Collatz Conjecture: The Race To One

Jochen Kiemes

2025-03-25

Abstract

This paper presents an approach that reinterprets the Collatz sequence by transforming it into a new sequence, highlighting the dynamic relationship between the trailing and leading bits of its elements. This mapping enables the study of a "bit-race," whose well-defined statistical properties rigorously guarantee the convergence of all Collatz sequences to 1.

1 Statement of the Collatz Conjecture

The **Collatz Conjecture** is a famous unsolved problem in mathematics. It states that for any positive integer $n \in \mathbb{N}$, the sequence defined by:

$$c_0 = n,$$

and recursively for all $i \geq 0$,

$$c_{i+1} = \begin{cases} \frac{c_i}{2} & \text{if } c_i \equiv 0 \pmod 2, \\ 3c_i + 1 & \text{if } c_i \equiv 1 \pmod 2, \end{cases} \tag{1}$$

will eventually reach the number 1. Once $c_k = 1$, subsequent terms enters the cycle $4, 2, 1$.

The conjecture is often phrased as:

"No matter what value of n you start with, you will always eventually reach 1."

We call the application of the transition from c_i to c_{i+1} a Collatz-step and the sequence $\{c_i\}$ the Collatz-sequence.

A thorough overview about the current status of mathematical investigations is presented in [Lagarias, 2021].

2 Key Observations and Challenges

- The conjecture has been verified computationally for all starting values up to 2^{68} [Barina, 2020], but no general proof exists.
- Despite its simple formulation, the problem exhibits chaotic behavior due to the interplay of additive $(3c_i + 1)$ and multiplicative $(c_i/2)$ operations.
- It belongs to the broader field of dynamical systems and is linked to problems in number theory like Syracuse sequences.

3 Proof Strategy

Our approach to proving the Collatz Conjecture involves a multi-step methodology, which can be succinctly outlined as follows:

- We commence by establishing the requisite mathematical foundations, including the introduction of relevant functions, definitions, and notation.

- Subsequently, we map the Collatz Conjecture to distinct number classes.
- A reformulation of the Collatz Conjecture is then presented, which facilitates a more manageable mathematical treatment and does not exhibit the chaotic behavior.
- We then identify the underlying race of leading and trailing bit in the reformulated Collatz-sequence.
- Furthermore, we are given the basis for proof by contradiction.
- Given the inherent race, we are determining the upper limit for the speed of the leading bit.
- Then we are estimating the speed of the trailing bit based on stochastic assumption.
- Finally we come up for a proof, that the stochastic assumption is valid.
- With the speed of leading and trailing bits, we can demonstrate the non-existence of trajectories not leading to 1

4 Mathematical Foundation for Analysis

4.1 Notation

Let $\{c_i\}$ denote the sequence starting with $c_0 \in \mathbb{N}$ and repeated application of the Collatz steps.

Let $\mathbb{O}$ denote the set of all odd numbers in $\mathbb{N}$, i.e., $n \in \mathbb{N}$ such that $n \bmod 2 \equiv 1$.

To facilitate our analysis, it is helpful to consider the binary representation of a number $n \in \mathbb{N}$ as $n = \sum_{i=0}^{k} b_i 2^i = b_k b_{k-1} \cdots b_0$, where $b_i \in \{0, 1\}$ for each $0 \leq i < k$ and $b_k = 1$.

4.2 Bit value: bit (n, i)

Based on the binary representation of n, the function bit (n, i) simply derives the value of b_i. Mathematical definition:

$$\text{bit}\,(n, i) = \lfloor n/2^i \rfloor \bmod 2 \tag{2}$$

4.3 Trailing bit position: $t(n)$

We define the function $t(n)$ for each $n \in \mathbb{N}$ to be the lowest index $\min\{i\}$ such that $b_i = 1$. This function is referred to as the trailing bit function. Mathematically, $t(n)$ can be defined as:

$$n \bmod 2^{t(n)} \equiv 0 \text{ and } \frac{n}{2^{t(n)}} \bmod 2 \equiv 1 \tag{3}$$

4.4 Mantissa: $m(n)$

Next, we define the function $m(n)$ for each $n \in \mathbb{N}$ to be the mantissa of n. The mathematical definition of $m(n)$ is:

$$m(n) = \frac{n}{2^{t(n)}} \tag{4}$$

It is important to note that the result of $m(n)$ always yields an element of $\mathbb{O}$.

Furthermore, we have the following properties:

- Invariance against multiplication of any arbitrary exponent of 2:

$$m(n) \equiv m(n \cdot 2^i) \text{ for any } i \in \mathbb{N} \tag{5}$$

- $m(n)$ is a homogeneous function in any degree of k, but only for elements $o \in \mathbb{O}$:

$$o^k \cdot m(n) \equiv m(o^k \cdot n) \text{ for any } o \in \mathbb{O} \text{ and } k \in \mathbb{N}_0 \tag{6}$$

- $m(n)$ is an idempotent function:

$$m(m(n)) \equiv m(n) \tag{7}$$

4.5 Leading bit position: $l(n)$

We define a function $l(n)$ that maps each natural number $n \in \mathbb{N}$ to its leading bit. The mathematical representation of this function is given by:

$$l(n) = \lfloor \log_2 n \rfloor \tag{8}$$

It can be verified that this definition implies the following inequality:

$$2^{l(n)} \leq n < 2^{l(n)+1} \tag{9}$$

4.6 Bit distance: $d(n)$

Finally, we define a distance function $d(n)$ as:

$$d(n) = l(n) - t(n) \tag{10}$$

It is worth noting:

- $d(2^i) \equiv 0$ for any $i \in \mathbb{N}_0$
- $d(n) = d(m(n))$ for any $n \in \mathbb{N}$.

4.7 Relation inbetween bit functions

Using these functions, any number n can be expressed as:

$$n = m(n) \cdot 2^{t(n)} \tag{11}$$

For the relation between mantissa and distance:

$$m(n) = \frac{n}{2^{t(n)}} \tag{12}$$

$$l(m(n)) = l(\frac{n}{2^{t(n)}}) = l(n) - t(n) = d(n) \tag{13}$$

With the inequality 9, this yields:

$$2^{d(n)} \leq m(n) < 2^{d(n)+1} \tag{14}$$

4.8 Example for $m(n), l(n), t(n), d(n)$

To illustrate the usage of these functions, consider the following example:

Bit	6	5	4	3	2	1	0
$n = 42 = 101010_b$ as binary=	0	1	0	1	0	1	0
$m(n) = 21 = 10101_b$		1	0	1	0	1	
$l(n) = 5$		5					
$t(n) = 1$						1	
$d(n) = 4 =$		5		-		1	

Table 1: Illustration for $m(n)$, $l(n)$, $t(n)$ and $d(n)$

4.9 Number classes

Beyond our preceding definitions, we introduce the notion of number classes. For each element $o \in \mathbb{O}$, we define its associated number class as the infinite set:

$$\{o \cdot 2^i : i \in \mathbb{N}_0\} \tag{15}$$

This definition establishes a bijective mapping between elements of $\mathbb{O}$ and number classes, where each class comprises an infinite sequence of powers of 2 scaled by the factor o.

