Number of Divisions by Two required for Collatz Cycles

Christian Koch[1,a*], Eldar Sultanow[2,b] and Sean Cox[3,c]

[1]Technische Hochschule Nürnberg Georg Simon Ohm, Nuremberg, Germany

[2]Capgemini, Nuremberg, Germany

[3]RatPac-Dune Entertainment, Los Angeles, USA

[a]christian.koch@th-nuernberg.de, [b]eldar.sultanow@capgemini.com, [c]sean.cox@ratpacent.com

Keywords: Collatz Cycles, Divisions by Two, Cayley Graph, Free Group, Reachability

Abstract. Using Data Science techniques, we identified empirically the number of divisions by two that are required for Collatz Cycles. We call this number cycle-alpha $\bar{\alpha}$. We provide a minimum and maximum condition for this cycle-alpha that must be proved in order to validate the correctness of this empirical finding. We prove the minimum condition for all $kx + 1$ variants of Collatz sequences. We prove the maximum condition for special corner cases. The underlying idea is to represent the inverted Collatz sequence as a tree that consists of nodes labeled with odd Collatz sequence numbers. We do not claim to solve the *Million Buck Problem*.

Introduction

Collatz introduced a surjective, non-injective function $g : \mathbb{N} \to \mathbb{N}$ as follows:

$$g(m) = \begin{cases} 3m+1 & 2 \nmid m \\ m/2 & \text{otherwise} \end{cases} \tag{1}$$

Let (a_k) be a numerical sequence with $a_k = g^{(k)}(m)$, then a reversion produces an infinite number of sequences of reversely-written Collatz members [1]. Let S be a set containing two elements q and r, which are bijective functions over $\mathbb{Q}$: $q(x) = 2x$ and $r(x) = \frac{1}{3}(x-1)$. Let a binary operation be the right-to-left composition of functions $q \circ r$, where $q \circ r(x) = q(r(x))$. The set, whose elements are all these compositions, forms a free group F of rank 2 with respect to the free generating set S, where the identity element is the identity function $id_{\mathbb{Q}} = e$. The corresponding Cayley graph $Cay(F, S) = G$ is a regular tree [2, p. 66]. We specify a subgraph H of G containing only vertices labeled by a string over alphabet $\{r, q\}$ without the inverses. This subgraph corresponds to the monoid S^*, which is freely generated by S. Let $Y^X = \{f \mid f \text{ is a map } X \to Y\}$ be the set of functions. We define the evaluation function $ev_{S^*} : S^* \times \{1\} \to \mathbb{Q}$ that evaluates an element of S^*, id est a composition of q and r, for the given input value 1. Furthermore we define the corestriction $ev^0_{S^*}$ of ev_{S^*} to $\mathbb{N}$, which operates on a subset $T \subset S^*$ containing only those compositions of q and r that return a natural number when inputting the value 1. Let $U \subset T$ be a subset of T, which does not contain a reduced word with two or more successive characters r. The corresponding tree $H_U \subset H_T$ reflects Collatz sequences. We define a tree H_C by removing all even labeled vertices from H_U via path contraction. A small section of the tree H_C is shown in figure 1.

Let v_1 and v_{n+1} be two vertices of H_C, where v_1 is reachable from v_{n+1} with $level(v_1) - level(v_{n+1}) = n$. Hence, a path $(v_{n+1}, \ldots, v_1)$ exists between these two vertices. Theorem 1 specifies the following relationship between v_1 and v_{n+1}.

Theorem 1. $l_{V(H_C)}(v_{n+1}) = 3^n l_{V(H_C)}(v_1) \prod_{i=1}^{n} \left(1 + \frac{1}{3 l_{V(H_C)}(v_i)}\right) 2^{-\alpha_i}$. *In order to simplify readability, we waive writing down the vertex label function and put it shortly:*
$v_{n+1} = 3^n v_1 \prod_{i=1}^{n} \left(1 + \frac{1}{3v_i}\right) 2^{-\alpha_i}$. *The value* $\alpha_i \in \mathbb{N}$ *is the number of divisions by two (the number of edges which have been contracted) between* v_i *and* v_{i+1} *in* H_U.

In order to correctly determine successive nodes using theorem 1, we must consider the halting conditions. These are specified in Definition 2.

