Article

Analyzing Collatz Conjecture Using the Mathematical Complete Induction Method

Mercedes Orús-Lacort [1] and Christophe Jouis [2,3]

[1] Online Teacher of College Mathematics: Universitat Oberta de Catalunya (Rambla del Poblenou 156, 08018 Barcelona – Spain), Universidad Nacional de Educación a Distancia (Calle Pintor Sorolla 21, 46002 Valencia – Spain); mercedes.orus@gmail.com
[2] Université de la Sorbonne Nouvelle Paris 3, Paris, France,
[3] Centre d'Analyse et de Mathématique Sociales—CAMS, paris, France
* Correspondence: Christophe Jouis; Christophe.jouis@sorbonne-nouvelle.fr*

Abstract: In this paper, we demonstrate the Collatz conjecture using the mathematical complete induction method. We show that this conjecture is satisfied for the first values of natural numbers, and in analyzing the sequence generated by odd numbers, we can deduce a formula for the general term of the Collatz sequence for any odd natural number n after several iterations. This formula is used in one case that we analyze using the mathematical complete induction method in the process of demonstrating the conjecture.

Keywords: number theory; Collatz conjecture

1. Introduction

Collatz conjecture is one of the best-known unsolved problems in sequences and series of number theory. It states:

> *"For any positive integer n, if a sequence is defined by recurrence, so that, if the previous term is even then the next term is obtained by dividing by 2 the previous term, and if it is odd it is obtained by multiplying by 3 the previous term and adding 1, this sequence always reaches the number 1, and therefore, its last terms will always be the cycle 4, 2, 1.".*

This conjecture was called the Collatz Conjecture, because it was Lothar Collatz who stated it in 1937 [1].

However, it is also known with other names, as the 3n+1 conjecture, or the Ulam conjecture, or Kakutani's problem, or the Thwaites conjecture, or Hasse's algorithm, or the Syracuse problem [2], or hailstone sequence or hailstone numbers, because the values are ascending or descending multiple times [3], or as wondrous numbers [4].

This conjecture has not been proven; however many mathematicians have studied it achieving important results, see [5–18]. Most of them, have argued that the conjecture is true, as a result of the experimental evidence and heuristic arguments [12].

In this paper, we demonstrate the Collatz conjecture using the mathematical complete induction method.

We show that this conjecture is satisfied for the first values of natural numbers. From this analysis, we can deduce a formula for the general term of Collatz sequence for any odd natural number n after several iterations, and this formula is used in one case that we analyze using the mathematical complete induction method in the process of demonstrating the conjecture.

The paper is organized as follows: In Section 2, we show that the conjecture is satisfied for the first values of natural numbers, and we deduce a formula for the general term of the Collatz sequence for any odd natural number n after several iterations. In Section 3, we demonstrate the Collatz conjecture using the mathematical complete induction method. Finally, we wrap up in Section 4 with our conclusions.

2. Sequences for the First Natural Numbers and Formula for the General Term of Collatz Sequence for Any Odd Natural Number *n*

Formally, each term of the sequence of numbers is equivalent to applying the following function to n, and to each term of the sequence:

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

Therefore, given any natural number, we can consider its orbit; that is, the successive images when iterating the function, in the following way.

For example, if $n = 13$:

$$\begin{aligned} x_1 &= f(13) = 3? 3+1 = 40, \\ x_2 &= f(40) = \frac{40}{2} = 20, \\ x_3 &= f(20) = \frac{20}{2} = 10; \text{ etc.} \end{aligned} \tag{1}$$

The conjecture says that we will always reach 1 (and therefore cycle 4, 2, 1) when starting with any natural number.

We will now present what happens with the first natural numbers.

For n = 1:

$$\begin{aligned} x_1 &= f(1) = 3? +1 = 4, \\ x_2 &= f(4) = \frac{4}{2} = 2, \\ x_3 &= f(2) = \frac{2}{2} = 1. \end{aligned}$$

For n = 2:

$$x_1 = f(2) = \frac{2}{2} = 1. \tag{3}$$

For n = 3:

$$\begin{aligned} x_1 &= f(3) = 3? +1 = 10, \\ x_2 &= f(10) = \frac{10}{2} = 5, \\ x_3 &= f(5) = 3? +1 = 16, \\ x_4 &= f(16) = \frac{16}{2} = 8, \\ x_5 &= f(8) = \frac{8}{2} = 4, \\ x_6 &= f(4) = \frac{4}{2} = 2, \\ x_7 &= f(2) = \frac{2}{2} = 1. \end{aligned} \tag{4}$$

Therefore, for the first values of n, we observe that the conjecture is satisfied. The conjecture is also satisfied for numerous other numbers greater than 3, and for numbers that are a power of 2, we will also always reach 1, dividing successively by 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2022 doi:10.20944/preprints202205.0174.v1

Given this, at a certain point in the process of demonstration using the mathematical complete induction method, we will need a formula that would represent the general term of the Collatz sequence for any odd natural number n after several iterations, and we examine how this formula would be below.

