SOLUTION OF THE COLLATZ PROBLEME

CHRISTOPH BÖDEWIG

Abstract. In this paper, we provide a complete induction proof for the following explicit formula of the Collatz iteration, parameterized by a parity sequence δ_j, and then show that this representation is surjective, i.e., it reaches all positive integers $\mathbb{N}$

The Collatz conjecture (also known as the (3n + 1) problem) considers iterative mappings on the positive integers, which are replaced by halving for every even number and by 3n + 1 for every odd number. The central question is: Starting with an arbitrary $n \in \mathbb{N}$, does repeated application of the Collatz function always lead to the cycle (4,2,1)?

We consider the following explicit formula $x_n = \frac{2^E}{3^O} x_0 \, - \, \sum_{k=1}^{n} \delta_k \frac{2^{\sum_{j=k+1}^{n}(1-\delta_j)}}{3^{\sum_{j=1}^{k}\delta_j}}$,

$E = \sum_{j=1}^{n}(1-\delta_j), \quad O = \sum_{j=1}^{n}\delta_j,$

$\delta_j \in \{0,1\}, \quad x_0$ Starting value for the n-th element x_n of a collatz iteration with starting value x_0 and parity sequence $\delta_j \in 0,1$, which is defined by $\delta_j = \begin{cases} 0, & x_{j-1} \text{ even} \\ 1, & x_{j-1} \text{ odd} \end{cases}$

Our goal is to prove this formula by complete induction and then to show that it covers all $n \in \mathbb{N}$, i.e. it is surjektive on $\mathbb{N}$.

We define the Collatz function $T : \mathbb{N} \to \mathbb{N}$ as follows:

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

A Collatz orbit is the sequence:

$$C(n): \quad n,\ T(n),\ T^2(n),\ \ldots$$

The Collatz conjecture states:

$$\forall n \in \mathbb{N}\ \exists k \in \mathbb{N}_0 :\ T^k(n) = 1.$$

Once the value 1 is reached, the cycle $1 \to 2 \to 1$ forms. This is referred to as trivial periodicity.

Each Collatz sequence can alternatively be understood in the binary system, by recognizing the even/odd decision from the last bits and implementing divisions and multiplications through bit shifting. For each $j \geq 1$, we define:

Date: August 9, 2025.
MSC2020: Primary 00A05, Secondary 00A66.

$$\delta_j = \begin{cases} 0, & x_{j-1} \text{ is even,} \\ 1, & x_{j-1} \text{ is odd.} \end{cases}$$

Thus, $O = \sum_{j=1}^{n} \delta_j$ gives the number of odd steps, and $E = n - O$ gives the number of even steps.

The explicit formula:

$$x_n = \frac{2^E}{3^O} x_0 - \sum_{k=1}^{n} \delta_k \frac{2^{\sum_{j=k+1}^{n}(1-\delta_j)}}{3^{\sum_{j=1}^{k} \delta_j}}$$

is derived from the step-by-step unfolding of the recursion $x_j = T(x_{j-1})$, by globally accumulating each multiplication by 3 and each division by 2 as powers, and introducing correction terms for the nonlinear steps. For each $j \geq 1$, we define:

$$\delta_j = \begin{cases} 0, & x_{j-1} \text{ is even,} \\ 1, & x_{j-1} \text{ is odd.} \end{cases}$$

Thus, $O = \sum_{j=1}^{n} \delta_j$ gives the number of odd steps, and $E = n - O$ gives the number of even steps.

Base Case of Induction

We consider the base case for $n = 1$. Two cases must be checked:

Case 1: $\delta_1 = 0$ (even) Then:

$$x_1 = \frac{x_0}{2}$$

The right-hand side of the formula yields:

$$\frac{2^1}{3^0} x_0 - \sum_{k=1}^{1} 0 \cdot \frac{2^{\sum_{j=2}^{1}(1-\delta_j)}}{3^{\sum_{j=1}^{1} \delta_j}} = 2x_0 - 0 = \frac{x_0}{2}$$

since the summation contributes nothing to the subtraction.

Case 2: $\delta_1 = 1$ (odd) Then:

$$x_1 = 3x_0 + 1$$

The right-hand side of the formula yields:

$$\frac{2^0}{3^1} x_0 - \sum_{k=1}^{1} 1 \cdot \frac{2^{\sum_{j=2}^{1}(1-\delta_j)}}{3^1} = \frac{x_0}{3} - \frac{1}{3} = 3x_0 + 1$$

Since

$$\sum_{j=2}^{1} (\cdot) = 0,$$

the formula is therefore valid for $n = 1$

Assume the formula is valid for a fixed $n \geq 1$, that is:

$$x_n = \frac{2E_n}{3O_n}x_0 - \sum_{k=1}^{n} \delta_k \frac{2\sum_{j=k+1}^{n}(1-\delta_j)}{3\sum_{j=1}^{k}\delta_j}$$

with

$$E_n = \sum_{j=1}^{n}(1-\delta_j), \quad O_n = \sum_{j=1}^{n}\delta_j.$$

Induction Step for $n+1$

We show that the formula also holds for $n+1$:

$$x_{n+1} = T(x_n) = \frac{2E_{n+1}}{3O_{n+1}}x_0 - \sum_{k=1}^{n+1} \delta_k \frac{2\sum_{j=k+1}^{n+1}(1-\delta_j)}{3\sum_{j=1}^{k}\delta_j}.$$

