



HAL
open science

Winning tickets from two Collatz graphs' structures.

Hubert Schaetzel

► **To cite this version:**

| Hubert Schaetzel. Winning tickets from two Collatz graphs' structures.. 2024. hal-04499940

HAL Id: hal-04499940

<https://hal.science/hal-04499940v1>

Preprint submitted on 11 Mar 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

WINNING TICKETS FROM TWO COLLATZ GRAPHS' STRUCTURES

HUBERT SCHAETZEL

ABSTRACT. The Collatz conjecture states that, starting with any strictly positive integer, the $(3x+1)$ algorithm leads systematically to the same cycle $(1, 4, 2, 1, \dots)$ after a finite number of steps. The only potential exceptions to this rule on the positive side of \mathbb{Z}^* are the existence of either separate cycles or separate infinite divergent series. We will analyse the constraints and impediments on these types of objects using the underlying structures and laws linked to the Collatz algorithm. We will first prove the link of any integer to a tree structure with a unique root consisting of a cycle, therefore confirming the impossibility of some open series of integers coming down from infinity and growing back up to infinity. We will then prove that the previously mentioned cycle is unique within the natural numbers by examining the two limit cases deduced from the parity vectors of the smallest element of any cycle with w elements, by using the cycles' rational formal solutions and by exploiting the Diophantine approximations' theory and a result linked to the property of the gcd of rational numbers that allows here to encompass all intermediary cases.

CONTENTS

1. The algorithm	2
2. Conventions and vocabulary	2
Part 1. The trees' structure	3
3. Graph crowns	3
4. Roots	4
5. Graphs	7
Part 2. The parity vectors' structure	9
6. Cycles' classification	9
7. Limit cases	13
8. The Collatz cycles' scarcity	16
Appendix A. Continued fraction of $\text{Ln}(3)/\text{Ln}(2)$	27
Appendix B. Evaluation of the gcd linked ratio	28
Literature and sources	29

Date: October 22, 2023.

2020 Mathematics Subject Classification. 11Y55, 40A05.

Key words and phrases. Collatz conjecture, Syracuse conjecture, graph.

1. THE ALGORITHM

Let us have i an index and u_i an integer different from zero (positive or negative). The Collatz algorithm consist in dividing u_i by 2 if u_i is even and to multiply it by 3 and adding 1 otherwise in order to get u_{i+1} recursively.

$$u_{i+1} = \begin{cases} u_i/2 & \text{if } u_i \equiv 0 \pmod{2}, \\ 3u_i + 1 & \text{if } u_i \equiv 1 \pmod{2}. \end{cases}$$

According to the Collatz conjecture, starting with any integer in \mathbb{N}^* , this algorithm leads to the same cycle (1, 4, 2, 1, \dots).

2. CONVENTIONS AND VOCABULARY

Vocabulary. *Part 1.*

We use standard graph vocabulary and a few additional conventions

- vertex : *any integer,*
- successor : *a successor vertex is obtain by applying the Collatz algorithm to an integer; an integer has one immediate successor; when speaking of successor in the singular we mean the immediate successor,*
- antecedent : *an antecedent vertex is obtain by applying the Collatz algorithm to an integer in the reverse way (upturn or upwards); an integer has one or two immediate antecedents according to its value; when speaking of antecedent in the singular we mean an immediate antecedent,*
- active vertex : *a green or blue integer (an integer equal to 1 or 2 modulo 3),*
- inactive vertex : *a yellow vertex (an integer equal to 0 modulo 3),*
- link : *an edge between two vertices,*
- active link : *a link giving a blue or green antecedent,*
- branch : *a set of vertices connected by some links,*
- inactive branch : *a branch with only yellow vertices,*
- graph : *an arbitrary initial choice of a vertex or cycle and then all the vertices and links formed by its successors and antecedents,*
- inactive graph : *a graph with ultimately only inactive ascendant branches; note : such graph has only a finite number of branches,*
- root : *the cycle or unique vertex (the latter being proven impossible) at the bottom of a graph,*
- rank : *the upturn step of antecedents from the root ; by extension, the term is also used for the number of upturn steps starting from some chosen integer,*
- graph crown : *the set of vertices and links of a graph except its root (or initial integer).*

Color code. Let us consider the values of the integers modulo 3. We associate the green color to the 1 (mod 3) integers, the blue color to the 2 (mod 3) integers and the yellow color to the 0 (mod 3) integers.



Vocabulary. Part 2.

- integers : \mathbb{N}^* designates the natural numbers, while \mathbb{N} includes 0. The same convention holds for \mathbb{Z}^* and \mathbb{Z} for integers.
- stopping time : executing the Collatz algorithm, the step when the absolute value of the resulting integer is equal or smaller than the chosen initial integer.
- odd step : a multiplicative operation $(3x+1)$ on x . The total number of odd steps is noted v at the stopping time.
- even step : a division operation $(x/2)$ on x . The total number of even steps is noted w at the stopping time.

Note. We will use regularly, in the second part of the article, the term "asymptotically" for diverse events. The reader will notice that such cases may however often occur quite soon despite its intrinsic remote meaning.

Part 1. The trees' structure

3. GRAPH CROWNS

Let us have the color code defined previously. There are 1 or 2 immediate antecedents for any integer and it is then straightforward to get the modulo values of these immediate antecedents and corresponding colors according to the six cases given in figure 1.

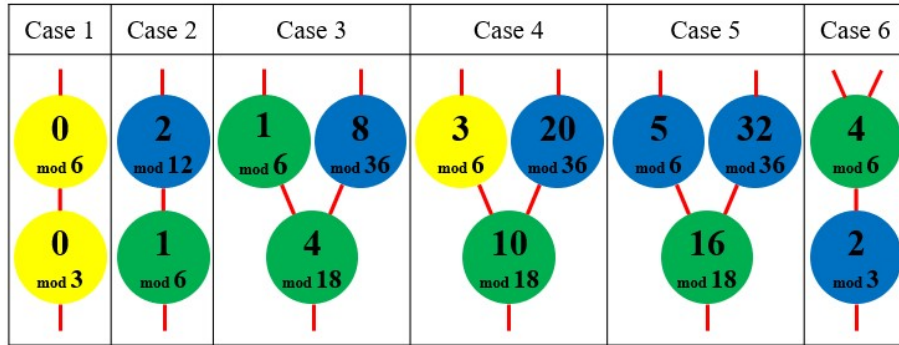


FIGURE 1. Collatz reverse algorithm : List of antecedents' modulo values.

The initial set of integers, of course, is chosen in order to cover a complete set of congruences:

Case 1	0 (mod 3)		
Case 2	1 (mod 3)	1 (mod 6)	
Case 3		4 (mod 6)	4 (mod 18)
Case 4			10 (mod 18)
Case 5			16 (mod 18)
Case 6	2 (mod 3)		

It is likewise easy to verify that the antecedents' set is also a complete set of congruences:

Case 1	0 (mod 6)		
Case 3	1 (mod 6)		
Case 2	2 (mod 6)	2 (mod 12)	
Case 3		8 (mod 12)	8 (mod 36)
Case 4			20 (mod 36)
Case 5			32 (mod 36)
Case 4	3 (mod 6)		
Case 6	4 (mod 6)		
Case 5	5 (mod 6)		

Lemma 1. *There is no blue antecedent to a blue vertex.*

Proof. Case 6 in figure 1 is the only alternative. It shows green vertices as antecedents. \square

Lemma 2. *A green vertex may have a green antecedent vertex. But the later cannot have another green antecedent.*

Proof. Case 3 followed by case 2 in figure 1 is the only alternative and shows a blue vertex after two consecutive green vertices. \square

Lemma 3. *Any branch is linked to an active branch.*

Proof. Recall, by our earlier definitions, a inactive branch contains only yellow vertices. The antecedent of a yellow vertex is unique (case 1 of figure 1) and is yellow, therefore the branch is inactive up to infinity once a yellow vertex appears. But the bottom yellow vertex of that branch has necessarily a green successor which is linked upwards to a blue vertex (case 4 of figure 1) and then again upwards to a green vertex (case 6 of figure 1). This last green vertex is equivalent to 4 (mod 6) and therefore has always 2 antecedents providing the début of an active branch as for there on, whatever follows, one encounters always at least a blue or a green vertex as antecedent, a yellow vertex never appearing alone (but with a blue vertex as shown in case 4). \square

4. ROOTS

Figure 1 shows that any integer has two or three links. The 3 links' pattern is the one that allows roots to thrive with a crown graph. The objective of this section is to prove that this blossom will always occurs : Any root has a crown graph.

Lemma 4. *The two lemmas 1 and 2 apply also in a cycle.*

Proof. The Collatz algorithm is the same in any circumstances (in cycles as in linear branches). \square

Lemma 5. *There is no root with one vertex. The only root with two vertices is $(-1, -2)$. The only root with three vertices is $(1, 2, 4)$. There is no root with four vertices. The only root with five vertices is $(-5, -14, -7, -20, -10)$. There is no root with six vertices.*

Proof. Let us verify all the 2^1 possibilities for one vertex $u_0 \rightarrow u_{-1} = u_0$ going from the initial integer upwards antecedents: we get $u_0 \rightarrow u_0 = \text{or}(2u_0, (u_0 - 1)/3)$. Hence $u_0 = \text{or}(0, -1/2)$ and therefore no solutions in \mathbb{Z}^* . The 2^2 possibilities for two vertices are $u_0 \rightarrow u_{-1} \rightarrow u_{-2} = u_0$ so that $u_0 \rightarrow \text{or}(2u_0, (u_0 - 1)/3) \rightarrow \text{or}(4u_0, 2(u_0 - 1)/3, (2u_0 - 1)/3, ((u_0 - 1)/3 - 1)/3)$. Hence $u_0 = \text{or}(0, -1, -2, -1/2)$. Only -1 and -2 are in our domain of definition \mathbb{Z}^* and give effectively a cycle with two vertices. For three vertices, the 2^3 initial solutions are $(0, 1, 2, -4/7, 4, -5/7, -8/7, -1/2)$ where only $(1, 2, 4)$ are in \mathbb{Z}^* and is effectively a cycle with 3 vertices. For four vertices, the 2^4 initial solutions are $(0, 1/5, 2/5, -4/5, 4/5, -1, -8/5, -13/25, 8/5, -7/5, -2, -14/25, -16/5, -17/25, -26/25, -1/2)$ where only $(-1, -2)$ are in \mathbb{Z}^* and is effectively a cycle, but only with 2 vertices, a redundancy with the previous search. Similarly, one can resolve the five and six vertices' cases. \square

Another cycle with 18 vertices is known in \mathbb{Z}^- to this day. Of course, to solve it with the given previous method would be quite cumbersome and painful due to the $2^{18} = 262144$ equations to solve. The 4 known cycles are represented in figures 3 to 6.