Every natural number $n \in \mathbb{N}$ is a member of precisely one number class, which is uniquely identified by its corresponding element $o \in \mathbb{O}$. The function $m(n)$ assigns to each natural number n its respective number class o, thereby facilitating the classification of all natural numbers into distinct classes based on their association with elements of $\mathbb{O}$.

More formally, we can derive the number class of n directly by $o = m(n)$ for any $n \in \mathbb{N}$.

4.10 Powers of 3

Lemma 1. *The sequence of all exponents of 3 denoted as $\{e_i = 3^i : i \in \mathbb{N}\}$ will start with:*

$$e_0 = 1,$$

and recursively for all $i \geq 0$,

$$e_{i+1} = 3e_i \tag{16}$$

will have these properties:

- $e_i \equiv 001_b \bmod 2^3$ *for even i,*
- $e_i \equiv 011_b \bmod 2^3$ *for odd i,*

Proof. Analyzing the individual cases:

- The starting element $e_0 = 1$ is 001_b with $i = 0$ being even. All further even indizes can be calculated as:

  $$e_{i+2} = 3 * e_{i+1} = 3 * 3 * e_i = 9e_i = 2^3 e_i + e_i$$

  Consequently, $e_{i+2} \equiv (2^3 e_i + e_i) \bmod 2^3 \equiv e_i \bmod 2^3$, thus e_{i+2} is 001_b for i even.

- The element $e_1 = 3e_0 = 3$ is 011_b with $i = 1$ being odd. All further odd indizes can be similarly calculated as:

  $$e_{i+2} = 3 * e_{i+1} = 3 * 3 * e_i = 9e_i = 2^3 e_i + e_i$$

  Consequently, $e_{i+2} \equiv (2^3 e_i + e_i) \bmod 2^3 \equiv e_i \bmod 2^3$, thus e_{i+2} is 011_b for i odd.

□

This gives reason for a different definition of $\{e_i\}$:

$$e_i = \begin{cases} 1 + 8 * \lfloor 3^i/8 \rfloor & \text{for } i \equiv 0 \mod 2 \\ 3 + 8 * \lfloor 3^i/8 \rfloor & \text{for } i \equiv 1 \mod 2 \end{cases} \tag{17}$$

Introducing $h_i = 8 * \lfloor 3^i/8 \rfloor$:

$$e_i = \begin{cases} 1 + 8 * h_i \text{ for } i \equiv 0 \mod 2 \\ 3 + 8 * h_i \text{ for } i \equiv 1 \mod 2 \end{cases} \tag{18}$$

Lemma 2. *Consider the sequence defined by $h_i = \left\lfloor \frac{3^i}{8} \right\rfloor \mod 2^k$ for $i = 0, 1, 2, \ldots$, where $k \geq 1$ is a positive integer. We prove that in every $2 \cdot 2^k$ consecutive terms, all integers n such that $0 \leq n < 2^k$ appear exactly twice.*

Proof. Define $r_i = 3^i \mod 2^{k+3}$. Since $8 = 2^3$ and $2^{k+3} = 8 \cdot 2^k$, write $3^i = q \cdot 2^{k+3} + r_i$, where $q \geq 0$ and $0 \leq r_i < 2^{k+3}$. Then,

$$\left\lfloor \frac{3^i}{8} \right\rfloor = \left\lfloor \frac{q \cdot 2^{k+3} + r_i}{8} \right\rfloor = \left\lfloor q \cdot 2^k + \frac{r_i}{8} \right\rfloor = q \cdot 2^k + \left\lfloor \frac{r_i}{8} \right\rfloor,$$

since $\frac{r_i}{8} < 2^k$. Thus,

$$h_i = \left\lfloor \frac{3^i}{8} \right\rfloor \mod 2^k = (q \cdot 2^k + \left\lfloor \frac{r_i}{8} \right\rfloor) \mod 2^k = \left\lfloor \frac{r_i}{8} \right\rfloor,$$

because $q \cdot 2^k \equiv 0 \mod 2^k$ and $0 \le \lfloor \frac{r_i}{8} \rfloor < 2^k$.

We now examine the periodicity of the sequence $h_i = \left\lfloor \frac{3^i}{8} \right\rfloor \mod 2^k$, which depends on $r_i = 3^i \mod 2^{k+3}$. To understand this, we need the concept of the *order* of 3 modulo 2^{k+3}, which is the smallest positive integer m such that $3^m \equiv 1 \mod 2^{k+3}$. This means $3^m - 1$ is divisible by 2^{k+3}, or equivalently, raising 3 to the power m brings us back to 1 when divided by 2^{k+3} and taking the remainder.

For $k \ge 1$ (so $k = 1, 2, 3, \ldots$), the exponent $k + 3$ is at least 4, since $1 + 3 = 4$, and increases with k (e.g., $k = 2$ gives $2 + 3 = 5$). Thus, we are working with moduli like $2^4 = 16$, $2^5 = 32$, and so on. The order of 3 modulo 2^{k+3} turns out to be 2^{k+1}. Let's verify this with examples:

- If $k = 1$, then $2^{k+3} = 2^4 = 16$ and $2^{k+1} = 2^{1+1} = 2^2 = 4$. Compute powers: $3^1 = 3$, $3^2 = 9$, $3^3 = 27 \equiv 11 \mod 16$ (since $27 - 16 = 11$), $3^4 = 81 \equiv 1 \mod 16$ (since $81 - 5 \cdot 16 = 81 - 80 = 1$). The smallest m where $3^m \equiv 1 \mod 16$ is 4, matching 2^{k+1}.
- If $k = 2$, then $2^{k+3} = 2^5 = 32$ and $2^{k+1} = 2^{2+1} = 2^3 = 8$. Check: $3^4 = 81 \equiv 17 \mod 32$ (since $81 - 2 \cdot 32 = 81 - 64 = 17$), $3^8 = 6561 \equiv 1 \mod 32$ (since $6561 - 205 \cdot 32 = 6561 - 6560 = 1$). The smallest m is 8, again matching 2^{k+1}.

This pattern holds: the order of 3 modulo 2^{k+3} is 2^{k+1}, meaning $3^{2^{k+1}} \equiv 1 \mod 2^{k+3}$, and no smaller positive exponent works.

Now, since $r_i = 3^i \mod 2^{k+3}$, it repeats when the exponent increases by the order: $3^{i+2^{k+1}} = 3^i \cdot 3^{2^{k+1}} \equiv 3^i \cdot 1 \equiv 3^i \mod 2^{k+3}$. Thus, $r_{i+2^{k+1}} = r_i$, and the period of r_i is exactly 2^{k+1} (e.g., 4 for $k = 1$, 8 for $k = 2$), as it's the smallest m making $3^m \equiv 1$.

Finally, $h_i = \left\lfloor \frac{3^i}{8} \right\rfloor \mod 2^k$ depends on 3^i, and since $3^i = q \cdot 2^{k+3} + r_i$, we have $\left\lfloor \frac{3^i}{8} \right\rfloor = q \cdot 2^k + \lfloor \frac{r_i}{8} \rfloor$, so $h_i = \lfloor \frac{r_i}{8} \rfloor$ modulo 2^k. When i increases by 2^{k+1}, $r_{i+2^{k+1}} = r_i$, so $h_{i+2^{k+1}} = h_i$. Hence, h_i also has period 2^{k+1}, which is $2 \cdot 2^k$ (e.g., $2 \cdot 2^1 = 4$, $2 \cdot 2^2 = 8$).