If n is an odd natural number, then using the definition of $f(n)$, the first iteration would be:

$$x_1 = 3n + 1.$$

Since $3n + 1$ is always an even number, the next iteration would be:

$$x_2 = \frac{3n+1}{2}.$$

The next iteration will depend on whether the previously obtained result is even or odd, and since it is possible to obtain an even number in the following or next iterations, let us assume that an even number is obtained in the following $r_1 - 1$ iterations, until we again obtain an odd number ($r_1 \geq 1$). Therefore, the result obtained after $r_1 - 1$ iterations would be:

$$x_{r_1+1} = \frac{3n+1}{2^{r_1}}.$$

Now, if $\frac{3n+1}{2^{r_1}}$ is an odd number, the next iteration would be:

$$x_{r_1+2} = 3 \cdot \frac{3n+1}{2^{r_1}} + 1$$

$$x_{r_1+2} = \frac{3^2 n + 3 + 2^{r_1}}{2^{r_1}}.$$

Since $\frac{3^2 n + 3 + 2^{r_1}}{2^{r_1}}$ is an even number, the next iteration would be:

$$x_{r_1+3} = \frac{3^2 n + 3 + 2^{r_1}}{2^{r_1+1}}.$$

Moreover, for the same reason as before, we will assume that we obtain an even number in the next $r_2 - 1$ iterations, until we obtain again an odd number ($r_2 \geq 1$). Therefore, the result would be:

$$x_{r_1+r_2+2} = \frac{3^2 n + 3 + 2^{r_1}}{2^{r_1+r_2}}.$$

Following the same reasoning, this process will continue, and after several iterations, such as for example $r_1 + r_2 + ... + r_k + k$ iterations, we would obtain that the term $x_{r_1+r_2+...+r_k+k}$ of the sequence would be:

$$x_{r_1+r_2+...+r_k+k} = \frac{3^k n + 3^{k-1} + 3^{k-2} \cdot 2^{r_1} + 3^{k-3} \cdot 2^{r_1+r_2} + ... + 2^{r_1+r_2+...+r_{k-1}}}{2^{r_1+r_2+...+r_k}}. \quad (5)$$

This formula will be used in the next section, when we analyze a specific case using the mathematical complete induction method.

3. Proof of Collatz Conjecture Using Mathematical Complete Induction

We need to prove that, for all $n \in$ N, the obtained sequence reaches 1.

In the previous section, we saw that this holds true for values of n from 1 to 3. Moreover, it is also proven to be true for values higher than $n = 3$, and it is found to also be true, for example, for all $n = 2^s$, $s \in$ N, because we always reach 1 when dividing successively by 2.

To apply the mathematical complete induction method, we will assume that, for a certain $m \in$ N that is sufficiently high in value and for any other natural number less than m, we can reach the number 1 with successive iterations.

Hence, if we can prove that it is true for $m + 1$, we can conclude that it is true for all $n \in$ N.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2022 doi:10.20944/preprints202205.0174.v1

Therefore, in our induction hypothesis, we assume that it is true for a sufficiently large $m \in \mathbb{N}$ value and for any other natural number less than m.

We explore below if this is true also for $m + 1$. Note that $m + 1$ can be an odd or even number, depending on whether m is even or odd, respectively. Hence, we will analyze both cases:

Case 1: $m + 1$ is an even number

If $m + 1$ is an even number, it is because m is an odd number; that is, $m = 2t + 1$ for $t \in \mathbb{N}$. Hence, $m + 1 = 2t + 2$ and the first iteration applying the definition of $f(n)$ would be:

$$x_1 = \frac{2t+2}{2} = t+1.$$

Note that $t + 1 < 2t + 1 = m$, and as we were assuming that we reach the number 1 for all natural numbers less or equal than m, then we reach the number 1 for $t + 1$.

Therefore, if $m + 1$ is an even number, the obtained sequence reaches 1.