Lemma 6. *There is no root containing a yellow vertex.*

Proof. The antecedent of an integer equal to 0 (mod 3) is unique and double its value (case 1 of figure 1). The next antecedent likewise and so up to infinity. Thus it cannot cycle back to its initial value. \square

Lemma 7. *In a root, there cannot be 2 blue vertices next to each other. In a root, there can be possibly 2 green vertices next to each other, but not 3.*

Proof. This is an immediate result of lemma 4 and figure 1. \square

Lemma 8. *In a root, a blue vertex has no link towards the outside of the root. In a root, an isolated green vertex has always a link towards the outside of the root. In a root, a pair of green vertex has one and only one of the vertex with a link towards the outside of the root.*

Proof. One gets again all the information from figure 1. For the blue vertex, which can only have two links (case 6), the two have to be inner links to get a cycle. For the isolated green vertex, the only contradictory case would be case 2, but then its value is equal to 1 (mod 6) and it is therefore linked

to the green vertex on the top of case 3 which will have a blue link to the outside. \square

Lemma 9. *Any root has active links towards the outside.*

Proof. For roots with less than six vertices, refer to figures 3 to 5 to confirm the claim. We know by figure 1 (cases 3, 4 and 5) that outside links from roots can only grow from green vertices. So then suppose that yellow vertices are growing out from two possible most narrow links. Figure 2 shows the two only possible cases of pieces of roots one can get on these premises. For each separate case in this figure, the vertices on the left side are those within the root itself and the vertices on the right side are the first items out of that root. White vertices may be green or blue, it doesn't matter. From figure 1, we know that the annotated green vertices must equal $10 \pmod{18}$ as the only case with two antecedents of which one is yellow is case 4. We then consider the following two alternatives.

Case 1: Start from the first green vertex towards the top equal to $10 + 18k_1$. It is even thus the vertex underneath is equal to $5 + 9k_1$. The next vertex is then either equal to $(5 + 9k_1)/2$ or $16 + 27k_1$. This vertex in the same time must equal $10 + 18k_2$. Therefore either $3(k_1 - 4k_2) = 5$ or $3(3k_1 - 2k_2) = -2$ which are both impossible with $k_1, k_2 \in \mathbb{Z}$.

Case 2: Start again from the first green vertex equal to $10 + 18k_1$. The vertex underneath is equal to $5 + 9k_1$. The next vertex is then either equal to $(5 + 9k_1)/2$ or $16 + 27k_1$. The next vertex is then either $(5 + 9k_1)/4$, $3(5 + 9k_1)/2 + 1$, $(16 + 27k_1)/2$ or $3(16 + 27k_1) + 1$. This last vertex in the same time must equal $10 + 18k_2$. Therefore either $9(k_1 - 8k_2) = 35$, $3(3k_1 - 4k_2) = 1$, $9(3k_1 - 4k_2) = 4$ or $3(9k_1 - 2k_2) = -13$ which are again all impossible. So we cannot have "adjacent" yellow vertices stemming from a root. We know also from lemma 8 that a supplementary intermediate blue-green-blue vertices' sequence without external link is impossible. Now considering a complete root, we know by lemma 7 that there are at least half of green vertices in any root. If these green vertices are systematically by pairs (which is certainly an absurd situation that we have not seek to object), there are anyway still at least $1/3$ of the vertices in the root having links with a crown graph. Now from the above discussion, less than half of these links are yellow. Therefore, the number of active links is at least $1/6$ of the root's cardinal. Thus with more than 6 vertices, a root has necessarily active links. \square

For the four known cycles, as the reader can check directly, all the first links towards the outside of the root are active links. Shalom Eliahou [1] has proven that any unknown cycle in \mathbb{N}^* would contain at least 17026679261 vertices (for the elements of the cycle only) and therefore would have, thanks to our own above study, more than 2837779877 active links.

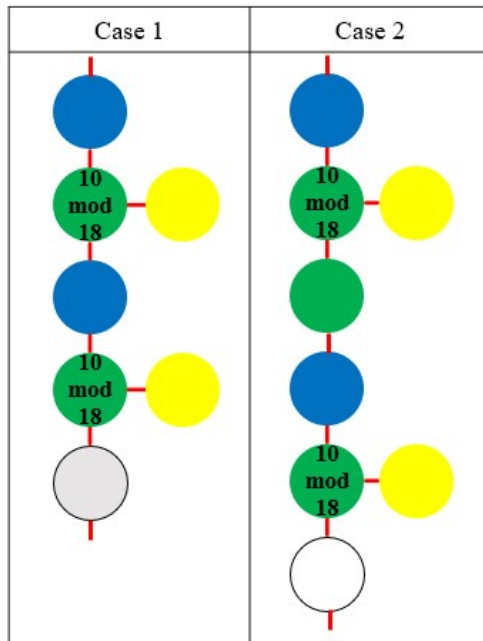


FIGURE 2. Pieces of roots.

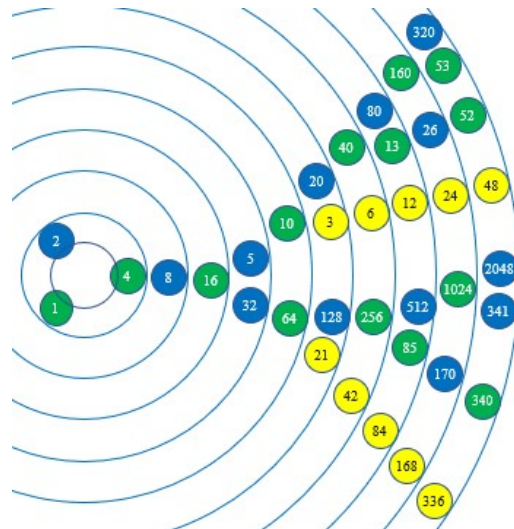
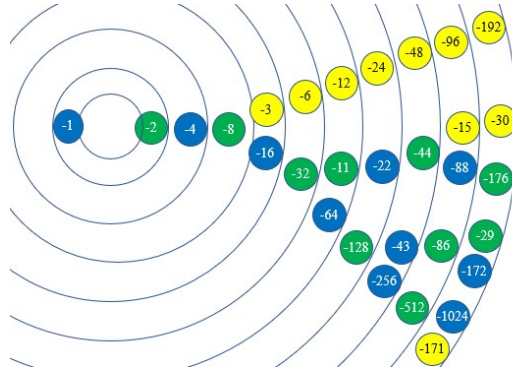


FIGURE 3. Cycle 1.

5. GRAPHS

Now that we know that any root has a crown graph, let us go further on this blossoming pattern.

Lemma 10. *A graph has one and only one root.*



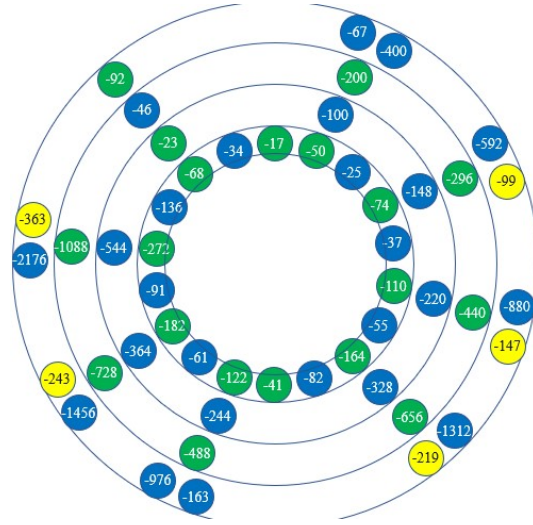


FIGURE 6. Cycle 4.

Case 2: Starting from any inactive vertex, its successors (in downwards Collatz algorithm) are divided by 2 as long as the result is even, otherwise the successor gets green and is hence an active vertex bringing us back to case 1. \square

Theorem 1. *Any integer belongs to an active graph with a unique root.*

Proof. This is the aggregated result of lemmas 9, 10, 11 and 12. \square

This proves our first point that there is no divergent series stemming from some given integer nor some series coming down from infinity and growing back up to infinity. There is always a root on the base of a Collatz tree structure and this root is unique.

Part 2. The parity vectors' structure

6. CYCLES' CLASSIFICATION

Lemma 13. *The composition of linear functions is a linear function.*

Proof. The lemma is obvious, but let us develop the precise result. Let us have $T_k(x) = a_kx + b_k$, $k = 1$ to i , a series of linear functions and let us consider $CT_i(x) = T_i \circ T_{i-1} \circ \dots \circ T_1(x)$. Then $CT_1(x) = a_1x + b_1$, $CT_2(x) = a_2a_1x + a_2b_1 + b_2$, ... and $CT_i(x) = a_i a_{i-1} \dots a_1 x + a_i a_{i-1} \dots a_2 b_1 + a_i a_{i-1} \dots a_3 b_2 + \dots + a_i a_{i-1} \dots a_{i-3} b_{i-4} + a_i a_{i-1} a_{i-2} b_{i-3} + a_i a_{i-1} b_{i-2} + a_i b_{i-1} + b_i$. \square

Lemma 14. *The composition of j linear functions with $a_k = 1/2$ and $b_k = 0$ and $i - j$ linear functions with $a_k = 3/2$ and $b_k = 1/2$, in that specific order,*

is equivalent to the linear function :

$$LT_i(x) = \frac{3^{i-j}}{2^i} (x + 1 - (\frac{2}{3})^{i-j}) \quad (1)$$

Proof. We have $T_k(x) = (3x + 1)/2$, $k = 1$ to $i - j$ and $T_k(x) = x/2$, $k = i - j + 1$ to i . Using the previous lemma, we get $LT_i(x) = (3^{i-j}x + 3^{i-j-1}2^0 + 3^{i-j-2}2^1 + 3^{i-j-3}2^2 + \dots + 3^3 2^{i-j-4} + 3^2 2^{i-j-3} + 3^1 2^{i-j-2} + 3^0 2^{i-j-1})/2^i = 3^{i-j} 2^{-i} (x + 3^{-1}(1 + (2/3)^1 + (2/3)^2 + \dots (2/3)^{i-j-1}))$. Thus the former result. \square