For $i = 0, 1, \ldots, 2^{k+1} - 1$, the values $r_i = 3^i \mod 2^{k+3}$ are distinct, forming a cyclic subgroup of $(\mathbb{Z}/2^{k+3}\mathbb{Z})^*$ of order 2^{k+1}, with $\phi(2^{k+3}) = 2^{k+2}$. Since 3 is odd, r_i is odd. Compute $r_i \mod 8$: if i is even, $3^i \equiv 1 \mod 8$; if i is odd, $3^i \equiv 3 \mod 8$. Among 2^{k+1} indices, there are 2^k even and 2^k odd i, so 2^k values of $r_i \equiv 1 \mod 8$ and 2^k values $r_i \equiv 3 \mod 8$.

For each n where $0 \le n < 2^k$, let $r = 8n + 1$ and $r = 8n + 3$, both less than 2^{k+3} since $8(2^k - 1) + 3 = 2^{k+3} - 5 < 2^{k+3}$. The subgroup includes all such r, with 2^k residues $\equiv 1 \mod 8$ and $2^k \equiv 3 \mod 8$. If $r_i = 8n + 1$, then $h_i = \lfloor \frac{8n+1}{8} \rfloor = n$; if $r_i = 8n + 3$, then $h_i = \lfloor \frac{8n+3}{8} \rfloor = n$. Each $8n + 1$ and $8n + 3$ appears once in the 2^{k+1} terms, so each n appears exactly twice.

Thus, in every $2 \cdot 2^k$ terms, each n from 0 to $2^k - 1$ appears exactly twice, ensuring all such n appear at least once. ☐

Lemma 3. *Consider the sequence defined by* $h'_i = \left\lfloor \frac{o3^i}{8} \right\rfloor \mod 2^k$ *for* $i = 0, 1, 2, \ldots,$ *where* $k \ge 1$ *is a positive integer and* $o \in \mathbb{O}$. *We prove that in every* $2 \cdot 2^k$ *consecutive terms, all integers* n *such that* $0 \le n < 2^k$ *appear exactly twice.*

Proof. The periodicity of $\left\lfloor \frac{3^i}{8} \right\rfloor$ of lemma 2 has already been proven. The additional multiplication with o will just spread out the sequence without changing its coverage. More precisely, there are two cases:

- o is a multiple of 3: $o \equiv 0 \mod 3$
- o is not a multiple of 3: $o \bmod 3 \neq 0$

In the first case we can rewrite with $o = 3^j o'$:

$$\left\lfloor \frac{o3^i}{8} \right\rfloor = \left\lfloor \frac{o'3^{i+j}}{8} \right\rfloor \tag{19}$$

This translates the first case just in the second case. And for that second case, we can claim, that o, 2 and 3 all do not share a common divisor. Consequently the periodicity of $2 \cdot 2^k$ is still applicable. ☐

5 Collatz Conjecture and Number Classes

In this section, we will transform the sequence $\{c_i\}$ generated by Collatz-steps into a sequence of number classes $\{o_j\}$. This helps to avoid the trivial part of the Collatz-sequence.

Any Collatz-Step can be broken down into two fundamental operations: a straightforward division by 2 and a more complex transformation involving multiplication by 3 and addition of 1. For any even number, the trivial operation is iteratively applied until an odd result is obtained. This means, the trivial operation maps any even number into its number class o. This can be directly achieved by applying $o = m(n)$. In contrast, the non-trivial step induces a change in the number class, with the exception of $n = 2^i$ with $i \in \mathbb{N}_0$.

The successor o_{i+1} for $o_i = m(c_i)$ can be found by application of a Collatz step:

$$o_{i+1} = m(c_{i+1}) = \begin{cases} m\left(\frac{c_i}{2}\right) = m(c_i) = o_i & \text{if } c_i \equiv 0 \pmod 2, \\ m(3c_i + 1) = m(3o_i + 1) & \text{if } c_i \equiv 1 \pmod 2, \end{cases} \tag{20}$$

It is obvious, that the trivial step translates into an identity function. This means that e.g. c_i, c_{i+1}, c_{i+2} all may translate to the same $o_i = o_{i+1} = o_{i+2}$. In contrary to this the non-trivial step yields $c_{i+1} \neq c_i$ for all $m(c_i) \neq 1$.

The identity function may yield repeated, duplicate values of o_i. If we drop the one-to-one relation from c_i to o_i and only use the deduplicated sequence $\{o_j\}$ then o_{j+1} can be derived by:

$$o_{j+1} = m(3o_j + 1) \tag{21}$$

Corollary 1. *For any $n \in \mathbb{N}$, the sequence of odd numbers defined by:*

$$o_0 = m(n),$$

and recursively for all $i \geq 0$,

$$o_{i+1} = m(3o_i + 1) \tag{22}$$

will eventually reach the number 1, which is equivalent to the validity of the Collatz Conjecture

Example for illustration with Collatz-sequence versus sequence as per corollary:

i	c_i	$m(c_i)$	j	o_j	o_{j+1}
0	26	13	0	13	5
1	13	13			
2	40	5	1	5	1
3	20	5			
4	10	5			
5	5	5			
6	16	1	2	1	1
7	8	1			
8	4	1			
9	2	1			
10	1	1	3	1	1
11	4	1	4	1	1
12	2	1	5	1	1
13	1	1	6	1	1

Table 2: Example for sequence c_i versus o_j

6 Reformulation of the Collatz Conjecture

Corollary 2. *Our reformulation states that for any positive integer n, the sequence defined by:*

$$a_0 = m(n),$$

and recursively for all $i \geq 0$,

$$a_{i+1} = 3a_i + 2^{t(a_i)} \tag{23}$$

will eventually reach the number class 1, which is equivalent to the validity of the Collatz Conjecture.

This means for any positive integer n, there exists $k \in \mathbb{N}$, that the following equations hold true for any $i \geq k$:

$$m(a_i) \equiv 1 \tag{24}$$

$$d(a_i) \equiv 0 \tag{25}$$

$$\log_2(a_i) \in \mathbb{N}_0 \tag{26}$$

We call the sequence $\{a_i : i \in \mathbb{N}_0\}$ the reformulated Collatz-sequence.

Proof. The evidence will be provided, that the sequence a_i will map to the sequence c_i and due to the Corollary 1 the three equations are valid. Any value a_i can be written as $a_i = m(a_i)2^{t(a_i)}$. Consequently, the sequence a_i can be transformed into a sequence of number classes like this $o_i = m(a_i)$. Especially the first element $o_0 = m(a_0) = m(n)$ matches to the definition of o_0 in Corollary 1.