Case 2: $m + 1$ is an odd number

If $m + 1$ is an odd number, it is because m is an even number; that is, $m = 2t$ for $t \in \mathbb{N}$, and $t > 1$. Hence, $m + 1 = 2t + 1$ and the first iteration applying the definition of $f(n)$ would be:

$$x_1 = 3 \cdot (2t + 1) + 1 = 6t + 4.$$

Given that $6t + 4$ is an even number, the next iteration would be:

$$x_2 = \frac{6t+4}{2} = 3t+2.$$

At this point, note that the next iteration depends on whether t is an even or odd number; if t is an even number then $3t + 2$ will be an even number; however, if t is an odd number then $3t + 2$ will be an odd number.

Next, we will analyze both options:

Option 1: t is an even number:

If t is an even number, then $3t + 2$ is an even number, and the next iteration applying the definition of $f(n)$ would be:

$$x_3 = \frac{3t+2}{2}.$$

Note that $\frac{3t+2}{2} \leq 2t = m$ because $t \geq 2$, and as we were assuming that we reach the number 1 for all natural numbers less or equal than m, then we can reach number 1 for $\frac{3t+2}{2}$ and for $m + 1$. Therefore, if $m + 1$ is an odd number and t an even number, the sequence obtained reaches 1.

Option 2: t is an odd number:

If t is an odd number, then $3t + 2$ is an odd number, and now using the formula Eq.5, which represents the general term of Collatz sequence for any odd natural number after several iterations, we have that after $r_1 + r_2 + \ldots + r_k + k$ iterations, we would obtain:

$$x_{r_1+r_2+\ldots+r_k+k} = \frac{3^k(3t+2) + 3^{k-1} + 3^{k-2} \cdot 2^{r_1} + 3^{k-3} \cdot 2^{r_1+r_2} + \ldots + 2^{r_1+r_2+\ldots+r_{k-1}}}{2^{r_1+r_2+\ldots+r_k}}. \quad (6)$$

Rewriting the above formula, we have:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 May 2022 doi:10.20944/preprints202205.0174.v1

$$\frac{3^k \cdot t + 3^{k-1} + 3^{k-2} \cdot 2^{r_1} + 3^{k-3} \cdot 2^{r_1+r_2} + \ldots + 2^{r_1+r_2+\ldots+r_{k-1}}}{2^{r_1+r_2+\ldots+r_k}} + \frac{3^k(2t+2)}{2^{r_1+r_2+\ldots+r_k}}. \quad (7)$$

At this point, note that for $m = 2t$, t an odd natural number, we were assuming that the sequence reaches number 1, and calculating the first terms for the sequence of $m = 2t$, we have:

$$x_1 = \frac{2t}{2} = t, \quad x_2 = 3t + 1, \quad x_3 = \frac{3t+1}{2}. \quad (8)$$

The next iteration will depend on whether the previously obtained result is even or odd, and since we can obtain an even number in the following or next iterations, then using the formula Equation (5) and after the next $s_1 + s_2 + \ldots + s_u + u - 3$ iterations, we would obtain:

$$x_{s_1+s_2+\ldots+s_u+u} = \frac{3^u t + 3^{u-1} + 3^{u-2} \cdot 2^{s_1} + 3^{u-3} \cdot 2^{s_1+s_2} + \ldots + 2^{s_1+s_2+\ldots+s_{u-1}}}{2^{s_1+s_2+\ldots+s_u}}. \quad (9)$$

If we call $S = \sum_{i=1}^{u} s_i$, and since for $m = 2t$ we were assuming that the sequence reaches the number 1, it means that the limit of x_{S+u} when S and u tend to infinity is equal to 1.

Thus, if we call $R = \sum_{i=1}^{k} r_i$ in equation (7) and we calculate the limit when R and k tend to infinity, we have:

$$\lim_{(R,k)\to(\infty,\infty)} \frac{3^k \cdot t + 3^{k-1} + 3^{k-2} \cdot 2^{r_1} + 3^{k-3} \cdot 2^{r_1+r_2} + \ldots + 2^{r_1+r_2+\ldots+r_{k-1}}}{2^{r_1+r_2+\ldots+r_k}} + \frac{3^k(2t+2)}{2^{r_1+r_2+\ldots+r_k}} =$$

$$= \lim_{(R,k)\to(\infty,\infty)} \frac{3^k \cdot t + 3^{k-1} + 3^{k-2} \cdot 2^{r_1} + 3^{k-3} \cdot 2^{r_1+r_2} + \ldots + 2^{r_1+r_2+\ldots+r_{k-1}}}{2^{r_1+r_2+\ldots+r_k}} + \lim_{(R,k)\to(\infty,\infty)} \frac{3^k(2t+2)}{2^{r_1+r_2+\ldots+r_k}} =$$

