : \alpha_k = -2^{\delta_j-1} \frac{\ln 2 - 2^{-\delta_j-1}}{1 - \sigma_{j+1} \ln 2/2}, \beta_k = \frac{2^{\delta_j-1}}{1 - \sigma_{j+1} \ln 2/2}$$

$$\ln 2\sigma_{k+1} = \alpha_k + \beta_k \ln 2\sigma_k$$

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

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

By condition the theorem

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

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

Suppose  $\forall i \in (1, n) \delta_i = 2 \Rightarrow$

$$\frac{\sigma_n \ln 2}{\prod_{l=0}^{l=n-1} \beta_{n-l}} + \sum_{m=0}^{m=n-1} \frac{\ln 2 - 1/2}{2^m} > \sigma_1 \ln 2 \Rightarrow$$

$$2(\ln 2 - 1/2) > \sigma_1 \Rightarrow$$

$\exists \delta_j > 2$

In common case

$$\frac{\sigma_n \ln 2}{\prod_{l=0}^{l=n-1} \beta_{n-l}} + \sum_{m=0}^{m=n-1} \frac{\ln 2 - \frac{2}{2^{\delta_m}}}{2^{\sum_{m=0}^{m=n-1} \beta_m}} > \sigma_1 \ln 2 \Rightarrow$$

and we see in case  $\delta_m > 2$  With the growth of  $\delta_m > 2$ , we see an exponential growth of the denominator, so the influence of  $\delta_m > 2$  terms with large values does not have a significant effect on the estimate of the left side of the inequality. The reason is the accumulation of the  $\beta_m$  value corresponding to  $\delta_m = 1$ . Therefore, the effect, taking into account  $\delta_m > 2$  with large values, is insignificant for estimating the left part of the inequality.  $\Rightarrow$  statement of Theorem

□

Theorem 4. Let

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

then

$$a_{8n} < a_n$$

Proof. Let's introduce operators defined formulas

$$Pf = f/2, \quad Tf = 3f + 1, \quad Zf = 3f$$

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

Let's consider all possible scenarios of the behavior of the Syracuse sequence, the same possible scenarios can be written in the following form

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

We need to calculate an estimate for every  $2n$ -th member of the Collatz sequence based on the number of applied operators  $P, T, Z$  over  $n$  step.

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

Let  $a_n$  have  $m$  ones in binary representation, then count the number of applications of the  $Z$  operator in the following formula.

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

and count the number of applications of the  $P$  operator in the following formula.

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

Because, each application  $Z$  is accompanied by an operator  $P$  and the number of applications of the operator  $P$  according to the zeros  $a_n$  has a corresponding  $n-m$  By rules of Collatz we have after  $n$  steps

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

$$B_n \leq \sum_{j=1, m} 3^j / 2^j < 23^m / 2^m \leq 3(3/2)^m 2^{-n} a_n$$

$$B_n \leq 3(3/4)^m 2^{m-n} a_n$$

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 ones at the \* step leads to an increase in the number of zeros in the binary representation according to the previous theorems. Where will the decrease in the following terms of the sequence follow

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

$$a_{2n} = \sum_{i=0}^{[\alpha_1]} \gamma_i 2^i, \quad \gamma_i \in \{0, 1\}, \quad \alpha_1 = m * \ln 3 / \ln 2 + \ln (2^{-n} a_n)$$

Let

$m^*$  is count of non zeros of  $\gamma_i$

$l^*$  is count of zeros of  $\gamma_i$

by theorem 2 we will have

$$m^* \leq [\alpha_1] / 2 + 5 = [m \ln 3 / \ln 2] / 2 + 5$$

$$l^* \geq [\alpha_1] / 2 - 5 = [m \ln 3 / \ln 2] / 2 - 5$$

After  $[\alpha_1]$  steps applying rules of Collatz we have

$$a_{2n+[\alpha_1]} \leq 3^m 3^{\alpha_1/2} 2^{-\alpha_1} 2^5 * 3^5 (a_n * 2^{-n} + B_n) = 3^m q_1 * a_n$$

where

$$q_1 = 3^m 3^{\alpha_1/2} 2^{-\alpha_1} 2^5 * 3^5$$

Repeating the process 3 times and using  $n > 1000 \Rightarrow q_3 < 1 \Rightarrow a_{8n} < a_n$  Theorem 5. Let

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

then for  $a_n$  Collatz conjecture is true Proof. Proof follows from theorem 1-4

## 6. Conclusions

Our assertion proves that after  $2n$  of steps the sequence comes to a number less than the start one, from which follows the solution of the Collatz conjecture.

## 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.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.