Let's check for the next element o_{i+1}:

$$\begin{aligned} o_{i+1} &= m(a_{i+1}) \\ &= m(3a_i + 2^{t(a_i)}) \\ &= m\left(3m(a_i)2^{t(a_i)} + 2^{t(a_i)}\right) \\ &= m\left((3m(a_i) + 1)2^{t(a_i)}\right) \\ &= m(3m(a_i) + 1) \\ &= m(3o_i + 1) \end{aligned}$$

This is exactly the definition as given in Corollary 1. □

As consequence of o_i eventually reaching 1 as defined in the corollary 1, the following equations hold true for any $i \geq k$:

$$\begin{aligned} m(a_i) &= m(1) = 1 \\ d(a_i) &= d(m(1)) = 0 \\ \log_2(a_i) &= \log_2(m(a_i)2^{t(a_i)}) = \log_2(1 * 2^{t(a_i)}) = t(a_i) \in \mathbb{N} \end{aligned}$$

Lemma 4. *The reformulation allows to write a closed form for a_k and $k \in \mathbb{N}$:*

$$a_k = 3^k a_0 + \sum_{i=0}^{k-1} 3^{k-1-i} 2^{t(a_i)} \tag{27}$$

$$= 3^k \left(a_0 + \frac{1}{3} \sum_{i=0}^{k-1} \frac{2^{t(a_i)}}{3^i} \right) \tag{28}$$

$$= 3^k \left(a_0 + \frac{1}{3} \sum_{i=0}^{k-1} \frac{a_i}{m(a_i)3^i} \right) \tag{29}$$

Proof. Let us first check the trivial version for $k = 1$. Based on the reformulation:

$$a_1 = 3a_0 + 2^{t(a_0)}$$

The closed form yields the correct result:

$$\begin{aligned} a_{k=1} &= 3^1 a_0 + \sum_{i=0}^{0} 3^{1-1-i} 2^{t(a_i)} \\ &= 3a_0 + 2^{t(a_0)} \end{aligned}$$

Now calculate the general step:

$$
\begin{aligned}
a_{k+1} &= 3a_k + 2^{t(a_k)} \\
&= 3 * \left(3^k a_0 + \sum_{i=0}^{k-1} 3^{k-1-i} 2^{t(a_i)} \right) + 2^{t(a_k)} \\
&= 3^{k+1} a_0 + \sum_{i=0}^{k-1} 3^{(k+1)-1-i} 2^{t(a_i)} + 2^{t(a_k)} \\
&= 3^{k+1} a_0 + \sum_{i=0}^{k-1} 3^{(k+1)-1-i} 2^{t(a_i)} + 3^{(k+1)-1-k} 2^{t(a_k)} \\
&= 3^{k+1} a_0 + \sum_{i=0}^{k-1} 3^{(k+1)-1-i} 2^{t(a_i)} + \sum_{i=k}^{(k+1)-1} 3^{(k+1)-1-i} 2^{t(a_i)} \\
&= 3^{k+1} a_0 + \sum_{i=0}^{(k+1)-1} 3^{(k+1)-1-i} 2^{t(a_i)}
\end{aligned}
$$

This is exactly the closed form for a_{k+1}. □

7 Collatz Conjecture: Actually a Race

In the following, we are investigating into some practical aspects of this sequence. It is well known, that the number 27 needs 111 Collatz-Steps to reach 1. Based on our reformulation, this translates into 41 steps, which are the non-trivial Collatz-steps. The sequence can be shown like this:

i	a_i	a_i in binary	$l(a_i)$	$t(a_i)$	Δt
0	27	11011	4	0	0
1	82	1010010	6	1	1
2	248	11111000	7	3	2
3	752	1011110000	9	4	1
4	2272	100011100000	11	5	1
5	6848	1101011000000	12	6	1
6	20608	101000010000000	14	7	1
7	61952	1111001000000000	15	9	2
8	186368	101101100000000000	17	11	2
9	561152	10001001000000000000	19	12	1
10	1687552	110011100000000000000	20	14	2
11	5079040	10011011000000000000000	22	15	1
12	15269888	111010010000000000000000	23	16	1
13	45875200	10101111000000000000000000	25	18	2
14	137887744	1000001110000000000000000000	27	19	1
15	414187520	11000101100000000000000000000	28	20	1
16	1243611136	1001010001000000000000000000000	30	21	1
17	3732930560	11011110100000000000000000000000	31	23	2
18	11207180288	1010011100000000000000000000000000	33	26	3
19	33688649728	11111011000000000000000000000000000	34	27	1
20	101200166912	1011110010000000000000000000000000000	36	28	1
21	303868936192	100011011000000000000000000000000000000	38	30	2
22	912680550400	1101010010000000000000000000000000000000	39	31	1
23	2740189134848	100111111000000000000000000000000000000000	41	33	2
24	8229157339136	1110111110000000000000000000000000000000000	42	34	1
25	24704651886592	101100111100000000000000000000000000000000000	44	35	1
26	74148315398144	10000110111000000000000000000000000000000000000	46	36	1
27	222513665671168	110010100110000000000000000000000000000000000000	47	37	1
28	667678435966976	10010111110100000000000000000000000000000000000000	49	38	1
29	2003310185807872	111000111100000000000000000000000000000000000000000	50	41	3
30	6012129580679168	10101010111000000000000000000000000000000000000000000	52	42	1
31	18040786788548608	1000000000110000000000000000000000000000000000000000000	54	43	1
32	54131156458668032	11000000010100000000000000000000000000000000000000000000	55	44	1
33	162411061562048512	1001000001000000000000000000000000000000000000000000000000	57	48	4
34	487514659662856192	11011000100000000000000000000000000000000000000000000000000	58	50	2
35	1463669878895411200	1010001010000000000000000000000000000000000000000000000000000	60	52	2
36	4395513236313604096	11110100000000000000000000000000000000000000000000000000000000	61	56	4
37	13258597302978740224	1011100000000000000000000000000000000000000000000000000000000000	63	59	3
38	40352252661239644160	100011000000000000000000000000000000000000000000000000000000000000	65	60	1
39	122209679488325779456	1101010000000000000000000000000000000000000000000000000000000000000	66	61	1
40	368934881474191032320	101000000000000000000000000000000000000000000000000000000000000000000	68	66	5
41	1180591620717411303424	10000000000000000000000000000000000000000000000000000000000000000000000	70	70	4

Table 3: Example for reformulated Collatz-Sequence

The average change of $t(a_i)$ can be calculated to $\Delta t = \dfrac{t(a_{41}) - t(a_0)}{41} \approx 1.707$. This can be generalized for any

k, which reaches end of the sequence aka smallest $k \in \mathbb{N}$ with $m(a_k) = 1$:

$$\begin{aligned}
\Delta t =& \frac{t(a_k) - t(a_0)}{k} && (30)\\
=& \frac{l(a_k) - (l(a_0) - d(a_0))}{k} && (31)\\
=& \frac{l(a_k) - l(a_0) + d(a_0)}{k} && (32)\\
\approx& \frac{\lfloor \log_2(a_0 3^k) \rfloor - \lfloor \log_2(a_0) \rfloor + d(a_0)}{k} && (33)\\
=& \frac{\lfloor \log_2(a_0) + \log_2(3^k) \rfloor - \lfloor \log_2(a_0) \rfloor + d(a_0)}{k} && (34)\\
\approx& \frac{\lfloor k \log_2(3) \rfloor + d(a_0)}{k} && (35)\\
\approx& \log_2(3) + \frac{d(a_0)}{k} && (36)
\end{aligned}$$

The binary sequence clearly illustrates a race between the leading bit and the trailing bit, where the leading bit begins with an initial advantage. According to the Collatz Conjecture, regardless of the starting value, the trailing bit ultimately prevails.