$$= \lim_{(S,u)\to(\infty,\infty)} \frac{3^u \cdot t + 3^{u-1} + 3^{u-2} \cdot 2^{s_1} + 3^{u-3} \cdot 2^{s_1+s_2} + \ldots + 2^{s_1+s_2+\ldots+s_{u-1}}}{2^{s_1+s_2+\ldots+s_u}} + \lim_{(R,k)\to(\infty,\infty)} \frac{3^k(2t+2)}{2^{r_1+r_2+\ldots+r_k}} =$$

$$= 1 + \lim_{(R,k)\to(\infty,\infty)} \frac{3^k(2t+2)}{2^R}.$$

Note now that the value of R cannot be less than k, because using the definition of the sequence, every time a term x_i in the sequence is odd, the next calculated term is always even.

The value of R cannot be equal to k either, because for this to happen, it would mean that all $r_i = 1$ that is, every time we divide by 2, the value obtained is an odd number. Moreover, it can be shown how this is only possible if the value of t is a value that increases as we calculate new iterations. However, we should remember that the value of t is an odd number that we take at random, so that $3t + 2$ is an odd number, but without changing the t value in each iteration.

Next, it is shown how this t value should be changing, so R would be equal to k.

Recall that $3t + 2$ is odd because t is also odd, and therefore, the next term of the sequence would be:

$$x_4 = 3 \cdot (3t + 2) + 1 = 9t + 7.$$

Since $9t + 7$ is even, the next term in the sequence would be:

$$x_5 = \frac{9t+7}{2}$$

For $\frac{9t+7}{2}$ to be odd, there must be an $a \in \mathbb{N}$ such that:

$$\frac{9t+7}{2} = 2a + 1$$

Hence, the question now is: Are there natural numbers t and a that satisfy the previous equation? The answer is yes, when $t = 4w + 3$ and $a = 9w + 8$ for $w \in$ N.

Therefore, assuming that $\frac{9t+7}{2}$ is odd because $t = 4w + 3$ for $w \in$ N, the next term in the sequence would be:

$$x_6 = 3 \cdot \frac{9t+7}{2} + 1 = \frac{27t+23}{2}$$

Since this value is even, the next term would be:

$$x_7 = \frac{27t+23}{2}$$

The question now is: Can the previous value be odd? That is, are there natural numbers t and a such that $27t + 23 = 8a + 4$?. The answer is yes when $t = 8w + 7$ for $w \in$ N.

Hence, we can observe that the value of t has increased. Previously, it was $t = 4w + 3 = 2^2w + (2^2 - 1)$, and now it should be $t = 8w + 7 = 2^3w + (2^3 - 1)$ for $w \in$ N, and so on.

Hence, to be $R = k$, the t value would be increasing, increasing the exponent of the power of 2, which means that in this case, it is not possible when setting an odd value of t but, rather, to make it possible, the value of t must be changing, increasing in the way we have seen. Hence, this implies that the case $R = k$ is not possible.

Therefore $R > k$, and then $R = k + v$ for $v \in$ N, and it can happen that $v > k$ or $v < k$ (this last option could happen if many $r_i = 1$).

Analyzing both cases, we have:

- First case $v > k$: If $v > k$ then $v = k + w$ for $w \in$ N, $R = 2k + w$, and the above limit when R and k tend to infinity would be calculated as:

$$1 + \lim_{(R,k)\to(\infty,\infty)} \frac{3^k(2t+2)}{2^R} = 1 + \lim_{k\to\infty} \frac{3^k(2t+2)}{2^{2k+w}} = 1 + \lim_{k\to\infty} \left(\frac{3}{4}\right)^k \frac{2t+2}{2^w} = 1 + 0 = 1.$$

- Second case $v < k$: If $v < k$ then $v = k - w$ for $w \in$ N and $w < k$, $R = 2k - w$, and the above limit when R and k tend to infinity would be calculated as:

$$1 + \lim_{(R,k)\to(\infty,\infty)} \frac{3^k(2t+2)}{2^R} = 1 + \lim_{k\to\infty} \frac{3^k(2t+2)}{2^{2k-w}} = 1 + \lim_{k\to\infty} \left(\frac{3}{4}\right)^k \frac{2t+2}{2^{-w}} = 1 + 0 = 1.$$

Thus, for an odd number $m + 1$ and an odd number t, the obtained sequence also reaches 1.