Lemma 15. *The function $HT_i(x) = LT_i(x)/x$ is an hyperbolic function, therefore strictly monotonous, defined everywhere except for $x = 0$. Its value is equal to 1 for the unique solution :*

$$x = -\frac{1 - (\frac{2}{3})^{i-j}}{1 - 2^j (\frac{2}{3})^{i-j}}$$

Posing $v = i - j$ the number of $3x + 1$ multiplications and $w = i$ the number of divisions by 2, we get also :

$$x = -\frac{1 - (\frac{2}{3})^v}{1 - 2^{w-v} (\frac{2}{3})^v} \quad (2)$$

Proof. We have $T_k(x) = 3^{i-j} 2^{-i} (1 + (1 - (2/3)^{i-j})(1/x))$ which is obviously a hyperbolic function. Its derivative is equal to $-3^{i-j} 2^{-i} (1 - (2/3)^{i-j})(1/x^2)$, therefore of the sign of the constant expression $-(1 - (2/3)^{i-j})$. Solving $T_k(x) = 1$ gives immediately the result x given in the lemma. \square

Lemma 16. *Let us pose*

$$w = \lfloor \frac{\ln(3)}{\ln(2)} v \rfloor + 1 - incr \quad (3)$$

using the floor function and $incr$ being an integer. Then

$$\begin{aligned} \text{if } incr > 1, \quad & -2 < x < 0, \\ \text{if } incr < 0, \quad & 0 < x < 1. \end{aligned}$$

Proof. Let us go back to equation 2. We get immediately $x = -(1 - (2/3)^v)/(1 - 2^w/3^v)$. For small values de v , we verify the proposition numerically and figure 7 illustrates the point. If $v \gg 1$, as v diverges, the numerator $1 - (2/3)^v$ will tend towards 1^- . Then $0 > x > -1/(1 - 1/2) = -2$ if $2^w/3^v < 1/2$. Solving $2^w/3^v < 1/2$, we get $w < (\ln(3)/\ln(2))v - 1$. Then replacing w with the expression of the lemma, we get $\lfloor (\ln(3)/\ln(2))v \rfloor + 1 - incr < (\ln(3)/\ln(2))v - 1$, therefore $1 \leq \lfloor (\ln(3)/\ln(2))v \rfloor - (\ln(3)/\ln(2))v + 2 < incr$ which is the announced lower limit value of $incr$. Studying the second condition, we observe that $0 < x < -1/(1 - 2) = 1$ if $2^w/3^v > 2$. Solving $2^w/3^v > 2$, we get $w > (\ln(3)/\ln(2))v + 1$. Then replacing w with the expression of the lemma, we get $\lfloor (\ln(3)/\ln(2))v \rfloor + 1 - incr > (\ln(3)/\ln(2))v + 1$, therefore $0 \geq \lfloor (\ln(3)/\ln(2))v \rfloor - (\ln(3)/\ln(2))v > incr$ which is this time the announced highest limit value of $incr$. \square

Note. The value of x tends towards -1 when $incr$ increases asymptotically ($incr \rightarrow +\infty$). The value of x tends towards 0 when $incr$ decreases asymptotically ($incr \rightarrow -\infty$).

Note. The figures 8 and 9 illustrate the two cases $incr = 0$ and $incr = 1$. The ordinates are in these two cases in logarithmic scales ($\ln(x)$ for $incr = 0$ and $\ln(-x)$ for $incr = 1$).

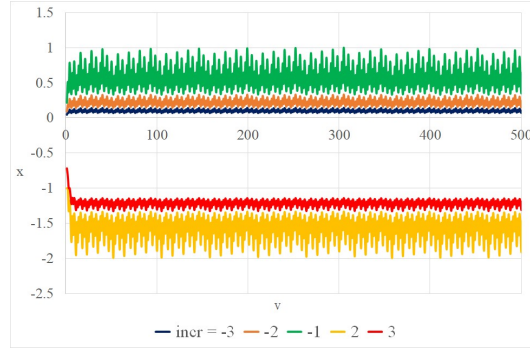


FIGURE 7.
Solutions to $HT_w(x) = 1$,
 $incr = -3, -2, -1, 2$ and 3 .

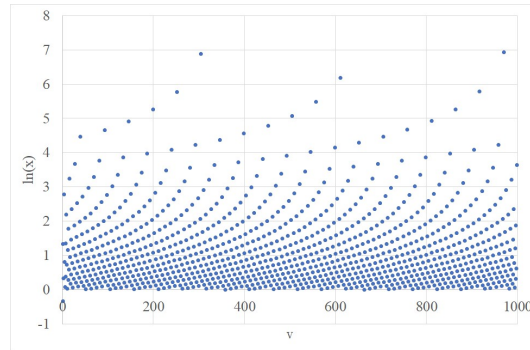
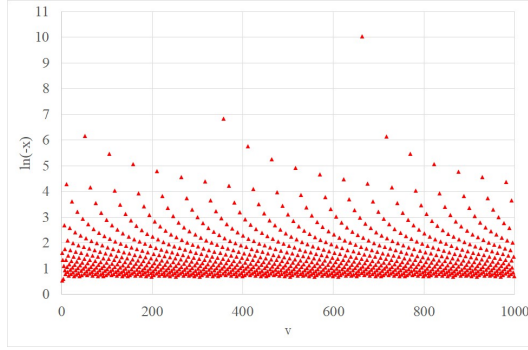


FIGURE 8. Solutions to $HT_w(x) = 1$, $incr = 0$.

Lemma 17. *The previous solution x is the largest in absolute value to the equation $PT_k(y)/y = 1$, where $PT_k(y)$ is any permutation of the composition $T_i \circ T_{i-1} \circ \dots \circ T_1(x)$ keeping here the same number of $(3x+1)/2$ multiplications and $x/2$ divisions.*

Proof. Let us have $a = 3^{i-j-1}2^0 + 3^{i-j-2}2^1 + 3^{i-j-3}2^2 + \dots + 3^3 2^{i-j-4} + 3^2 2^{i-j-3} + 3^1 2^{i-j-2} + 3^0 2^{i-j-1}$, thus $HT_i(x) = (3^{i-j}x + a)/2^i x = 1$ has solution $x = a/(2^i - 3^{i-j})$. Here the denominator has a fixed value and therefore the absolute value of x is maximal if the absolute value of a diminish when

FIGURE 9. Solutions to $HT_w(x) = 1$, $incr = 1$.

the permutation is applied (giving a smaller alternative value y). In order to get the final composition of the linear functions, we apply a finite number of elementary permutations such that each one switches two members $3^{n_1}2^{m_1}$ and $3^{n_2}2^{m_2}$ to $3^{n_1-1}2^{m_1+1}$ and $3^{n_2+1}2^{m_2-1}$, where $n_1 > n_2$ and $m_1 < m_2$, systematically reducing the value of the initial a (because $3 > 2$). Thus the result. \square

Theorem 2. *The Collatz algorithm may lead to a cycle in \mathbb{N}^* if and only if the number of $(3x+1)$ multiplications, noted v , to the number of $(x/2)$ divisions, noted w , meets the condition*

$$\text{Type1} : w = \lfloor \frac{\ln(3)}{\ln(2)} v \rfloor + 1 \quad (4)$$

and may generate a cycle in $\mathbb{Z} - \mathbb{N}$ if and only if it meets the condition

$$\text{Type0} : w = \lfloor \frac{\ln(3)}{\ln(2)} v \rfloor. \quad (5)$$

Proof. According to the lemmas 16 and 17, which hold in \mathfrak{R} and therefore also in \mathbb{Z} , the largest absolute value y to equation $PT_w(y) = 1$ is smaller than 2 (for any value of $incr$ different from 0 or 1), a finite interval that one can easily totally explore numerically for solutions and check that there are none. One can then conclude on the value of w in \mathbb{Z}^* using again lemma 16 which distinguishes the two cases on $incr$. \square

Note. Checking the known cycles (except 0 which meets the type 1), using $\ln(3)/\ln(2) \approx 1.58496$, we get the following numerical results

Type 1 : $y = 1$, $v = 1$, $w = 2$ and $2 = \lfloor (\ln(3)/\ln(2)).1 \rfloor + 1$,

Type 0 : $y = -1$, $v = 1$, $w = 1$ and $1 = \lfloor (\ln(3)/\ln(2)).1 \rfloor$,

Type 0 : $y = -5$, $v = 2$, $w = 3$ and $3 = \lfloor (\ln(3)/\ln(2)).2 \rfloor$,

Type 0 : $y = -17$, $v = 7$, $w = 11$ and $11 = \lfloor (\ln(3)/\ln(2)).7 \rfloor$.

The reader may refer to the figures 3 to 6 to check that the cycles meet the number of $(3x+1)$ multiplications and $(x/2)$ divisions.