The leading bit runs at a constant pace of $\log_2(3)$ bits per step. While the trailing bit follows with steps drawn from a pseudo-random process, which on average are larger than $\log_2(3)$ bits per steps.

The focus of our proof is not the sequence length. The Collatz Conjecture postulates only finite length. Our focus is that the pace of the trailing bit is statistically larger than $\log_2(3)$. Due to this statistical approach, at some point the trailing bit will always win the race.

8 Basis for Proof by Contradiction

For our proof, we are negating corollary 2 and assume there exists a sequence $\{a_i\}$ with a starting value $a_0 \in \mathbb{O}$, which does not lead to $m(a_i) = 1$. Let the set of all elements in this sequence be denoted as $\mathbb{V}$. This implies that for all $a_i \in \mathbb{V}$, there is no element for which $d(a_i) \equiv 0$.

The Collatz Conjecture states that such sequences and consequently any $\mathbb{V}$ do not exist. Our proof is based on the postulation, that $\mathbb{V}$ exists, which leads to a logical contradiction, which negates the postulation. Consequently the Collatz Conjecture must be correct.

Assuming there exists at least one $\mathbb{V}$, we select one element from $\mathbb{V}$ as a_0 with the following property:

$$d(a) \geq d(a_0) \text{ for each } a \in \mathbb{V} \quad (37)$$

We denote this minimum value $d(a_0)$ as d_0.

In course of this paper, d_0 is used to derive a reasonable upper limit for the elements $\{a_k\}$. Our proof relies on violation of the inequality 37. This means, the sequence $\{d(a_i)\}$ will eventually produce values smaller than d_0. This will prove the non-existence of $\mathbb{V}$.

Lemma 5. *Any starting value a_0 chosen by equation 37 fulfills:*

$$a_0 \bmod 3 \neq 1 \quad (38)$$

Proof. If a_0 would be of the form $1 + 3n$, then n would be a predecessor of a_0 in the sequence $\{a_i\}$ and $d(n) < d(a_0)$. This violates inequality 37 and as such the lemma is correct. □

9 Upper Bound for $l(a_i)$

In the following we are deriving bounds for a_k as a function of k with the assumption of equation 37.

$$a_1 = 3a_0 + 2^{t(a_0)} = 3a_0 + \frac{2^{l(a_0)}}{2^{d(a_0)}} = 3a_0 + \frac{2^{l(a_0)}}{2^{d_0}} \leq 3a_0 + \frac{a_0}{2^{d_0}} = a_0(3 + \frac{1}{2^{d_0}}) \quad (39)$$

Based on equation 37, we can limit $d(a_i) \geq d_0$ for all $i > 0$:

$$a_{i+1} = 3a_i + 2^{t(a_i)} = 3a_i + \frac{2^{l(a_i)}}{2^{d(a_i)}} \leq 3a_i + \frac{2^{l(a_i)}}{2^{d_0}} \leq 3a_i + \frac{a_i}{2^{d_0}} = a_i(3 + \frac{1}{2^{d_0}}) \tag{40}$$

Closed form:

$$a_i \leq a_0 \left(3 + \frac{1}{2^{d_0}}\right)^i \tag{41}$$

An obvious lower limit is:

$$a_i > a_0 3^i \tag{42}$$

Based on the two equations, the function $l(a_i)$ can be limited to:

$$\lfloor \log_2(a_0) + i\log_2(3) \rfloor < l(a_i) \leq \left\lfloor \log_2(a_0) + i\log_2\left(3 + \frac{1}{2^{d_0}}\right)\right\rfloor \tag{43}$$

Slightly rewritten:

$$\lfloor \log_2(a_0) + i\log_2(3) \rfloor < l(a_i) \leq \left\lfloor \log_2(a_0) + i\log_2(3) + i\log_2\left(1 + \frac{1}{3 * 2^{d_0}}\right)\right\rfloor \tag{44}$$

This indicates that $l(a_i)$ is monotonically increasing in average on each step with $\frac{\Delta}{\Delta i} l(a_i) \approx \log_2(3) \approx 1.59$. [1]

Remark for the case $d(a_i) = 0$: In this case $l(a_i)$ will grow exactly by 2 on each step.

10 Estimation for $t(a_i)$

In order to work on an estimation for the upper limit for the trailing bit of a_i, a practical calculation with the least five bits of a_i will be provided. Hereby we ignore the bit shift $2^{t(a_i)}$.

$a_i \mod 32$	$a_{i+1} \mod 32$	$\Delta t_{i\to i+1}$	$a_{i+2} \mod 32$	$\Delta t_{i\to i+2}$	$a_{i+3} \mod 32$	$\Delta t_{i\to i+3}$	pace
$1 = 00001_b$	$4 = 00100_b$	2	$16 = 10000_b$	4	$0 = 00000_b$	≥ 5	faster
$3 = 00011_b$	$10 = 01010_b$	1	$0 = 00000_b$	≥ 5			faster
$5 = 00101_b$	$16 = 10000_b$	4	$0 = 00000_b$	≥ 5			faster
$7 = 00111_b$	$22 = 10110_b$	1	$4 = 00100_b$	2	$16 = 10000_b$	4	slower
$9 = 01001_b$	$28 = 11100_b$	2	$24 = 11000_b$	3	$16 = 10000_b$	4	slower
$11 = 01011_b$	$2 = 00010_b$	1	$8 = 01000_b$	3	$0 = 00000_b$	≥ 5	faster
$13 = 01101_b$	$8 = 00100_b$	2	$16 = 10000_b$	4	$0 = 00000_b$	≥ 5	faster
$15 = 01111_b$	$14 = 01110_b$	1	$12 = 01100_b$	2	$8 = 01000_b$	3	slower
$17 = 10001_b$	$20 = 10100_b$	2	$0 = 00000_b$	≥ 5			faster
$19 = 10011_b$	$26 = 11010_b$	1	$16 = 10000_b$	4	$0 = 00000_b$	≥ 5	faster
$21 = 10101_b$	$0 = 00000_b$	≥ 5					faster
$23 = 10111_b$	$6 = 00110_b$	1	$20 = 10100_b$	2	$0 = 00000_b$	≥ 5	faster
$25 = 11001_b$	$12 = 01100_b$	2	$8 = 01000_b$	3	$0 = 00000_b$	≥ 5	faster
$27 = 11011_b$	$18 = 10010_b$	1	$24 = 11000_b$	3	$16 = 10000_b$	4	slower
$29 = 11101_b$	$24 = 11000_b$	3	$16 = 10000_b$	4	$0 = 00000_b$	≥ 5	faster
$31 = 11111_b$	$30 = 11110_b$	1	$28 = 11100_b$	2	$24 = 11000_b$	3	slower

Table 4: Trailing bit change for least five bits of o_i

Five of the entries are slower and the others are faster than the leading bit, which runs with $\log_2(3) \approx 1.585$ per step.