Definition 2. When determining successive nodes starting at v_1 according to theorem 1, we halt if one of the following two conditions is fulfilled:

1. $v_{n+1} = 1$
2. $v_{n+1} \in \{v_1, v_2, \ldots, v_n\}$

If the first condition applies, the Collatz conjecture is true for a specific sequence. When the second condition is fulfilled, the sequence has led to a cycle. For every starting node, except the root node (labeled with 1), the Collatz conjecture is consequently falsified. Let us consider the example $v_1 = 13$, where the algorithm halts after two iterations, because the first condition is met:

$$v_{n+1} = 3^2 \cdot \left(1 + \frac{1}{3 \cdot 13}\right)\left(1 + \frac{1}{3 \cdot 5}\right) \cdot 2^{-7} = 1$$

If we examine the case $v_1 = 1$, we realize that the algorithm finishes after the first iteration, since both halting conditions are true. The sequence stops because the final node labeled with 1 is reached. Furthermore, the sequence has led to a cycle:

$$v_{n+1} = 3 \cdot \left(1 + \frac{1}{3}\right) 2^{-2} = 1$$

The trivial cycle is the only sequence where both conditions are fulfilled.

Collatz Cycles

A path of length $n \geq 1$ that starts and ends at the same vertex, in which no vertex is repeated with the sole exception that the initial vertex is the terminal vertex, is called a cycle. A cycle of length n is referred to as an n-cycle. When different nodes collapse on one, the graph is no longer necessarily a tree. Non-trivial cycles do not originate from the root, but cause the graph to be a disconnected graph.

Figure 2 depicts a section of $H_{C,5}$, the $5x + 1$ variant of H_C. Because of the two non-trivial cycles $43, 17, 27$ and $83, 33, 13$, in $H_{C,5}$ there does not exist a path between the root and the vertex 43 and between the root and the vertex 83. Utilizing the example of the graph $H_{C,5}$ we are able to deduct from the cycle $43, 17, 27$ the simple and self-evident equality *left-child*$^3(43) = 43$:

$$\begin{aligned}
\textit{left-child}(43) &= \tfrac{1}{5} * (43 * 2^1 - 1) = 17 \\
\textit{left-child}(17) &= \tfrac{1}{5} * (17 * 2^3 - 1) = 27 \\
\textit{left-child}(27) &= \tfrac{1}{5} * (27 * 2^3 - 1) = 43
\end{aligned}$$

Obviously, the authors note, it would be interesting to find out what circumstances enable a graph to have non-trivial cycles, whether it be the $5x + 1$ variant of H_C, the $7x + 1$ variant of H_C or any variant of H_C; let us say the $kx + 1$ variant of H_C with $k \geq 1$.

Let us refer to a $kx + 1$ variant of H_C as $H_{C,k}$. By having introduced and proven theorem 1 we already started an assertion about the reachability of successive nodes in H_C. This reachability relationship can be generalized for any graph $H_{C,k}$ as follows:

$$v_{n+1} = k^n v_1 \prod_{i=1}^{n} \left(1 + \frac{1}{k v_i}\right) 2^{-\alpha_i} \tag{2}$$

This generalization leads to the condition for an existence of an n-cycle in any $kx + 1$ variant of H_C, which looks analogous to the condition given by equation **??** that specifies H_C has a cycle:

$$2^{\alpha} = \prod_{i=1}^{n} \left(k + \frac{1}{v_i} \right) \tag{3}$$

The natural number α is the sum of edges that have been contracted between the vertices v_i forming the cycle, in other words α is the number of divisions by 2 within the sequence. The natural number n is the cycle length and k obviously specifies the variant of H_C. Since between each vertex at least one edge has been contracted (at least one division by 2 took place), we know that our exponent alpha is greater than or equal to the sequence length:

$$\alpha \geq n \tag{4}$$

Using incremental search, one can calculate cycles through trial and error. Table 1 lists all empirically discovered cycles having a length up to 100 that appear in $kx+1$ variants of H_C for $k \in [1, 1000]$. Within each of these variants, the cycles have been searched at potential starting nodes v_1 with a label between 1 and 1000. Note that the cycles in table 1 are written in reverse order, i.e. in the order which corresponds to the Collatz sequence. To obtain the cycles in terms of graph theory referring to the graph H_C, read them from right to left.

k	**cycle**	α	**non-trivial**
1	1	1	
3	1	2	
5	1,3	5	
5	13,33,83	7	✓
5	27,17,43	7	✓
7	1	3	
15	1	4	
31	1	5	
63	1	6	
127	1	7	
181	27,611	15	✓
181	35,99	15	✓
255	1	8	
511	1	9	

Table 1: Known n-cycles in $kx + 1$ variants of H_C for $k \leq 1000$, $n \leq 100$

Based on the results shown in table 1 we state the following theorem 3 that renders more precisely the prerequisite for cycles that may occur in variants of H_C.