7. LIMIT CASES

Lemma 18. *The ratio $\ln(3)/\ln(2)$ is irrational.*

Proof. Let us suppose $\ln(2)/\ln(3) = p/q$, where p and q are integers. Then $q.\ln(2) = p.\ln(3)$, so that $\ln(2^q) = \ln(3^p)$ and finally $2^q = 3^p$, which is obviously false. Thus $\ln(3)/\ln(2) \notin \mathbb{Q}$. \square

Lemma 19. *Let us consider x_0 any positive integer. Applying v odd steps and w even steps in some order of the Collatz algorithm to x_0 , the nearest result y_0 to the initial integer x_0 among all combination of the said odd and even steps, is equal to*

$$y_0 = \frac{3^v}{2^w}x_0 + \frac{1}{2^{w-v}}\left(\left(\frac{3}{2}\right)^v - 1\right) \quad (6)$$

Proof. Applying first all the $(3x+1)/2$ multiplications, we get $y_0 = (3^v x_0 + 3^{v-1}2^0 + 3^{v-2}2^1 + 3^{v-3}2^2 + \dots + 3^1 2^{v-2} + 3^0 2^{v-1})/2^v / 2^{w-v} = (3^v/2^w)x_0 + (1/2^{w-v})((3/2)^v - 1)$. Applying first all the $(x/2)$ divisions, we get $y_0 = (3^v x_0 + 3^{v-1}2^{w-v} + 3^{v-2}2^{w-v+1} + 3^{v-3}2^{w-v+2} + \dots + 3^1 2^{w-2} + 3^0 2^{w-1})/2^{w-v} / 2^v = (3^v/2^w)x_0 + ((3/2)^v - 1)$. The other combinations give intermediary values between these two results and the first expression is the nearest result y_0 to x_0 because of the additional ratio $(1/2^{w-v})$ smaller than 1 in front of $((3/2)^v - 1)$. \square

Lemma 20. *The ratio x_w/x_0 , where $x_0 > 0$ and x_w is resulting from a Collatz algorithm is systematically such that*

$$\frac{x_w}{x_0} > \frac{3^v}{2^w}$$

Proof. The nearest y_0 to x_0 means that $x_w/x_0 \geq y_0/x_0$. Thus using the previous lemma result

$$\frac{x_w}{x_0} \geq \frac{y_0}{x_0} = \frac{3^v}{2^w} + \frac{1}{x_0} \frac{1}{2^{w-v}} \left(\left(\frac{3}{2} \right)^v - 1 \right) > \frac{3^v}{2^w}.$$

\square

Lemma 21. *Let us have x_0 , x_w , v and w some fixed strictly positive values meeting the condition of lemma 20, that is $3^v/2^w < x_w/x_0$. The function $f(k) = (x_w + k.3^v)/(x_0 + k.2^w)$ is continuous over positive or null k , decreasing monotonously from x_w/x_0 towards $3^v/2^w$, the later an asymptotic value.*

Proof. The function $f(k)$ is a hyperbolic function which is undefined at the unique strictly negative value $k = -x_0/2^w$, therefore is continuous on \mathbb{R}^+ . The derivative is $f'(k) = (x_0.3^v - x_w.2^w)/(x_0 + k.2^w)^2$, therefore of the sign of $x_0.3^v - x_w.2^w$ which is strictly negative by the chosen hypothesis. The function therefore evolves over \mathbb{R}^+ monotonously from x_w/x_0 at $k = 0$ to the limit value $3^v/2^w$ when $k \rightarrow \infty$. \square

Lemma 22. *Let us have some positive integer x_0 and x_w its result by the Collatz algorithm at its stopping time. If the stopping time is finite then the ratio x_w/x_0 is such that $1/2 < x_w/x_0 \leq 1$.*

Proof. By definition of the stopping time, we have $x_w/x_0 \leq 1$. The last step of the algorithm is necessarily an even step which either gives exactly the value of x_0 or a strictly greater value of its half. Using $(3x+1)/2$ multiplications and $(x/2)$ divisions, each step of the process includes a division by 2, therefore w is the appropriate index to count them and v will be the number of multiplications. \square

Lemma 23. *Let us consider the set $\{x_0+k.2^w, k \in \mathbb{Z}\}$. Then, if x_w exist for a finite w , the elements of the set $\{x_w+k.3^v, k \in \mathbb{Z}\}$ are the resulting values of the initial set at their respective stopping time and moreover $x_0 < 2^w$ (recall also that $0 < x_0$ by hypothesis) and $1/2 < (x_w+k.3^v)/(x_0+k.2^w) \leq 1$.*

Proof. This is mostly a well-known result but is worth reviewing. By hypothesis, it is clear that the elements resulting from x_0 being divisible w times by 2 then $x_0+k.2^w$ is also divisible in the same condition w times. Moreover at each step the distance 2^t between the intermediary results $x_i+k.2^t$ is constant and equal to $2^{w-s}3^m$ where s is the number of steps at that stage and m the number of $(x+1)/2$ multiplications, hence a distance 2^03^v at the stopping time. Now for $k=0$, the number of steps v and w is necessarily such that $3^v/2^w \leq 1$ (in fact $3^v/2^w < 1$ by lemma 18) but in the closest way as a division by 2 is always the last step of the stopping time process and therefore $3^v/2^w > 1/2$. Now according to lemma 21, $x_w/x_0 > 3^v/2^w$ so that $x_0.3^v - x_0.x_w < x_w.2^w - x_0.x_w$ and so $x_0(3^v - x_w) < x_w(2^w - x_0)$ is equivalent to $(3^v - x_w)/(2^w - x_0) < x_w/x_0$ because $x_0 < 2^w$ providing the first sample of the $(x_w+k.3^v)/(x_0+k.2^w)$ were k is negative. Here $k = -1 < -x_0/2^w$ which is the undefined abscissa of the hyperbolic function. Therefore, as we know that the function is strictly decreasing, the ratio is increasing from $(3^v - x_w)/(2^w - x_0)$ up asymptotically towards $3^v/2^w$ as $k \rightarrow -\infty$. \square

Theorem 3. *At its stopping time, for a finite non-cyclic event, the number of $(3x+1)$ multiplications, noted v , to the $(x/2)$ divisions, noted w , is such that*

$$w = \lfloor \frac{\ln(3)}{\ln(2)} v \rfloor + 1 \quad (7)$$

over the whole domain \mathbb{Z} .

Proof. According to lemma 23, $1/2 < (x_w+k.3^v)/(x_0+k.2^w) \leq 1$ and the elements of set $x_0+k.2^w, k \in \mathbb{Z}$ have all the same number of odd and even steps at the stopping time. We get also, for $k=0$, $1/2 < 3^v/2^w \leq 1$ which is equivalent to $w = \lfloor \frac{\ln(3)}{\ln(2)} v \rfloor + 1$. \square

Lemma 24.

$$\frac{1}{2} < \frac{3^v}{2^w} < 1 \Leftrightarrow w = \lfloor \frac{\ln(3)}{\ln(2)} v \rfloor + 1$$

Proof. The only point to complete from the previous proof is that $\frac{2^w}{3^v} \neq 1$ which is obvious. \square

Having established the type of most of the integers, let us study the underlying structure derived from the stopping time w which depends on v by the relationship $w = \lfloor \frac{\ln(3)}{\ln(2)}v \rfloor + 1$. Starting from the set \mathbb{Z} , let us remove all the elements such that $\{v = 0, w = 1\}$ that are in the interval $[0, 2^w - 1 = 1[$. Only 0 complies and the other elements satisfying $\{v = 0, w = 1\}$ are separated by a distance $2^w = 2$, therefore the even integers. Then we discard the elements such that $\{v = 1, w = 2\}$ in the interval $[0, 2^w - 1 = 3[$, where only 1 meets the requirement and those separated by a distance $2^w = 4$ from the formers, hence all $1 \pmod 4$ integers. Going to step $\{v = 2, w = 4\}$, we consider the integers in interval $[0, 2^w - 1 = 15[$, where only 3 meets the requirement and the complement separated by a distance $2^w = 16$, hence all $3 \pmod 16$ integers. At next step $\{v = 3, w = 5\}$, we consider the integers in interval $[0, 2^w - 1 = 31[$, where only 11 and 23 meets the requirement and all those separated by a distance $2^w = 32$, hence all $11 \pmod 32$ and $23 \pmod 32$ integers. This removal process is illustrate in the table underneath by lowering the initial integers to the corresponding v -indexed line. Of course the integers $-1, -5$ and -17 in red in the second line can never be affected as those comply with $w = \lfloor \frac{\ln(3)}{\ln(2)}v \rfloor$ instead of $w = \lfloor \frac{\ln(3)}{\ln(2)}v \rfloor + 1$.

TABLE 1

v	-19	-17	-15	-13	-11	-9	-7	-5	-3	-1	1	3	5	7	9	11
		-17						-5		-1						
1	-19		-15		-11		-7		-3		1		5		9	
2				-13								3				
3						-9										11

Definition. We will call parity vector the binary representation 1 or 0 of the sequence of odd (as $(3x+1)/2$ multiplications) and even steps of the Collatz algorithm applied to some initial integer until the stopping time in the order of apparition. The parity vector is of size w .

It includes v digits 1 and $w - v$ digits 0.

Definition. We call a licit parity vector that one that doesn't break any rule of the Collatz algorithm during an altitude flight routine.

Let us consider, for example, the parity vector 100. It is not licit because the number of even steps written here (that is 2) is greater than the correct value $w - v = \lfloor (\ln(3)/\ln(2).v) \rfloor + 1 - v$ here (that is $1 + 1 - 1 = 1$). The altitude flight time is exceeded in this writing. Similarly, writing 10111100 is not licit, even if we do have globally $w - v = \lfloor (\ln(3)/\ln(2).v) \rfloor + 1 - v = 7 + 1 - 5 = 3$, as the altitude flight time is met prematurely by writing 10 at the beginning of the sequence. For some given parity vector, in the same way, one has to check its validity at each new even intermediate step. For example, with parity

vector 111111101000100, the three intermediate necessary checks for a premature non licit parity vector are the following ones between corresponding parentheses $((11111110)1000)100$. We have $8-7 = 1 < \lfloor (\ln(3)/\ln(2)).v1 \rfloor + 1 - v1 = 12+1-7 = 6$, $12-8 = 4 < \lfloor (\ln(3)/\ln(2)).v2 \rfloor + 1 - v2 = 13+1-8 = 5$ and $15-9 = 6 = \lfloor (\ln(3)/\ln(2)).v3 \rfloor + 1 - v3 = 14+1-9$ thus corresponding effectively to a licit parity vector.

This summarizes as follows.

Lemma 25. *The rule linking w to v being respected, there are two limit cases for the licit parity vectors. The first one is where the components 1 are all on the left and the components 0 follow.*

$$\underbrace{11\dots1}_{v \text{ times}} \underbrace{00\dots0}_{w-v \text{ times}}$$

The second is where the components 1 are shifted on the right in such a way that at each step it stays a licit vector. For those, an easy algorithm is proposed in order to construct them. It consist to use the limit parity vector at step $v-1$ and replace the last 0 by 1 and complete to the right with the necessary number of 0 to get $w-v$ of them in total.

10
1100
11010
1101100
11011010
1101101100
...

Proof. The proof is in the lemma's self-explanation. □

Note. Theorem 3 and Table 1 explain why there can be more than one cycle within the negative integers. It is because the smallest element of the cycles is "outside" the standard parity vectors' structure which is subject to $w = \lfloor \frac{\ln(3)}{\ln(2)}v \rfloor + 1$.