If all entries are occurring with same rate, then the average Δt per step is faster than the leading bit:

$$\Delta t \approx \frac{\frac{5}{3} + \frac{5}{2} + \frac{5}{2} + \frac{4}{3} + \frac{4}{3} + \frac{5}{3} + \frac{5}{3} + \frac{3}{3} + \frac{5}{2} + \frac{5}{3} + \frac{5}{1} + \frac{5}{3} + \frac{5}{3} + \frac{4}{3} + \frac{5}{3} + \frac{3}{3}}{16} = \frac{\frac{53}{3} + \frac{15}{2} + 5}{16} \approx 1.885$$

[1] Based on literature, it has been proven that all numbers $n \leq 2^{68}$ are fulfilling the Collatz Conjecture. Consequently, there cannot be any element $n \in \mathbb{V}$ with $d(n) = d_0 = 67$. This means: $\log_2\left(3 + \frac{1}{2^{d_0}}\right) - \log_2(3) < 3.259 * 10^{-21}$

The average after one step ($\Delta t_{i \to i+1}$) is still faster than the leading bit:

$$\Delta t \geq \frac{2+1+4+1+2+1+2+1+2+1+5+1+2+1+3+1}{16} = \frac{30}{16} = 1.875$$

The average of the slower steps are:

$$\Delta t_{\text{slower}} = \frac{\frac{4}{3}+\frac{4}{3}+\frac{3}{3}+\frac{4}{3}+\frac{3}{3}}{5} = \frac{6}{5} = 1.2$$

and the faster:

$$\Delta t_{\text{faster}} \geq \frac{\frac{5}{3}+\frac{5}{2}+\frac{5}{2}+\frac{5}{3}+\frac{5}{3}+\frac{5}{2}+\frac{5}{3}+\frac{5}{1}+\frac{5}{3}+\frac{5}{3}+\frac{5}{3}}{11} = \frac{\frac{35}{3}+\frac{15}{2}+5}{11} \approx 2.197$$

This observation gives reason for the following lemma:

Lemma 6. *If the bits inbetween leading and trailing bits of a_i are 1 with a probability of 0.5 and else 0 - e.g. due to a pseudo-random process - then the following shall be true:*

$$\lim_{i \to \infty} \frac{t(a_i) - t(a_0)}{i} \to 2 \tag{45}$$

Proof. Let's consider $o_i = m(a_i)$ instead of a_i, which can be written in a binary representation:

$$o_i = \sum_{j=0}^{l(o_i)} b_j 2^j \text{ with } b_0 = b_{l(o_i)} = 1 \tag{46}$$

Based on the corollary 2 the next element can be calculated as:

$$a_{i+1} = 3a_i + 2^{t(a_i)} \tag{47}$$

So we are calculating

$$a_{i+1} = 3o_i + 1 \tag{48}$$

o_i is per definition odd, the resulting a_{i+1} is even. This means the trailing bit of o_i aka $t(o_i) = 0$ has been increased by a minimum of 1 aka $t(a_{i+1}) \geq t(o_i) + 1 = 1$.

In the following we are focussing on bit position p:

$$b'_p = \text{bit}\left(3o_i + 1, p\right) = \text{bit}\left(1 + 3\sum_{j=0}^{l(o_i)} b_j 2^j, p\right) \tag{49}$$

It is obvious, that all elements with $i > p$ have no influence on bit b'_p. So we can rewrite:

$$b'_p = \text{bit}\left(3o_i + 1, p\right) = \text{bit}\left(1 + 3\sum_{j=0}^{p} b_j 2^j, p\right) \tag{50}$$

Now we isolate the actual bit b_p from the sum:

$$b'_p = \text{bit}\left(3o_i + 1, p\right) = \text{bit}\left(1 + 3b_p 2^p + 3\sum_{j=0}^{p-1} b_j 2^j, p\right) \tag{51}$$

Due to the fact that $3 * 2^p$ yields bit p and $p+1$, while only p is relevant, it can be simplified to:

$$b'_p = \text{bit}\left(3o_i + 1, p\right) = \text{bit}\left(1 + b_p 2^p + 3\sum_{j=0}^{p-1} b_j 2^j, p\right) \tag{52}$$

$bit(1+3\sum)$	b_p		b'_p
0	0	$\rightarrow$	0
1	0	$\rightarrow$	1
0	1	$\rightarrow$	1
1	1	$\rightarrow$	0

Table 5: Bit transitions b_p to b'_p

With the algebraic rules for bit (i,p), we can rewrite to:

$$b'_p = \text{bit}\left(3o_i+1,p\right) = \left(\text{bit}\left(1+3\sum_{j=0}^{p-1} b_j 2^j, p\right) + \text{bit}\left(b_p 2^p, p\right)\right) \bmod 2 \tag{53}$$

$$= \left(\text{bit}\left(1+3\sum_{j=0}^{p-1} b_j 2^j, p\right) + b_p\right) \bmod 2 \tag{54}$$

The consequence of this equation is quite significant. Let's have a short overview of the outcomes: This means in logical terms b_p is exor'ed with the original result. Consequently, independent from all b_i with $i < p$, the bit b_p can always force the state of b'_p to any value $\{0,1\}$.

In order to assess an estimate for $t(a'_{i+1})$ we need to count the 0s starting with $i=0$ until the first 1 bit is met. Based on the assumption, that the occurence rate of 0 is 0.5, the estimate for $t(a'_{i+1})$ is:

$$X\left(t(a_{i+1}) - t(a_i)\right) \approx 1 + \sum_{i=1}^{l(o_i)} (0.5)^i = 1.11111\cdots_b \approx 2 \tag{55}$$

Based on the law of large numbers the following is valid:

$$\lim_{i\to\infty} \frac{t(a_i) - t(a_0)}{i} = \lim_{i\to\infty} \frac{\sum_{j=0}^{i-1} t(a_{j+1}) - t(a_j)}{i} \approx \lim_{i\to\infty} \frac{i * X\left(t(a_{i+1}) - t(a_i)\right)}{i} \rightarrow 2 \tag{56}$$

□

11 Equidistribution

In the following subsections we are focussing on any five consecutive bits ($2^5 = 32$) defined as:

$$f_{i,j} = \left\lfloor \frac{a_i}{2^j} \right\rfloor \bmod 32 \text{ with } j \in \mathbb{N} \tag{57}$$

Our choice on five bits is just arbitrarily for better illustration and explanation.

In case j is chosen, that $j \leq l(a_0)$, then $f_{0,i}$ will equal the respective bits of a_0 as starting value. Otherwise $f_{0,j}$ is 0. For illustration in mathematical terms:

$$f_{0,j} = \begin{cases} \left\lfloor \frac{a_0}{2^j} \right\rfloor \bmod 32 & \text{if } j \leq l(a_0) \\ 0 & \text{if } j > l(a_0) \end{cases} \tag{58}$$

Moreover we are splitting the reformulated sequence at the chosen j. For illustration in table 6, we are using the same starting number $a_0 = 27$ and select $j = 4$ as split point.

This table already displays two more sequences $\{r_{i,j}\}$ and $\{v_{i,j}\}$. r is the remainder of the division of a_i by 2^j. In other words the bits of a_i on position $0\ldots < j$. The sequence $\{v_i\}$ collects all overflows from the remainder $r_{i,j}$ into $f_{i,j}$.