Theorem 3. *An n-cycle can only exist in a graph $H_{C,k}$, that means in a $kx + 1$ variant of H_C, if the following equation holds:*

$$2^{\bar{\alpha}} = 2^{\lfloor n \log_2 k \rfloor + 1} = \prod_{i=1}^{n} \left(k + \frac{1}{v_i} \right)$$

The key of theorem 3 consists in the claim that, in order for an n-cycle to occur, the exponent α has to be $\bar{\alpha} = \lfloor n \log_2 k \rfloor + 1$. We approach a proof by expressing formally that $\bar{\alpha}$ is not allowed to be smaller and it is not allowed to be greater than $\lfloor n \log_2 k \rfloor + 1$, in other words we indicate a lower and an upper limit for $\bar{\alpha}$ as follows:

$$\bar{\alpha} > \lfloor n \log_2 k \rfloor \tag{5}$$
$$\bar{\alpha} < \lfloor n \log_2 k \rfloor + 2 \tag{6}$$

The validity of the first part (5), which specifies $\lfloor n \log_2 k \rfloor + 1$ as the lower limit for $\bar{\alpha}$, can be demonstrated in a fairly simple way: Our starting point is equation 2, which describes the relationship of successive vertices in $H_{C,k}$. Having a cycle, requires us to consider the first and the last vertex being one and the same $v_{n+1} = v_1$. Setting a smaller exponent $\bar{\alpha} = \lfloor n \log_2 k \rfloor$ into equation 2 results in the inequality $v_{n+1} > v_1$, which is in any case a true statement:

$$\begin{aligned}
&k^n v_1 2^{-\lfloor n \log_2 k \rfloor} \textstyle\prod_{i=1}^{n} \left(1 + \frac{1}{kv_i}\right) > v_1 \\
&k^n \textstyle\prod_{i=1}^{n} \left(1 + \frac{1}{kv_i}\right) > 2^{\lfloor \ nlog_2 k \rfloor} \\
&\log_2 \left(k^n \textstyle\prod_{i=1}^{n} \left(1 + \frac{1}{kv_i}\right)\right) > \lfloor n \log_2 k \rfloor \\
&n \log_2 k + log_2 \left(\textstyle\prod_{i=1}^{n} \left(1 + \frac{1}{kv_i}\right)\right) > \lfloor n \log_2 k \rfloor
\end{aligned}$$

The validity of the second part (6) is not so trivial to prove. Analogous to the above-shown proof of the cylce-alpha's lower limit, we again refer to equation 2 as our starting point and we need to show that v_{n+1} is smaller than v_1 if $\alpha = \lfloor \ nlog_2 k \rfloor + 2$:

$$\begin{aligned}
&k^n v_1 2^{-(\lfloor n \log_2 k \rfloor + 2)} \textstyle\prod_{i=1}^{n} \left(1 + \frac{1}{kv_i}\right) < v_1 \\
&k^n \textstyle\prod_{i=1}^{n} \left(1 + \frac{1}{kv_i}\right) < 2^{(\lfloor n \log_2 k \rfloor + 2)}
\end{aligned}$$

This leads to the following general condition for the validity of the cycle-alpha's upper limit:

$$n \log_2 k - \lfloor n \log_2 k \rfloor < 2 - log_2 \left(\prod_{i=1}^{n} \left(1 + \frac{1}{kv_i}\right)\right) \tag{7}$$

A product $\prod(1 + a_n)$ with positive terms a_n is convergent if the series $\sum a_n$ converges, see Knopp [**?**, p. 220]. Thus, to verify whether the product in condition 7 is converging towards a limiting value, it is sufficient to examine the following sum:

$$\sum_{i=1}^{n} \frac{1}{kv_i}$$

Conclusion and outlook

We defined an algebraic graph structure that expresses the Collatz sequences in the form of a tree. Next, the vertex reachability properties were unveiled by examining the relationship between successive nodes in H_C. Moreover, we dealt with graphs that represent other variants of Collatz sequences, for instance $5x + 1$ or $181x + 1$. The interesting part of both variants just mentioned is that for these sequences the existence of cycles is known. With regard to a proof of the Collatz conjecture, theorem 3 seem promising. They serve as the basis for further investigations of the problem.