8. THE COLLATZ CYCLES' SCARCITY

Lemma 26. *The smallest rational value solution to the smallest element x_0 of a cycle on the \mathbb{N}^* side of \mathbb{Z} is equal to*

$$x_0 = \frac{3^v - 2^v}{2^w - 3^v}. \tag{8}$$

Proof. As solution to a hyperbolic equation, the cycle solution to some given order combination of odd and even steps is unique. In lemma 19, we got the following intermediary result $y_0 = (3^v/2^w)x_0 + 1/2^{w-v}((3/2)^v - 1)$ when applying first all the $(3x+1)/2$ multiplications while meeting the relationship $w = \lfloor (\ln(3)/\ln(2)).v \rfloor + 1$. The solution to $y_0 = x_0$ is therefore the proposed one because applying any other combinations of the even and odd steps will increase the solution x_0 . □

Lemma 27. *Using a logarithmic scale on the ordinates, the expression $f(v) = \frac{3^v - 2^v}{2^w - 3^v} - 1$ is located around the horizontal axis 1, in some apparently symmetrical way, with diverging points (towards $+\infty$ or towards 0^+) depending largely on the rational approximations of $\frac{\ln(2)}{\ln(3)}$.*

Proof. The figure 10 visualize the lemma for the reader and provides an answer to the "symmetry" around ordinate 1-axis. The said expression diverges if and only if $2^w - 3^v \rightarrow 0$, which is equivalent to $2^w/3^v \rightarrow 1$. Replacing w by its value, we get

$$\begin{aligned} \frac{2^{\lfloor (\ln(3)/\ln(2))v \rfloor + 1}}{3^v} &= 2^{\lfloor (\ln(3)/\ln(2))v \rfloor + 1 - (\ln(3)/\ln(2))v} \frac{2^{(\ln(3)/\ln(2))v}}{3^v} \\ &= 2^{\lfloor (\ln(3)/\ln(2))v \rfloor + 1 - (\ln(3)/\ln(2))v} \\ &\rightarrow 1 \end{aligned}$$

and therefore $\lfloor \frac{\ln(3)}{\ln(2)}v \rfloor + 1 - \frac{\ln(3)}{\ln(2)}v \rightarrow 0$.

These kind of events occurs of course only if $\frac{\ln(3)}{\ln(2)}v$ approaches an integer value, so that $\frac{\ln(3)}{\ln(2)}v \rightarrow n$ equivalent to $\frac{\ln(2)}{\ln(3)} \rightarrow v/n$ for some $n \in \mathbb{N}$ (and $v \in \mathbb{N}$). Let us observe that $\frac{\ln(2)}{\ln(3)} \approx 0.63092975$ and that $\{53/84 \approx 0.63095238, 306/485 \approx 0.630927835, 665/1054 \approx 0.63092979\}$ and therefore the diverging locations' representation on the figure is only approximative with $v = 0 \pmod{53}$ being one of the proposed locations (in green) and $v = 0 \pmod{306}$ a stronger one (in red) and $v = 0 \pmod{665}$ even more so (in yellow) as the fraction narrows the goal in a better way. \square

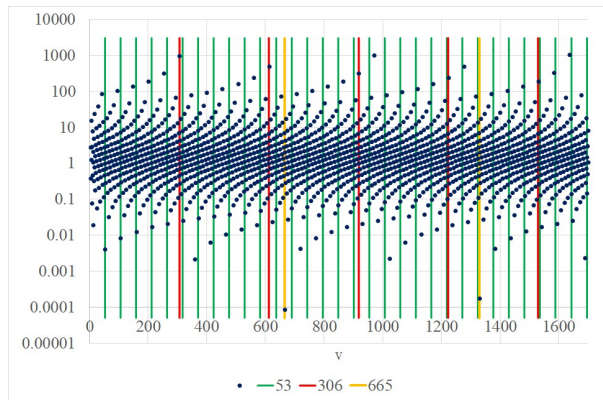


FIGURE 10. Data $\frac{3^v - 2^v}{2^w - 3^v} - 1$.

Note. The previous lemma shows the importance to get the best rational approximations of the the real number $\frac{\ln(2)}{\ln(3)}$. For any continued fraction, the best rational approximations are also called the convergents of the continued fraction [4] and are represented by the Gaussian brackets [5].

Lemma 28. *The coefficients cf_i of the continued fraction of $\frac{\ln(2)}{\ln(3)}$ follow fairly a Gauss-Kuzmin discrete probability distribution. That is*

$$p(cf_i) \rightarrow -\log_2\left(1 - \frac{1}{(cf_i + 1)^2}\right)$$

Proof. The Gauss-Kuzmin discrete probability distribution $p(cf_i)$ arises as the limit probability distribution of the coefficients in the continued fraction expansion of a random variable uniformly distributed in $(0, 1)$ [5]. A numerical verification, with a small size sample (2000 elements), shows that the coefficients of the continued fraction of $\frac{\ln(2)}{\ln(3)}$ follow fairly that distribution. \square

Lemma 29. *The offset Δr of $\frac{\ln(2)}{\ln(3)}$ with its best rational approximations is approximatively*

$$\Delta r = \left(\frac{\ln(2)}{\ln(3)} \frac{1}{v}\right)^2.$$

Proof. Here the continued fraction cf starts with the Gaussian bracket $[0; 1, 1, 1, 2, 2, 3, 1, 5, 2, 23, 2, 2, 1, 1, 55, 1, 4, 3, 1, \dots]$. Table 2 gives the corresponding resulting extracted fractions. Using the theory of Diophantine approximations, the Dirichlet theorem states that there exists for any positive irrational number ir an infinity of couples $(p, q) \in \mathbb{N} * \mathbb{N}^*$ such that $|ir - \frac{p}{q}| < \frac{a}{q^2}$, a being some finite value. This theorem is of course optimum for the best rational approximations. Applying this to $ir = \frac{\ln(2)}{\ln(3)}$, we can thus find two infinite series of integers $\{v_i\}$, $\{wr_i\}$ and an infinite series of real numbers $\{a_i\}$ such that $\frac{\ln(2)}{\ln(3)} - \frac{v_i}{wr_i} = \frac{a_i}{wr_i^2}$. The index i is here a dummy index and the values of v_i are the one corresponding to v for which we get the said best approximations. Besides, the denominator wr_i is either equal to w or $w - 1$ according to the cases where $\frac{\ln(3)}{\ln(2)}v_i$ tends to an integer from beneath or from above. An approximate value $|a_i| \approx 1$ is then obtained by numerical verification. Asymptotically $w_i = \lfloor \frac{\ln(3)}{\ln(2)}v_i \rfloor + 1 \approx \frac{\ln(3)}{\ln(2)}v_i$ and therefore $|\frac{\ln(2)}{\ln(3)} - \frac{v_i}{wr_i}| \approx (\frac{\ln(2)}{\ln(3)} \frac{1}{v_i})^2$. The figures 11, 12, 13 and 14 show the excellent match by making the choice $|a_i| \approx 1$ which is likely the asymptotic exact value. In the 3 last figures, we indicate the values of the coefficients of the continued fraction next to the representative points. We remark that the higher these coefficients the better the "fine tuning" at that step with the exact excepted value. \square

i	cf	fraction	v	w	approx	$x_0(v)$	$x_0(v, w^-)$
0	0	0/1	0	1	0	0	
1	1	1/1	1	2	1	1	-1
2	1	1/2	1	2	0.5	1	
3	1	2/3	2	4	0.66666667	0.71429	-5
4	2	5/8	5	8	0.62500000	16.23077	
5	2	12/19	12	20	0.63157895	1.01974	-73.72361
6	3	41/65	41	65	0.63076923	86.7389	
7	1	53/84	53	85	0.63095238	1.00419	-479.39702
8	5	306/485	306	485	0.63092784	977.7448	
9	2	665/1054	665	1055	0.63092979	1.00009	-22907.85023
10	23	15601/24727	15601	24727	0.63092975	54960.9	
11	2	31867/50508	31867	50509	0.63092975	1.00001	-137648.0025
12	2	79335/125743	79335	125743	0.63092975	272871.59	
13	1	111202/176251	111202	176252	0.63092975	1.00001	-277761.83
14	1	190537/301994	190537	301994	0.63092975	15502072.2	
15	55	10590737/16785921	10590737	16785922	0.63092975	1.00000	-19120269.3
16	1	10781274/17087915	10781274	17087915	0.63092975	81920324.8	
17	4	53715833/85137581	53715833	85137582	0.63092975	1.00000	-287969592.7
18	3	171928773/272500658	171928773	272500658	0.63092975	558903955.	
19	1	225644606/357638239	225644606	357638240	0.63092975	1.00000	-594045517.

TABLE 2

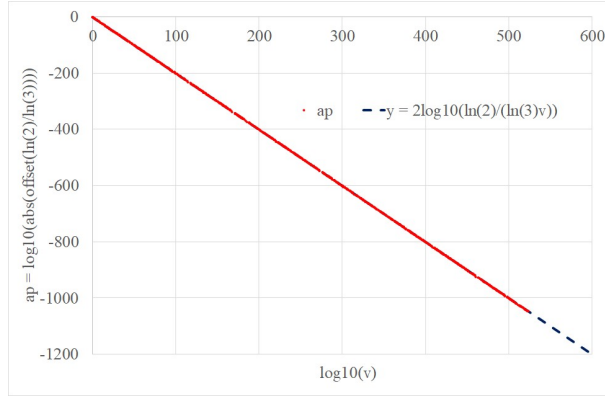


FIGURE 11. Comparison of $ap = \log_{10}(\text{abs}(\frac{\ln(2)}{\ln(3)} - \frac{v}{wr}))$ with $2\log_{10}(\frac{\ln(2)}{\ln(3)} \frac{1}{v})$.

Lemma 30. *The following strictly inequality is true for all $v > 2$*

$$\frac{1}{2} < \frac{2^{w-1} + 2^{v-1}}{3^v} < 1.$$

Proof. If $v = 0$, $\frac{2^{w-1} + 2^{v-1}}{3^v} = \frac{3}{2}$. If $v = 1$, $\frac{2^{w-1} + 2^{v-1}}{3^v} = 1$. If $v = 2$, $\frac{2^{w-1} + 2^{v-1}}{3^v} = \frac{10}{9}$. These cases are excluded.