Therefore, for an $m + 1$ even or odd number, the sequence reaches 1, and this means that, for all natural numbers n, the Collatz sequence always reaches 1, as we sought to prove.

4. Conclusions

We have shown how to use the mathematical complete induction method to prove the Collatz conjecture. We show that this conjecture is satisfied for the first values of natural numbers. From this analysis, we can deduce a formula for the general term of a Collatz sequence for any odd natural number n after several iterations. This formula is used in one case that we analyzed by mathematical complete induction during the demonstration. Thus, using the mathematical complete induction method, we have demonstrated that the Collatz conjecture is true.

Acknowledgments: We acknowledge Roman Orus for proposing the dissemination of this result.

References

1. O'Connor, J.J.; Robertson, E.F. *Lothar Collatz*; St Andrews University School of Mathematics and Statistics: St Andrews, UK, 2006.
2. Maddux, C.D.; Johnson, D.L. *Logo: A Retrospective*; Haworth Press: New York, NY, USA, 1997; p. 160, ISBN 0-7890-0374-0. The problem is also known by several other names, including: Ulam's conjecture, the Hailstone problem, the Syracuse problem, Kakutani's problem, Hasse's algorithm, and the Collatz problem.

 doi:10.20944/preprints202205.0174.v1

3. Pickover, C.A. *Wonders of Numbers*; Oxford University Press: Oxford, UK, 2001; pp. 116–118.
4. Hofstadter, D.R. *Gödel, Escher, Bach*; Basic Books: New York, NY, USA, 1979; pp. 400–402, ISBN 0-465-02685-0.
5. Guy, R.K. E17: Permutation Sequences. In *Unsolved Problems in Number Theory*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 336–337, ISBN 0-387-20860-7.
6. Guy, R.K. Don't try to solve these problems. *Amer. Math. Mon*. 1983, *90*, 35–41. https://doi.org/10.2307/2975688.
7. Lagarias, J.C. (Ed.) *The Ultimate Challenge: The 3x + 1 Problem*; American Mathematical Society: Providence, RI, USA, 2010; p. 4, ISBN 0821849409.
8. Leavens, G.T.; Vermeulen, M. $3x + 1$ Search Programs. *Comput. Math. Appl*. 1992, *24*, 79–99. https://doi.org/10.1016/0898-1221(92)90034-F.
9. Simons, J.; de Weger, B. Theoretical and computational bounds for m–cycles of the $3n + 1$ problem (PDF). *Acta Arithmetica* 2003, *117*, 51–70. https://doi.org/10.4064/aa117-1-3.
10. Steiner, R.P. A Theorem on the Syracuse Problem. In Proceedings of the 7th Manitoba Conference on Numerical Mathematics, 29 September–1 October 1977; pp. 553–559.
11. Simons, J.L. On the nonexistence of 2–cycles for the $3x + 1$ problem. *Math. Comp.* 2005, *74*, 1565–1572. https://doi.org/10.1090/s0025-5718-04-01728-4.
12. Krasikov, I.; Lagarias, J.C. Bounds for the $3x + 1$ problem using difference inequalities. *arXiv* 2003, arXiv:math/0205002.
13. Colussi, L. The convergence classes of Collatz function. *Theor. Comput. Sci*. 2011, *412*, 5409–5419. https://doi.org/10.1016/j.tcs.2011.05.056.
14. Hew, P.C. Working in binary protects the repetends of 1/3h: Comment on Colussi's The convergence classes of Collatz function. *Theor. Comput. Sci*. 2016, *618*, 135–141. https://doi.org/10.1016/j.tcs.2015.12.033.
15. Terras, R. A stopping time problem on the positive integers (PDF). *Pol. Akad. Nauk.* 1976, *30*, 241–252.
16. Chamberland, M. A continuous extension of the $3x + 1$ problem to the real line. *Dynam. Contin. Discret. Impuls Syst*. 1996, *2*, 495–509.
17. Letherman, S.; Schleicher, D.; Wood, R. The $(3n + 1)$-Problem and Holomorphic Dynamics. *Exp. Math*. 1999, *8*, 241–252. https://doi.org/10.1080/10586458.1999.10504402.
18. Conway, J.H. Unpredictable Iterations. In Proceedings of the 1972 Number Theory Conference, Boulder, CO, USA, 14–18 August 1972; pp. 49–52.