$3n + 1$ conjecture: a proof or almost

Alexandre MAKARENKO

April 2, 2022

The original Collatz algorithm is rewritten to avoid division by two and to transform it from hailstone to a steadily growing value. Then the problem is reverted and solved in combinatorial way to find all integers leading to the sequence end.

1 Collatz differently

1.1 Notations

In the original formulation for any integer number $X_i > 0$ to obtain X_{i+1} we either multiply X_i by 3 and add 1 if it is odd, or divide it by 2 until the result remains even. Such an algorithm leads to a, so called, hailstone behaviour of X_i.

For any integer number $X_i > 0$ represented in binary basis we will use H_i (Head) to designate the most significant bit position and T_i(Tail) the least significant bit position (number of trailing zeros). For example for a binary number

```
10001010101000
H000000000T000
```

$H = 13$ and $T = 3$.

1.2 Key statement

The new equivalent sequence will be

$$X_{i+1} = 3X_i + 2^{T_i} \tag{1.1}$$

and Collatz states it will eventually lead to $H_n = T_n$

$$X_n = 2^{Tn} \tag{1.2}$$

In other words to a single 1 shifted left by T_n bits. Any additional step for $i > n$ will merely multiply X_i by 4

$$X_{i+1} = 3X_i + 2^{T_i} = 3 \cdot 2^{T_i} + 2^{T_i} = 4 \cdot 2^{T_i} = 2^{T_i+2}$$

or shift it left by two positions.

Example for 49:

i	binary X_i	decimal X_i	original X_i
0	110001	49	49
1	10010100	148	37
2	111000000	448	7
3	10110000000	1408	11
4	1000100000000	4352	17
5	11010000000000	13312	13
6	1010000000000000	40960	5
7	100000000000000000	131072	(end)1
8	10000000000000000000	524288	(useless)1

Let us demonstrate in details a step for $X_2 = 448 = 111000000$. After multiplication by 3 instead of dividing the result by 2 we add $2^6 = 1000000$:

i	binary X_i	decimal X_i	original X_i
2	111000000	448	7
	10101000000	448*3	
	+1000000	+2^6	
3	10110000000	1408	11

With this new formulation the recursion will be:

$$X_1 = 3X_0 + 2^{T_0}$$

$$X_2 = 3X_1 + 2^{T_1} = 3\left(3X_0 + 2^{T_0}\right) + 2^{T_1} = 3^2 \cdot X_0 + 3^1 \cdot 2^{T_0} + 3^0 \cdot 2^{T_1}$$

$$X_n = 3^n \cdot X_0 + 3^{n-1} \cdot 2^{T_0} + \cdots + 3^1 \cdot 2^{T_{n-2}} + 3^0 \cdot 2^{T_{n-1}} \tag{1.3}$$

1.3 Sequence properties

Fact 1. *The number of steps to complete the sequence is exactly the number of odd values in the original Collatz.*

Fact 2. *For any $j > i$:*

$$T_j > T_i \tag{1.4}$$

One could say: the problem is no longer a hailstone.

Fact 3. *On a side note, between two neighbour steps:*

$$H_{i+1} - H_i = 1 \, or \, 2$$

and in average the head speed before it meats the tail is $S_H = Av\,(H_{i+1} - H_i) = \log_2 3$. Meanwhile the tail moves with average speed $S_T = 2$ (for $H_i - T_i > 2$).

So, intuitively, we would expect the tail to catch and substitute the head (this is exactly what Collatz is about).

2 The proof

For a given sequence end $X_n = 2^{T_n}$ there are generally many starting points X_0 leading to X_n. For instance, both $X_0 = 26$ and $X_0 = 85$ end with $X_n = 256 = 2^8$.

2.1 Key moment

Instead of generating values according to Eq.1.2 we will look for all possible paths back from $2^{T_n} = 1 \ll T_n$.

Reverting Eq.1.2 yields

$$X_i = \left(X_{i+1} - 2^{T_i}\right)/3 \tag{2.1}$$

where

$$0 \leq T_i < T_{i+1} \tag{2.2}$$

and

$$\mod\left(X_{i+1} - 2^{T_i}, 3\right) = 0 \tag{2.3}$$

This means that starting from a $X_n = 2^{T_n}$ we can find all possible values for X_{n-1} by testing T_{n-1} against Eq.2.2 and Eq.2.3. Then repeat for each T_{n-1}. And so on we will discover all values leading to X_n.

For example, observing closer a value $2^{T_n} = 1000000...0000$ one can see that the number of suitable values for T_{n-1} is $T_n/2$ (number of zero pairs). Moreover, the lowest acceptable $T_{n-1} = 1$ if T_n is even otherwise $T_{n-1} = 2$. While the highest is always $T_{n-1} = T_n - 2$.

All child values of 2^{T_n} with even T_n and odd T_n never overlap (see Example). Thus picking up two large starting points 2^{T_n-1} and 2^{T_n} will seed uniquely values situated below $2^{T_n}/3$. Tending T_n to infinity then will fill the integers from 1 to ∞.

Proof. If for any integer X_0 there is always a way to reach it from a 2^{T_n} according to Eq.2.1 the same path can be followed back by means of Eq.1.2. ☐

3 Example

Values reverted from 2^7 and 2^8 with Eq.2.1:

```
128        10000000  odd  Tn=7
  42         101010  =  (10000000-10)/11          =  1111110/11
  40         101000  =  (10000000-1000)/11        =  1111000/11
    13         1101  =  (101000-1)/11             =  100111/11
    12         1100  =  (101000-100)/11           =  100100/11
  32         100000  =  (10000000-100000)/11      =  1100000/11
    10         1010  =  (100000-10)/11            =  11110/11
      3          11  =  (1010-1)/11               =  1001/11
    8          1000  =  (100000-1000)/11          =  11000/11
      2          10  =  (1000-10)/11              =  110/11

256       100000000  even  Tn=8
  85        1010101  =  (100000000-1)/11          =  11111111/11
  84        1010100  =  (100000000-100)/11        =  11111100/11
  80        1010000  =  (100000000-10000)/11      =  11110000/11
    26        11010  =  (1010000-10)/11           =  1001110/11
    24        11000  =  (1010000-1000)/11         =  1001000/11
  64        1000000  =  (100000000-1000000)/11    =  11000000/11
    21        10101  =  (1000000-1)/11            =  111111/11
    20        10100  =  (1000000-100)/11          =  111100/11
      6         110  =  (10100-10)/11             =  10010/11
    16        10000  =  (1000000-10000)/11        =  110000/11
      5         101  =  (10000-1)/11              =  1111/11
      4         100  =  (10000-100)/11            =  1100/11
        1         1  =  (100-1)/11                =  11/11
```

4 Source code

This document and computer programs may be found here:

https://github.com/sashamakarenko/collatz