The equality $2^{w-1} + 2^{v-1} = 3^v$ is obviously false as soon as $v > 1$ for contradiction on parity between the two members of the equation, the first one being even, the second odd. One can then check the inequality for a significant number of values of v . Using the same approach as in the proof of lemma 29, we can write rigorously $\frac{\ln(3)}{\ln(2)} - \frac{wr_i}{v_i} = \frac{c_i}{v_i^2}$, where $|c_i| \approx 1$ as soon

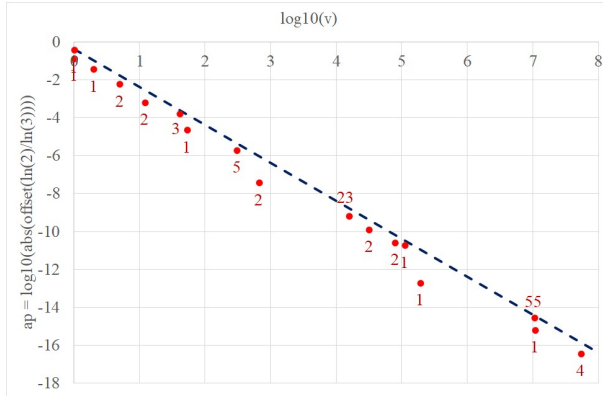


FIGURE 12. Detail relative to figure 11.

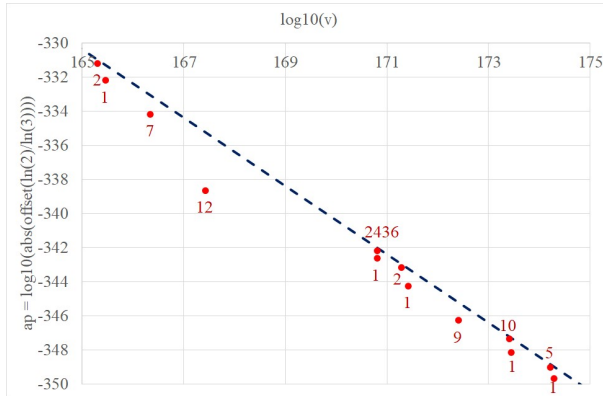


FIGURE 13. Detail relative to figure 11.

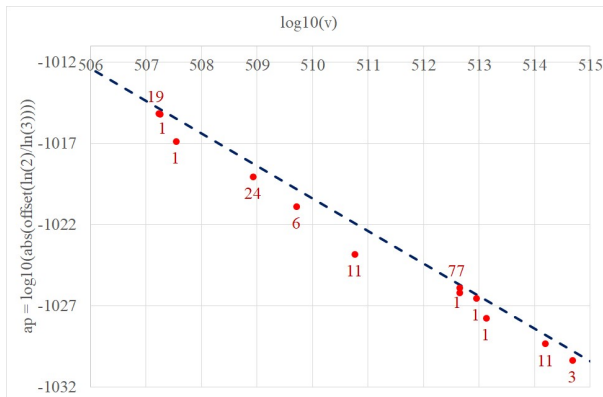


FIGURE 14. Detail relative to figure 11.

as for example $v = 10$. That is equivalent to $\frac{\ln(3)}{\ln(2)}v_i - wr_i = \frac{c_i}{v_i}$ where wr_i is

either equal to w or $w - 1$ according to the cases where $\frac{\ln(3)}{\ln(2)}v_i$ tends to an integer from beneath or from above. Therefore, getting rid of indices, $w = \frac{\ln(3)}{\ln(2)}v + or(0, 1) - \frac{c}{v}$, $|c| \approx 1$. As $w = \lfloor \frac{\ln(3)}{\ln(2)}v \rfloor + 1 = \frac{\ln(3)}{\ln(2)}v + \lfloor \frac{\ln(3)}{\ln(2)}v \rfloor - \frac{\ln(3)}{\ln(2)}v + 1$, we get $or(0, 1) - \frac{c}{v} = \lfloor \frac{\ln(3)}{\ln(2)}v \rfloor - \frac{\ln(3)}{\ln(2)}v + 1$. Therefore if $\frac{\ln(3)}{\ln(2)}v$ tends to an integer from beneath, we get $\frac{c}{v} = \frac{\ln(3)}{\ln(2)}v - \lfloor \frac{\ln(3)}{\ln(2)}v \rfloor - 1$, $c \approx -1$ and if $\frac{\ln(3)}{\ln(2)}v$ tends to an integer from above, we get $\frac{c}{v} = \frac{\ln(3)}{\ln(2)}v - \lfloor \frac{\ln(3)}{\ln(2)}v \rfloor$, $c \approx 1$. This can be summarized, i being some positive integer that $\frac{\ln(3)}{\ln(2)}v$ is the nearest by, with

$$\begin{aligned} \frac{\ln(3)}{\ln(2)}v - \lfloor \frac{\ln(3)}{\ln(2)}v \rfloor &\approx \frac{c}{v} && \text{if } \frac{\ln(3)}{\ln(2)}v \rightarrow i^+ \\ \frac{\ln(3)}{\ln(2)}v - \lfloor \frac{\ln(3)}{\ln(2)}v \rfloor &\approx 1 - \frac{c}{v} && \text{if } \frac{\ln(3)}{\ln(2)}v \rightarrow i^- \end{aligned}$$

where $c \approx 1$.

Now $\frac{2^{w-1}}{3^v} = 2^{\lfloor \frac{\ln(3)}{\ln(2)}v \rfloor} \frac{2^{-\frac{\ln(3)}{\ln(2)}v}}{3^v} = 2^{\lfloor \frac{\ln(3)}{\ln(2)}v \rfloor - \frac{\ln(3)}{\ln(2)}v} = 2^{\lfloor \frac{\ln(3)}{\ln(2)}v \rfloor - \frac{\ln(3)}{\ln(2)}v} = 2^{or(-\frac{c}{v}, -1 + \frac{c}{v})}$.

Then $\frac{2^{w-1} + 2^{v-1}}{3^v} = 2^{or(-\frac{c}{v}, -1 + \frac{c}{v})} + \frac{2^{v-1}}{3^v}$, hence two cases.

If $\frac{\ln(3)}{\ln(2)}v \rightarrow i^+$ provides $2^{-\frac{c}{v}} + \frac{2^{v-1}}{3^v} = e^{-\ln(2)\frac{c}{v}} + \frac{2^{v-1}}{3^v} \approx 1 - \ln(2)\frac{c}{v} + \frac{1}{2}(\frac{2}{3})^v$.

As v is exponentiated in the last term, this one will converges to 0 faster than the term preceding it. Therefore $\frac{2^{w-1} + 2^{v-1}}{3^v} \rightarrow 1^-$.

If $\frac{\ln(3)}{\ln(2)}v \rightarrow i^-$ provides $2^{-1 + \frac{c}{v}} + \frac{2^{v-1}}{3^v} = \frac{1}{2}e^{\ln(2)\frac{c}{v}} + \frac{2^{v-1}}{3^v} \approx \frac{1}{2} + \frac{1}{2}\ln(2)\frac{c}{v} + \frac{2^{v-1}}{3^v} \rightarrow \frac{1}{2}^+$. These are of course limit cases but are the only ones we have to be concerned with as the intermediary cases' values will land between these two.

Let us be even more cautious and go back to the approximate value of c and have a look at the inconsistency of its value if the limit cases were to be met. If we reconsider the above first case, the limit situation would be to write the equality $2^{-\frac{c}{v}} + \frac{2^{v-1}}{3^v} = 1$. When v increases asymptotically, bringing the second term on the left side of the equation near 0, obviously to meet the equality would require to bring the first one up near 1 and therefore c nearer and nearer to 0. More precisely, using $\ln(1+x) \rightarrow x$ for small x , will conduct to $c = -v \frac{\ln(1 - \frac{1}{2}(\frac{2}{3})^v)}{\ln(2)} \rightarrow \frac{v}{2\ln(2)}(\frac{2}{3})^v$. This contradicts already the requirement that c is a constant. Let us however measure the resulting offset. We get $\log_{10}(c) \rightarrow \log_{10}(\frac{v}{2\ln(2)}) + v \log_{10}(\frac{2}{3}) \approx v \log_{10}(\frac{2}{3}) \approx -0.176v$. This is to be compared to $\log_{10}(c) = \log_{10}(1) = 0$ that we used for figure 11. The basis formula $|\frac{\ln(2)}{\ln(3)} - \frac{v}{wr}| \approx (\frac{\ln(2)}{\ln(3)} \frac{1}{v})^2$ for this figure is to be replaced by $|\frac{\ln(2)}{\ln(3)} - \frac{v}{wr}| \approx c \cdot (\frac{\ln(2)}{\ln(3)} \frac{1}{v})^2$ which is equivalent to $\log_{10}|\frac{\ln(2)}{\ln(3)} - \frac{v}{wr}| \approx \log_{10}(c) + \log_{10}((\frac{\ln(2)}{\ln(3)} \frac{1}{v})^2) \approx -0.176v + \log_{10}((\frac{\ln(2)}{\ln(3)} \frac{1}{v})^2)$. This gives an offset as represented in figure 15. The dark blue dashed line is now replaced by the clear blue dashed line. This is totally incompatible with the Gauss-Kuzmin discrete probability distribution's red (almost) line (remembering especially that the representation is in logarithmic scale). Let us not forget here also,

as mentioned previously [1], that any unknown cycle in \mathbb{N}^* would contain at least 17026679261 vertices, that is $v \geq 10742638550$ or $\log_{10}(v) \geq 10.03$. At this stage, to meet the limit case, $c \approx 1$ has to be replaced already by the way off value $c \approx 1.4 \cdot 10^{-1891684748}$ to compensate (which is quite absurd).