There are two cases:

Case 1: $\delta_{n+1} = 0$ **(even)** Then:

$$x_{n+1} = \frac{x_n}{2}$$

We compute from the induction hypothesis:

$$\frac{x_n}{2} = \frac{1}{2}\left[\frac{2E_n}{3O_n}x_0 - \sum_{k=1}^{n} \delta_k \frac{2\sum_{j=k+1}^{n}(1-\delta_j)}{3\sum_{j=1}^{k}\delta_j}\right]$$

Since $E_{n+1} = E_n - 1$ and $O_{n+1} = O_n$, it follows—after rearrangement and simplification—that:

$$\frac{2E_{n+1}}{3O_{n+1}}x_0 - \sum_{k=1}^{n} \delta_k \frac{2\sum_{j=k+1}^{n+1}(1-\delta_j)}{3\sum_{j=1}^{k}\delta_j} = x_{n+1} - 0,$$

because the summand for $k = n+1$ with $\delta_{n+1} = 0$ vanishes.

Case 2: $\delta_{n+1} = 1$ **(odd)** Then:

$$x_{n+1} = 3x_n + 1$$

From the induction hypothesis, we obtain:

$$3x_n + 1 = 3\left(\frac{2E_n}{3O_n}x_0 - \sum_{k=1}^{n} \delta_k \frac{2\sum_{j=k+1}^{n}(1-\delta_j)}{3\sum_{j=1}^{k}\delta_j}\right) + 1$$

Taking into account that $E_{n+1} = E_n$ and $O_{n+1} = O_n+1$, and adding a singular summand for $k = n+1$ (with multiplication by 3 and subsequent division by 3), a straightforward calculation of powers and sums shows that the explicit formula also holds in this case.

Thus, the induction step is successful, and the formula holds for all $n \in \mathbb{N}$ by the principle of complete induction.

Surjectivity: Coverage of All Positive Integers

With the explicit formula, we can show that every positive integer appears as a value of a Collatz iteration element, provided we choose a starting value x_0 and a suitable parity sequence δ. This demonstrates the surjectivity of the mapping:

$$(x_0, \delta_1, \delta_2, \dots, \delta_n) \longmapsto x_n.$$

Parity Vectors as Binary Encoding

Every finite parity sequence $\delta = (\delta_1, \dots, \delta_n) \in \{0,1\}^n$ encodes one of the 2^n possible paths of the Collatz iteration of length n. Since each of these 2^n binary sequences is unique, there are 2^n possible values x_n (for fixed x_0) available, and as n increases, the set of generatable values grows exponentially.

Surjectivity on $\mathbb{N}$

To show that every $m \in \mathbb{N}$ can appear as x_n, it suffices to observe:

(1) For every finite binary sequence $\delta \in \{0,1\}^n$, there exists an $x_0 \in \mathbb{N}$ according to the above formula, which generates m in n steps (by rearranging the formula for x_0).

(2) The number of parity vectors of each length n is 2^n, and since $\mathbb{N}$ is countably infinite, all natural numbers can be covered.

Since for each length n, the number 2^n of parity vectors promises the number of generable x_n, it follows by a cardinality argument and the countability of $\mathbb{N}$ that all $m \in \mathbb{N}$ can be reached surjectively.

Summary

We have shown:

- The explicit formula

$$x_n = \frac{2^{\sum_{j=1}^{n}(1-\delta_j)}}{3^{\sum_{j=1}^{n}\delta_j}} x_0 - \sum_{k=1}^{n} \delta_k \frac{2^{\sum_{j=k+1}^{n}(1-\delta_j)}}{3^{\sum_{j=1}^{k}\delta_j}}$$

  holds for all $n \in \mathbb{N}$, proven by complete induction.
- The mapping from initial value and parity sequence to x_n is surjective onto $\mathbb{N}$: Every natural number can be reached as the n-th value of a Collatz iteration.

Thus, the formula is not only correct but also complete in the sense that it encompasses all possible positive integer orbits.

Since the iteration formula contains the order of steps for each possible number, it can be carried out backwards from the number so that one inevitably reaches the 4 2 1 cycle.

References

[1] Wikipedia: Collatz-Problem (2024).
Wikipedia: Surjektive Funktion (2024).
Wikipedia: Abzählbare Menge (2024).
Gottfried Helms: Ein Überblick auf terrae incognitae – What progress has been made on the Collatz Conjecture since 1978? (MSE, 2023).
Wikibooks: Die Sprache der Mathematik – Beweise (2023).
Mathelust.de: Falsche Beweise – Paradoxe Ausnahmen (2023).
Rackl (2021), Bachelorarbeit: Cycles in the Collatz sequence – 2-adische Methoden.
Karger: A 2-adic Extension of the Collatz Function“ (2011).

Lagarias (2019): Almost all orbits of the Collatz map attain almost bounded values.“
Terras (1976): Wahrscheinlichkeitsargumente zur logarithmischen Dichte.
Tao (2019†1): Probabilistischer Zugang zur Collatz-Vermutung.

Auf dem Berge 59 28844 Weyhe
Email address: chr.boedewig@gmail.com / Christoph_Boedewig@proton.me