Mathematical definition:

$$r_{i,j} = a_i \bmod 2^j \text{ with } j \in \mathbb{N} \tag{59}$$

$$v_{i,j} = \left\lfloor \frac{3r_{i,j} + 2^{t(a_i)}}{2^j} \right\rfloor \tag{60}$$

It is important to note, that $0 \leq v_{i,j} \leq 2$.

i	a_i	a_i in binary			$v_{i,4} = \left\lfloor \frac{3r_{i,4}+2^{t(a_i)}}{2^4} \right\rfloor$
			$f_{i,4}$	$r_{i,4}$	
0	27		00001	1011	1
1	82		00101	0010	0
2	248		01111	1000	1
3	752	1	01111	0000	0
4	2272	100	01110	0000	0
5	6848	1101	01100	0000	0

Table 6: Example for $f_{i,j}$, $r_{i,j}$ and $v_{i,j}$

11.1 Transition of 5-Bit Values

Given the previous definitions, we analyze the transitions between the 32 possible values of the 5-bit representation. The following table details the transformation of each 5-bit value f_i based on the value of v_i, resulting in the next 5-bit value f_{i+1}.

f_i	$v_i = 0$		$v_i = 1$		$v_i = 2$	
	v_{i+1}	f_{i+1}	v_{i+1}	f_{i+1}	v_{i+1}	f_{i+1}
00000	0	00000	0	00001	0	00010
00001	0	00011	0	00100	0	00101
00010	0	00110	0	00111	0	01000
00011	0	01001	0	01010	0	01011
00100	0	01100	0	01101	0	01110
00101	0	01111	0	10000	0	10001
00110	0	10010	0	10011	0	10100
00111	0	10101	0	10110	0	10111
01000	0	11000	0	11001	0	11010
01001	0	11011	0	11100	0	11101
01010	0	11110	0	11111	1	00000
01011	1	00001	1	00010	1	00011
01100	1	00100	1	00101	1	00110
01101	1	00111	1	01000	1	01001
01110	1	01010	1	01011	1	01100
01111	1	01101	1	01110	1	01111
10000	1	10000	1	10001	1	10010
10001	1	10011	1	10100	1	10101
10010	1	10110	1	10111	1	11000
10011	1	11001	1	11010	1	11011
10100	1	11100	1	11101	1	11110
10101	1	11111	2	00000	2	00001
10110	2	00010	2	00011	2	00100
10111	2	00101	2	00110	2	00111
11000	2	01000	2	01001	2	01010
11001	2	01011	2	01100	2	01101
11010	2	01110	2	01111	2	10000
11011	2	10001	2	10010	2	10011
11100	2	10100	2	10101	2	10110
11101	2	10111	2	11000	2	11001
11110	2	11010	2	11011	2	11100
11111	2	11101	2	11110	2	11111

Table 7: Transition table for five bits

11.2 Stochastic Modeling of f_i and v_i

To analyze the behavior of the sequence $\{f_{i,j}\}$, we consider the frequency of occurrence of values within the range $0 \ldots 31$. We define a probability distribution over these values, assigning a probability $p_{i,k}$ to the event

$f_{i,j} = k$, where $k \in \{0, 1, \ldots, 31\}$. Since one value must occur, the probabilities must sum up to 1:

$$\sum_{k=0}^{31} p_{i,k} = 1 \tag{61}$$

Next, we consider the values $v_{i,j}$ resulting from the Collatz iteration. We define the occurrence rates for $v_{i,j} \in \{0, 1, 2\}$ as p_0, p_1, and p_2, respectively, subject to the constraint $p_0 + p_1 + p_2 = 1$. The probability vector $\vec{p_i} = (p_{i,0}, p_{i,1}, \ldots, p_{i,31})$ and the transition probabilities p_0, p_1, p_2 determine the subsequent probability vector $\vec{p}_{i+1}$. This relationship can be expressed as a matrix transformation:

$$A(p_0, p_1, p_2) \cdot \vec{p_i} = \vec{p}_{i+1}$$

In the following we use matrix indices starting from 0.

The matrix A represents the transition probabilities between the states defined by the values of f_i. Specifically, the row indexed by k of A represents the probability of transitioning to the value k. To construct A, we consider the transition table 7. For example, if $f_i = 10$, the next value f_{i+1} can be 30, 31, or 0 depending on v_i. Therefore, the 30th, 31st, and 0th entries in the 10th column of A (or $a_{30,10}$,$a_{31,10}$, $a_{0,10}$) are p_0, p_1, and p_2, respectively, with all other entries being zero.

Here is the complete matrix:

$$\begin{bmatrix}
p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2
\end{bmatrix} \cdot \begin{bmatrix} p_{i,0} \\ p_{i,1} \\ \vdots \\ p_{i,31} \end{bmatrix} = \begin{bmatrix} p_{i+1,0} \\ p_{i+1,1} \\ \vdots \\ p_{i+1,31} \end{bmatrix}$$

For each step from i to $i+1$ this matrix multiplication occurs. This means, we can expect, that the probability vector $\vec{p_i}$ converges to $\vec{p}$ for large i. With the matrix D being the diagonal matrix with all elements 1, we can write:

$$A(p0, p1, p2) \cdot \vec{p} = \vec{p} \tag{62}$$

$$(A(p0, p1, p2) - D) \cdot \vec{p} = 0 \tag{63}$$

The matrix difference is:

$$A{-}D = \begin{bmatrix}
p0-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
p1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
p2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & p0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & p1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & p2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & p0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & p1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & p1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & p2 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & p0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & p1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & p2 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & p0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & p1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & p2-1
\end{bmatrix}$$

Without diving into the complexities of solving this matrix, there is an obvious, non-trivial solution: Every row contains once each of -1, $p0$, $p1$, $p2$, or a partial sum if any $p0$, $p1$, $p2$ coincide with -1. A multiplication with any vector $\vec{x} = c * \vec{1}$ will yield for each row: $c * (-1 + p0 + p1 + p2) = c * 0 = 0$. This means any $c \neq 0$ fulfills the equation $A \cdot \vec{x} = \vec{0}$. The sum of all vector elements must be 1. So $c = \frac{1}{32}$.

Consequently, the repeated application of $f_{i+1} = f_i * 3 + \{0, 1, 2\}$ will converge to an equi-distribution of the 32 values for large values of i, if the values 0, 1, 2 are themselves evenly distributed. This means, if there is no correlation pattern between f_i and v_i, that means $p0/p1/p2$ are not linked to $p_{i,j}$.