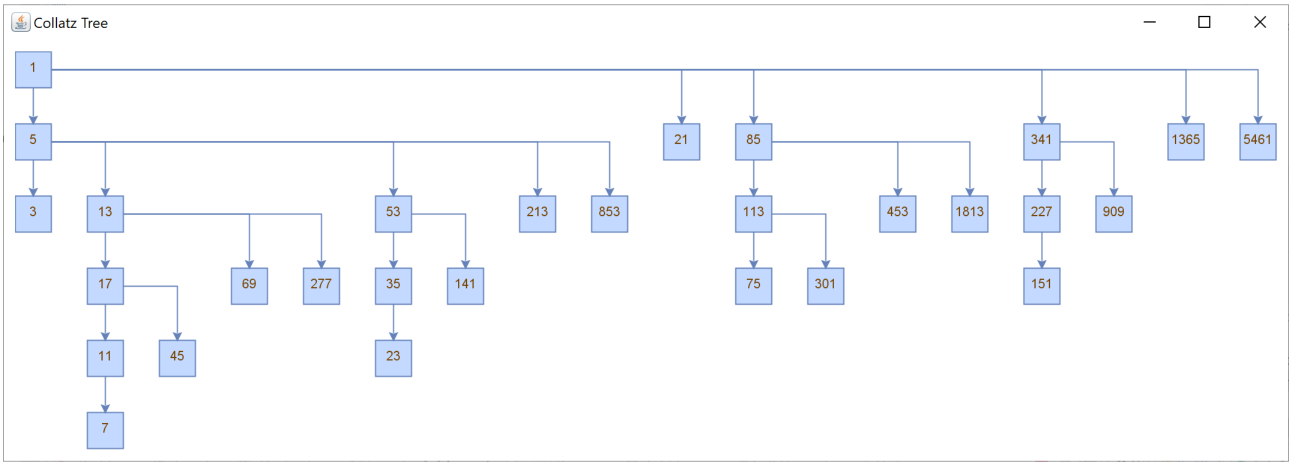

Fig. 1: Small section of H_C

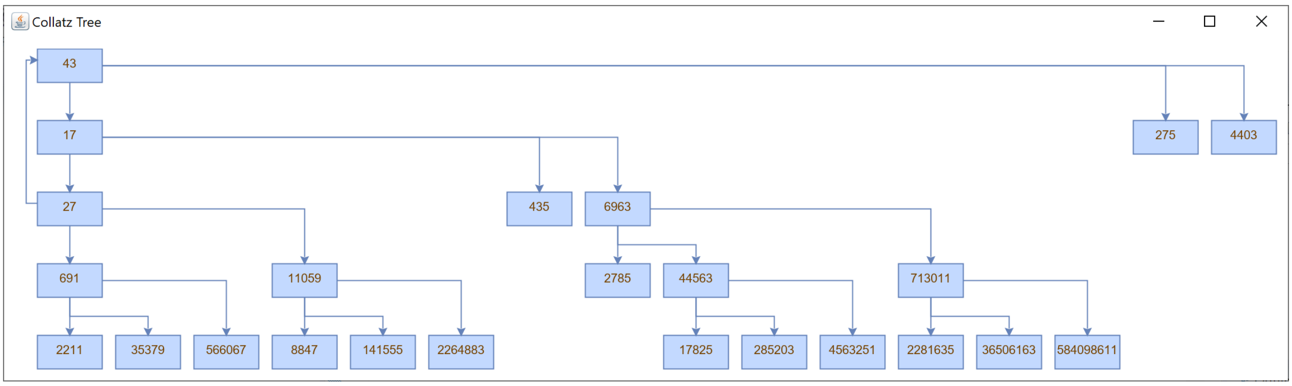

Fig. 2: Section of $H_{C,5}$ including the 3-cycle $43, 17, 27$

References

[1] M. Klisse, Das Collatz-Problem: Lösungs- und Erklärungsansätze für die 1937 von Lothar Collatz entdeckte (3n+1)-Vermutung. (2010)