We can resume the former argument for the second case. The result is then very close from the previous one, with only a change in sign of c and a slight change in absolute value which does not require further analysis. \square

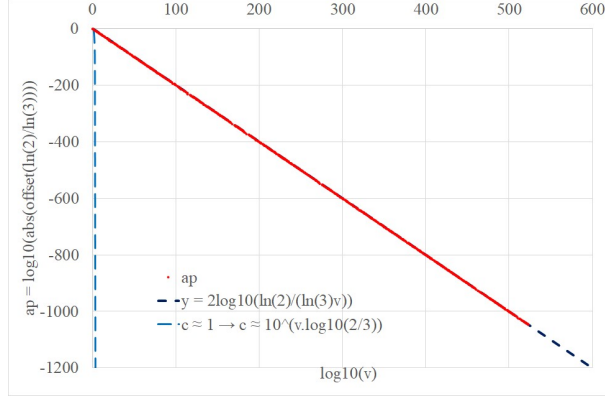


FIGURE 15. Comparison of $ap = \log_{10}\left(\text{abs}\left(\frac{\ln(2)}{\ln(3)} - \frac{v}{wr}\right)\right)$ with $-0.176v + 2\log_{10}\left(\frac{\ln(2)}{\ln(3)} \frac{1}{v}\right)$ (clear blue line).

Theorem 4. *There are no cycle of rank $v > 1$ in \mathbb{N}^* .*

Proof. Let us consider, for some given rank $v > 1$, the two limit cases x_0 and x'_0 obtained within the domain of the licit parity vectors such as provided by lemma 25.

We get

$$x_0 = \frac{3^{v-1}2^0 + 3^{v-2}2^1 + 3^{v-3}2^2 + \dots + 3^12^{v-2} + 3^02^{v-1}}{2^w - 3^v}$$

and

$$x'_0 = \frac{3^{v-1}2^0 + 3^{v-2}2^1 + 3^{v-3}2^3 + 3^{v-4}2^4 + 3^{v-5}2^6 + 3^{v-6}2^7 + \dots + 3^02^{w-1-\text{or}(1,2)}}{2^w - 3^v},$$

the term $\text{or}(1, 2)$ depending on the limit parity vector at rank v ending with 1 or 2 zeroes. We intend to study the greatest common divisor between numerator (num) and denominator (den) of these expressions. In order to make the understanding easier let us start with an example. Let us choose $v = 4$, $w = 7$ and one of the corresponding licit parity vector 1110100. Therefore we get $\text{num} = 3^32^0 + 2^13^2 + 2^23^1 + 2^43^0$ and $\text{den} = 2^w - 3^v$. Then

$$\begin{aligned}
gcd(den, num) &= gcd(2^7 - 3^4, 2^0 3^3 + 2^1 3^2 + 2^2 3^1 + 2^4 3^0) & (1) \\
&= gcd(2^7 - 3^4, 3(3^3 2^0 + 2^1 3^2 + 2^2 3^1 + 2^4 3^0)) \\
&= gcd(2^7 - 3^4, 3^4 2^0 + 2^1 3^3 + 2^2 3^2 + 2^4 3^1) \\
&= gcd(2^7 - 3^4, 3^4 2^0 + 2^1 3^3 + 2^2 3^2 + 2^4 3^1 + 2^7 - 3^4) \\
&= gcd(2^7 - 3^4, 2^1 3^3 + 2^2 3^2 + 2^4 3^1 + 2^7 3^0) \\
&= gcd(2^7 - 3^4, 3^3 + 2^1 3^2 + 2^3 3^1 + 2^6 3^0) & (2) \\
&= gcd(2^7 - 3^4, 3^3 + 2^2 3^2 + 2^5 3^1 + 2^6 3^0) & (3) \\
&= gcd(2^7 - 3^4, 3^3 + 2^3 3^2 + 2^4 3^1 + 2^5 3^0) & (4) \\
&= gcd(2^7 - 3^4, 3^3 + 2^1 3^2 + 2^2 3^1 + 2^4 3^0) & (1)
\end{aligned}$$

We give the detail from step (1) to step (2), the other ones being entirely similar. In each main steps (1) to (4), the exponents of 3 are unchanged and decrease from $v - 1 = 3$ to 0. The only "challenge" is to handle the exponents of 2. One has to proceed as follows. Start with $(0, 1, 2, 4)$ which is the initial list of the exponents of 2 and add $w = 7$ at the end of the list $(0, 1, 2, 4; 7)$. Shift the list by one to the left and subtract the value of the first item to each number and add $w = 7$ at the end of the list again $(1-1, 2-1, 4-1, 7-1; 7) \equiv (0, 1, 3, 6; 7)$. Continue $(1-1, 3-1, 6-1, 7-1; 7) \equiv (0, 2, 5, 6; 7)$. Again $(2-2, 5-2, 6-2, 7-2; 7) \equiv (0, 3, 4, 5; 7)$ until going back to the initial expression $(3-3, 4-3, 5-3, 7-3; 7) \equiv (0, 1, 2, 4; 7)$. Here, during the process, the total number of subtractions is $v = 4$ and add to $w = 1 + 1 + 2 + 3 = 7$. Obviously it is a general pattern. Now, for any value v , if t_0 is some smallest elements of a cycle in \mathbb{N} , we will have $den = 2^w - 3^v$ and $num = 3^{v-1}2^0 + 3^{v-2}2^1 + \dots + 3^0 2^{w-1-\text{or}(1,2)} = 2^w - 3^v$. Let us have $(0, i_1, i_2, i_3, \dots, i_{v-1}; w)$ the initial corresponding list of exponents. Then we inherit of a numerators' list containing v lines which each must correspond to exact multiples of $2^w - 3^v$

$$\begin{aligned}
&(0, i_1, i_2, i_3, \dots, i_{v-2}, i_{v-1}; w) \\
&(0, i_2 - i_1, i_3 - i_1, i_4 - i_1, \dots, i_{v-1} - i_1, w - i_1; w) \\
&(0, i_3 - i_2, i_4 - i_2, i_5 - i_2, \dots, w - i_2, w + i_1 - i_2; w)
\end{aligned}$$

...

$$(0, w - i_{v-1}, w + i_1 - i_{v-1}, w + i_2 - i_{v-1}, \dots, w + i_{v-3} - i_{v-1}, w + i_{v-2} - i_{v-1}; w)$$

before getting back to the initial

$$(0, i_1, i_2, i_3, \dots, i_{v-2}, i_{v-1}; w).$$

The case that each of the corresponding gcd be equal to $2^w - 3^v$ bends largely to absurdity. But let us continue anyway.

We can step by step use these v expressions to reduce by (linear) subtractions and multiplications the powers of 3. That construction leads, in a systematic way, to expressions between commas containing each 4 members (if $v > 3$) as 4 terms disappear at each step being equal by pairs (and yes the 2 last ones $2^{w-i_4+i_1} - 2^{w-i_4+i_2}$ underneath emerge always at the end of the process whatever the value of v)

$$3 + 2^{i_{v-1}-1} - 2^{i_{v-2}-1} + 2^{w-i_4+i_1} - 2^{w-i_4+i_2}$$

and therefore to the following greatest common divisor to be studied

$$\gcd(2^w - 3^v, 3 + 2^{i_{v-2}-1}(2^{i_{v-1}-i_{v-2}} - 1) - 2^{w-i_4+i_1}(2^{i_2-i_1} - 1)).$$

Although interesting by its simplicity, the second component of this \gcd is not systematically smaller than $2^w - 3^v$, an event which would constituted a lucky first condition to prove our aim. Therefore, let us focus instead only on the two first items of the previous list:

$$(0, i_1, i_2, i_3, \dots, i_{v-1}; w) \\ (0, i_2 - i_1, i_3 - i_1, i_4 - i_1, \dots, w - i_1; w).$$

As we mentioned, in order to get a Collatz' cycle, the first item corresponds to a value that equals exactly $2^w - 3^v$ while the second one must then correspond to an exact non-zero multiple of the same value. So let us have the two corresponding terms:

$$3^{v-1}2^0 + 3^{v-2}2^{i_1} + 3^{v-3}2^{i_2} + 3^{v-4}2^{i_3} + \dots + 3^1 2^{i_{v-2}} + 3^0 2^{i_{v-1}} \\ 3^{v-1}2^0 + 3^{v-2}2^{i_2-i_1} + 3^{v-3}2^{i_3-i_1} + 3^{v-4}2^{i_4-i_1} + \dots 3^1 2^{i_{v-1}-i_1} + 3^0 2^{w-i_1}$$

The \gcd linked ratio r of the second term to the first one is

$$r = \frac{3^{v-1}2^0 + 3^{v-2}2^{i_2-i_1} + 3^{v-3}2^{i_3-i_1} + 3^{v-4}2^{i_4-i_1} + \dots 3^1 2^{i_{v-1}-i_1} + 3^0 2^{w-i_1}}{3^{v-1}2^0 + 3^{v-2}2^{i_1} + 3^{v-3}2^{i_2} + 3^{v-4}2^{i_3} + \dots + 3^1 2^{i_{v-2}} + 3^0 2^{i_{v-1}}} \\ = \frac{1 + \frac{2^{i_2-i_1}}{3^1} + \frac{2^{i_3-i_1}}{3^2} + \dots + \frac{2^{i_{v-1}-i_1}}{3^{v-2}} + \frac{2^{w-i_1}}{3^{v-1}}}{1 + \frac{2^{i_1}}{3^1} + \frac{2^{i_2}}{3^2} + \frac{2^{i_3}}{3^3} + \dots + \frac{2^{i_{v-2}}}{3^{v-2}} + \frac{2^{i_{v-1}}}{3^{v-1}}}$$

Now the exponents are issued from a parity vector and therefore we have systematically $i_1 = 1$. So that

$$r = \frac{1 + \frac{2^{i_2-1}}{3^1} + \frac{2^{i_3-1}}{3^2} + \dots + \frac{2^{i_{v-1}-1}}{3^{v-2}} + \frac{2^{w-1}}{3^{v-1}}}{1 + \frac{2^{i_1}}{3^1} + \frac{2^{i_2}}{3^2} + \frac{2^{i_3}}{3^3} + \dots + \frac{2^{i_{v-2}}}{3^{v-2}} + \frac{2^{i_{v-1}}}{3^{v-1}}} \\ = \frac{1 + \frac{2^{i_1-1}}{3^0} + \frac{2^{i_2-1}}{3^1} + \frac{2^{i_3-1}}{3^2} + \dots + \frac{2^{i_{v-1}-1}}{3^{v-2}} + (\frac{2^{w-1}}{3^{v-1}} - 1)}{1 + \frac{2^{i_1}}{3^1} + \frac{2^{i_2}}{3^2} + \frac{2^{i_3}}{3^3} + \dots + \frac{2^{i_{v-2}}}{3^{v-2}} + \frac{2^{i_{v-1}}}{3^{v-1}}}.$$

Let us have

$$c_0 = \frac{2^{i_1}}{3^1} + \frac{2^{i_2}}{3^2} + \frac{2^{i_3}}{3^3} + \dots + \frac{2^{i_{v-2}}}{3^{v-2}} + \frac{2^{i_{v-1}}}{3^{v-1}} \quad \text{and} \quad g = \frac{2^{w-1}}{3^{v-1}} - 1.$$

Then

$$r = \frac{1+3c_0/2+g}{1+c_0}.$$

Therefore the conditions underneath, if one of them is true, are equivalent

$$3/2 < r < 2 \Leftrightarrow \{g > 1/2 \text{ and } c_0 > 2(g-1)\} \\ \Leftrightarrow \{\frac{2^w}{3^v} > 1 \text{ and } c_0 > 3(\frac{2^w}{3^v} - \frac{4}{3})\}.$$

Let us observe that starting with a lower and upper bound on r , we end with only a lower bound on c_0 . This is a main simplifying event as we will see very soon. Meanwhile, let us examine the two limit cases.