[2] C. Löh, Geometric Group Theory: An Introduction. Springer, 2010, DOI https://doi.org/10.1007/978-3-319-72254-2

[3] J. A. Bondy and U. S. R. Murty: Graph Theory with Applications. Elsevier Science, 1976, ISBN 0-444-19451-7

[4] C. P. Bonnington and C. H.C. Little: The Foundations of Topological Graph Theory. Springer, 1995, DOl: 10.1007/978-1-4612-2540-9

[5] E. A. Bender and S. G. Williamson: Mathematics for Algorithm and System Analysis. Dover, 2005, ISBN 0-486-44250-0.

[6] M. Trümper: The Collatz Problem in the Light of an Infinite Free Semigroup. Chinese Journal of Mathematics, Volume 2014, DOI: http://dx.doi.org/10.1155/2014/756917

[7] J. Almeida: Profinite semigroups and applications. In V. B. Kudryavtsev and I. G. Rosenberg (eds.), Structural Theory of Automata, Semigroups, and Universal Algebra. Springer, 2005.

[8] R. Johnsonbaugh: Discrete Mathematics (Eighth Edition). Pearson, 2017, ISBN 0-321-96468-3.

[9] S. Mac Lane and G. Birkhoff: Algebra (Third Edition). AMS Chelsea Publishing, 1999, ISBN 0821816462.

[10] V. Novák, I. Perfilieva, and J. Močkoř: Mathematical Principles of Fuzzy Logic. Springer, 1999, DOI 10.1007/978-1-4615-5217-8

[11] R. Angot-Pellissier: The Relation Between Logic, Set Theory and Topos Theory as It Is Used by Alain Badiou. In A. Koslow and A. Buchsbaum (eds.), The Road to Universal Logic: Festschrift for the 50th Birthday of Jean-Yves Beziau (Volume II). Birkhäuser, 2015, DOI 10.1007/978-3-319-15368-1

[12] A. Ya. Helemskii: Lectures and Exercises on Functional Analysis. American Mathematical Society, 2006, ISBN 0-8218-4098-3

[13] K. H. Rosen: Discrete Mathematics and Its Applications (Seventh Edition). McGraw-Hill, 2011, ISBN 978-0-07-338309-5

[14] D. Makinson: Sets, Logic and Maths for Computing (Second Edition). Springer, 2012, DOI 10.1007/978-1-4471-2500-6

[15] B. Korte and J. Vygen: Combinatorial Optimization: Theory and Algorithms (Sixth Edition). Springer, 2018, DOI https://doi.org/10.1007/978-3-662-56039-6

[16] K. Mehlhorn and P. Sanders: Algorithms and Data Structures: The Basic Toolbox. Springer, 2008, DOI 10.1007/978-3-540-77978-0

[17] D.-Z. Du, K.-I Ko, and Z. Hu: Design and Analysis of Approximation Algorithms. Springer, 2012, DOI 10.1007/978-1-4614-1701-9

[18] H. Ehrig, K. Ehrig, U. Prange, and G. Taentzer: Fundamentals of Algebraic Graph Transformation. Springer, 2006, DOI 10.1007/3-540-31188-2

[19] L. N. Childs: A Concrete Introduction to Higher Algebra (Third Edition). Springer, 2006, DOI 10.1007/978-0-387-74725-5

[20] V. I. Voloshin: Introduction to Graph and Hypergraph Theory. Nova Science Publishers, 2011, ISBN 978-1-61470-112-5

[21] N. A. Loehr: Combinatorics (Second Edition). CRC Press, 2017, ISBN 978-1-4987-8025-4

[22] K. Conrow: The Structure of the Collatz Graph; A Recursive Production of the Predecessor Tree; Proof of the Collatz 3x+1 Conjecture. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.423.3396

[23] F. L. Bauer: Historische Notizen zur Informatik. Springer, 2009, DOI 10.1007/978-3-540-85790-7

[24] C. Hercher: Über die Länge nicht-trivialer Collatz-Zyklen. Die Wurzel, Hefte 6 und 7/2018