For $i_1 = 1, i_2 = 2, i_3 = 3, \dots, i_{v-1} = v - 1$, we get

$$\begin{aligned}
 c_0 &= 1 + \frac{2^1}{3^1} + \frac{2^2}{3^2} + \frac{2^3}{3^3} + \dots + \frac{2^{v-2}}{3^{v-2}} + \frac{2^{v-1}}{3^{v-1}} - 1 \\
 &= \frac{1 - \frac{2^v}{3^v}}{1 - \frac{2}{3}} - 1 \\
 &= 3\left(\frac{2}{3} - \frac{2^v}{3^v}\right)
 \end{aligned}$$

Now

$$\begin{aligned}
 w = \lfloor \frac{\ln(3)}{\ln(2)} v \rfloor + 1 &\Rightarrow \frac{2^w}{3^v} > 1 \\
 \frac{2^{w-1} + 2^{v-1}}{3^v} < 1 &\Rightarrow 3\left(\frac{2}{3} - \frac{2^v}{3^v}\right) > 3\left(\frac{2^w}{3^v} - \frac{4}{3}\right)
 \end{aligned}$$

The first implication results from lemma 24. The second implication, an equivalence in fact, results from lemma 30. Therefore the term r , lying strictly between $\frac{3}{2}$ and 2, cannot be an integer which proves the first limit case. Now for the second case (and all intermediary ones) the term

$$c_1 = \frac{2^{i_1}}{3^1} + \frac{2^{i_2}}{3^2} + \frac{2^{i_3}}{3^3} + \dots + \frac{2^{i_{v-2}}}{3^{v-2}} + \frac{2^{i_{v-1}}}{3^{v-1}}$$

is necessary superior, as $i_k \geq k$ for all $k \in 1, 2, \dots, v-1$, to the term

$$c_0 = \frac{2^1}{3^1} + \frac{2^2}{3^2} + \frac{2^3}{3^3} + \dots + \frac{2^{v-2}}{3^{v-2}} + \frac{2^{v-1}}{3^{v-1}}$$

Hence $c_1 > c_0 > 3\left(\frac{2^w}{3^v} - \frac{4}{3}\right)$ proving the second case and all intermediary cases which corresponding values c_i will be strictly between c_0 and c_1 . \square

Figure 16 shows a sample of the value of $r - 1$. It shows the constraint $r \rightarrow \frac{3}{2}^+$ in the case of the second limit case which is a consequence of $\frac{1}{2} < \frac{2^{w-1} + 2^{v-1}}{3^v}$ (see lemma 30). Note also the common value $r - 1 = \frac{6}{5}$ for $v = 2$ in the first and second limit cases.

Appendix B provides a computer program to evaluate the ratio $r - 1$ for some initial choice of x_0 that the reader may implement.

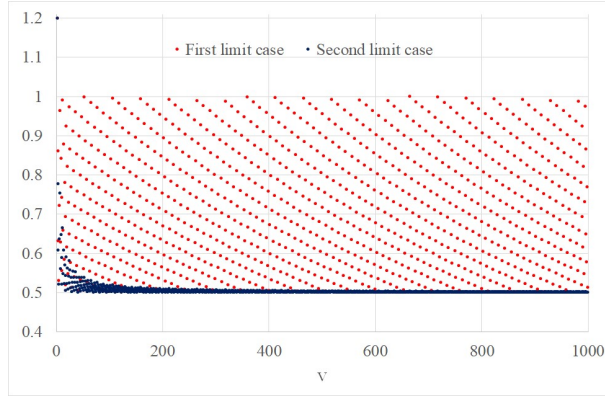


FIGURE 16. Ratio $r - 1$.

Theorem 5. *The Collatz conjecture is true.*

Proof. This is an immediate consequence of the two theorems 1 and 4. \square

The reader may refer to [6] for a longer version of this article with additional arguments confirming the Collatz conjecture.

Mainly, we show that any Collatz tree grows at the same asymptotic pace, a result rather obvious when admitting that the reverse Collatz algorithm is asymptotically a random process. A second "equal" sized tree should have then be easily detected in \mathbb{N}^* .

Exploiting the 2^w periodicity of the integers with equal stopping time and studying the smallest ones, that we call 2^w -seeds, the precise evaluation of the number of the said seeds, considering a random distribution, allows us to show that their abscissas in \mathbb{N}^* grows exponentially while the corresponding abscissa of the smallest element of a cycle grows polynomially which is again asymptotically incompatible with the Collatz conjecture to be false.

APPENDIX A. CONTINUED FRACTION OF $\text{LN}(3)/\text{LN}(2)$

The following program provides the continued fraction of $\frac{\ln(2)}{\ln(3)}$ and the corresponding successive fractions' approximations. For $\frac{\ln(3)}{\ln(2)}$, the continued fraction starts with [1; 1, 1, 2, ...] instead of [0; 1, 1, 1, 2, ...].

PARI/GP programming code.

```
\p 5000
{nb = 20;
vct = vector(nb); x = log(2)/log(3);
for(i = 1, nb, vct[i] = x\1;
y = 1/(x-vct[i]); x = y);
print(vct);
for(i = 2, nb, t = i; x = vct[t];
for(j = 2, i, k = t-j+1; x = vct[k]+1/x);
print(numerator(x)" "denominator(x)))}
```

The following program provides an alternative way to get the continued fraction of $\frac{\ln(2)}{\ln(3)}$ and the corresponding successive fractions' approximations. It may be analogous to the Terence Jackson / Keith Matthews algorithm in the article "On Shanks' algorithm for computing the continued fraction of $\log_b(a)$ " (see reference [2]).

PARI/GP programming code.

```
{infty = 10000000000;
e1 = 1; e2 = 0;
v1 = 0; w1 = 1; v2 = 1; w2 = 1;
printtex(cf" "num" "den);
print(0" "v1" "w1);
print(1" "v2" "w2);
for(n = 1, 50, v3 = v1+v2;
for(cf = 1, infty, m = cf;
v3 = v3+v2;
w3 = (log(3)/log(2)*v3)\1+1-e2;
r1 = (w3-w1)/(v3-v1);
r2 = w2/v2;
if(r1 == r2, ,v3 = v3-v2; break));
w3 = (log(3)/log(2)*v3)\1+1-e2;
print(m" "v3" "w3);
e1 = 1-e1; e2 = 1-e2;
v1 = v2; v2 = v3;
w1 = (log(3)/log(2)*v1)\1+e1;
w2 = (log(3)/log(2)*v2)\1+e2)}
```

APPENDIX B. EVALUATION OF THE *gcd* LINKED RATIO

The following program provides the evaluation of the ratio $r - 1$ for some initial values x_0 equal to $3 \pmod 4$. The ratio $n^{\circ}1$ is always in the interval $]0, 1[$ except here if $v = 2$ for the said initial values.

Note that sometimes the exponentiation sign \wedge won't copy successfully and has to be retyped manually (4 corrections to implement in that case).

PARI/GP programming code.

```
{infty = 10000;
for(i = 1, 10, \\ make choice
x0 = 4*i+3;
parv = addp = shft = vector(1);
t = (3*x0+1)/2 ; parv[1] = 1; shft[1] = 0; v = 1;
for(k = 1, infty, if(t < x0, break,
if(t/2 == t\2, t = t/2;
addp[1] = 0, v = v+1; t = (3*t+1)/2; addp[1] = 1);
parv = concat(parv, addp[1]));
print("Initial value = "x0", v = "v);
print("Parity vector "parv);
w = (log(3)/log(2)*v)\1+1; addp[1] = 0; m = 1;
for(k = 1, v, n = m;
for(j = 1, w-n, if(parv[n+j] == 0, m = m+1, break)); m = m+1; addp[1] = m-1;
shft = concat(shft, addp[1])); shftp = shftn = shft; tot0 = 0.0;
for(k = 1, v, tot0 = tot0+(3 ^ (v-k))*(2 ^ shftn[k]));
for(j = 1, v-1,
for(k = 3, v+1, shftp[k-1] = shftn[k]-shftn[2]);
print("Exponents vector "shftp); tot1 = 0.0;
for(k = 1, v, tot1 = tot1+(3 ^ (v-k))*(2 ^ shftp[k]));
r = tot1/tot0; print("Ratio n^{\circ}j": "r-1); shftn = shftp))}
```

LITERATURE AND SOURCES

- [1] Shalom Eliahou. 2011-2013. <http://images.math.cnrs.fr/Le-probleme-3n-1-y-a-t-il-des-cycles-non-triviaux-III>
- [2] Terence Jackson, Keith Matthews. Journal of Integer Sequences, Vol. 5 (2002), Article 02.2.7. On Shanks' algorithm for computing the continued fraction of $\log_b(a)$.
- [3] Riho Terras. A stopping time problem on the positive integers. Acta Arithmetica 30 (1976), 241–252
- [4] Wikipedia. Continued fraction. https://en.wikipedia.org/wiki/Continued_fraction
- [5] Wikipedia. Gaussian brackets. https://en.wikipedia.org/wiki/Gaussian_brackets
- [6] Hubert Schaezel. <https://hubertschaetzel.wixsite.com/website> Terras sheet.

INPG GRENOBLE

Email address: `hubert.schaetzel@wanadoo.fr`

URL: <https://hubertschaetzel.wixsite.com/website>