Similar calculation for the v_i aka $p0$,$p1$,$p2$: The values 11 or 10 are results by counting the rows in table 7 with $C = 0, 1, 2$ in the respective column.

$$\begin{bmatrix} 11c & 11c & 10c \\ 11c & 10c & 11c \\ 10c & 11c & 11c \end{bmatrix} \cdot \begin{bmatrix} p0_i \\ p1_i \\ p2_i \end{bmatrix} = \begin{bmatrix} p0_{i+1} \\ p1_{i+1} \\ p2_{i+1} \end{bmatrix}$$

And for the stable situation:

$$\begin{bmatrix} 11c-1 & 11c & 10c \\ 11c & 10c-1 & 11c \\ 10c & 11c & 11c-1 \end{bmatrix} \cdot \begin{bmatrix} p0 \\ p1 \\ p2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The matrix is then:

$$A = c \begin{bmatrix} 11-32 & 11 & 10 \\ 11 & 10-32 & 11 \\ 10 & 11 & 11-32 \end{bmatrix} = c \begin{bmatrix} -21 & 11 & 10 \\ 11 & -22 & 11 \\ 10 & 11 & -21 \end{bmatrix}$$

Again, the trivial solution is $p0 = p1 = p2 = \frac{1}{3}$.

Therefore, the assertion that the bit sequence f_i generated by the Collatz iteration can be treated as statistically independent and unbiased is rigorously justified by the preceding analysis. This validation forms a crucial foundation for our final conclusion.

12 Conclusion

We have established a bound on the growth rate of the leading bit position, $l(a_n)$, as described by inequality (44). This analysis reveals that $l(a_k)$ grows linearly with a rate of approximately $\log_2(3) \approx 1.59$. Comparing this to the estimated growth rate of the trailing bit position, $t(a_k) \approx 2$ (equation (55)), we observe a differential growth.

We can approximate this differential growth aka distance $d(a_k) = l(a_k) - t(a_k)$ as:

$$\frac{\Delta}{\Delta k} d(a_k) = \frac{\Delta}{\Delta k}(l(a_k) - t(a_k)) \approx 1.59 - 2 = -0.41 < 0 \tag{64}$$

This negative rate of change indicates that, for sufficiently large k, the distance $d(a_k)$ will decrease. Consequently, there exists a value k^* such that $d(a_{k^*}) < d_0$, where d_0 is the initial minimum distance defined in our assumptions.

More formally, this can be expressed as:

$$\lim_{i \to k} d(a_i) = \lim_{i \to k}(l(a_i) - t(a_i)) \tag{65}$$

$$\approx \lim_{i \to k}(l(a_0) + i \log_2(3) - t(a_i)) \tag{66}$$

$$= \lim_{i \to k}(d_0 + i \log_2(3) - t(a_i)) \tag{67}$$

$$\approx d_0 + i \log_2(3) - 2i < d_0 \tag{68}$$

It is noteworthy that this analysis holds as long as $m(a_i) \neq 1$. When $m(a_i) = 1$, the speed of the leading bit jumps to 2, and the distance $d(a_i)$ remains zero. We avoid considering the limit as $i \to \infty$ because our estimate for the leading bit speed is predicated on the condition $m(a_i) \neq 1$. Given our initial assumption that the Collatz Conjecture is false, an infinite sequence not converging to 1 cannot include the case where $m(a_i) = 1$.

Therefore, assuming the Collatz Conjecture is false, we arrive at a contradiction. Specifically:

- We initially assume the Collatz Conjecture is invalid, implying the existence of a set $\mathbb{V}$ with a minimum distance d_0.
- Based on this minimum d_0, we derive bounds for $l(a_k)$ for any $k \in \mathbb{N}_0$.
- We estimate $t(a_k)$ for large k based on the assumption of stochastic bit values.
- Verification confirms the validity of our stochastic assumption.
- For sufficiently large k, the value of $d(a_k)$ decreases below the initial minimum distance d_0.
- This contradicts our initial assumption regarding the existence of a minimum distance d_0.
- Consequently, the assumption of the existence of the set $\mathbb{V}$ must be incorrect.
- Therefore, the Collatz Conjecture must be valid.

13 Computational Verification

To validate our predictions regarding the differential growth rate, we perform computational experiments. We execute the reformulated Collatz sequence for a set of odd numbers with a fixed value of d. For example, for $d = d(a_0) = 3$, we test all the numbers represented as 1001_b, 1011_b, 1101_b, and 1111_b. We then average the number of steps required for each number to reach $d(a_i) \equiv 0$.

We determined the rate of increase in the average number of steps per unit change in d using equation 64, which provides the expected reduction of d per step k. As d increases, the average number of steps, k_{avg}, also increases, as more steps are needed to reduce $d(a_i)$ to 0. This relationship is approximated by:

$$d(a_{k_{avg}}) \approx d(a_0) + k_{avg} \frac{\Delta}{\Delta k} d(a_k) \approx 0 \tag{69}$$

This allows us to estimate k_{avg} as:

$$k_{avg} \approx \frac{-d(a_0)}{\frac{\Delta d}{\Delta k}} d \tag{70}$$

$$= \frac{-d(a_0)}{\log_2(3) - 2} \tag{71}$$

$$\approx 2.409 * d(a_0) \tag{72}$$

Therefore, we predicted that the average number of steps would increase by approximately 2.409 for each unit increase in d.

The following table and figure present the empirical data obtained from our experiments:

d	average steps
3	4.25
4	13.250
5	15.000
6	18.406
7	20.719
8	21.898
9	24.547
10	27.469
11	30.138
12	32.256
13	34.198
14	36.270
15	38.819
16	41.369
17	43.784
18	46.194
19	48.524
20	50.917
21	53.328
22	55.717
23	58.133

Table 8: Emperical data: average steps for all o with $d(o) = d$

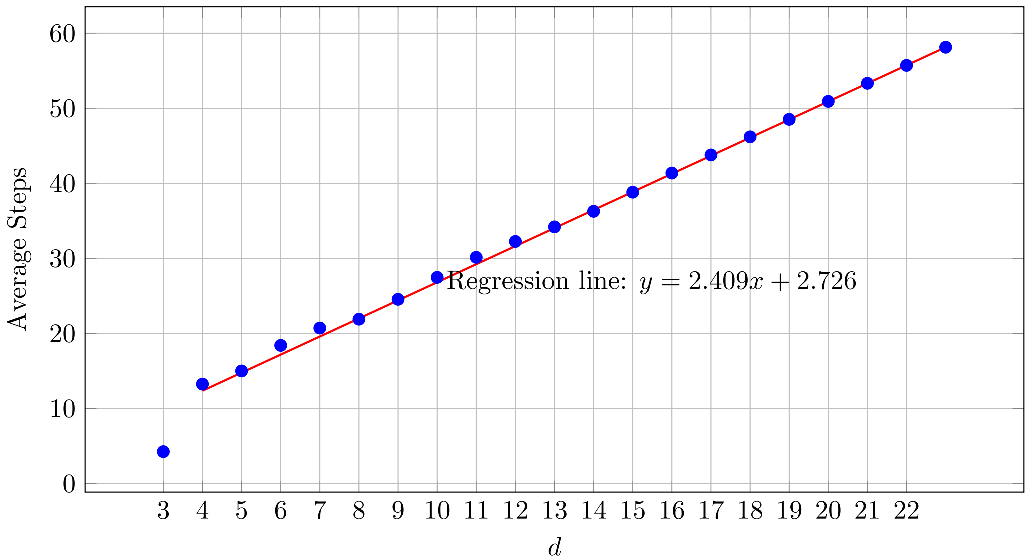

Figure 1: Plot of Average Steps against d

The experimental data confirms our analysis. The observed y-intercept of approximately 2.726 may warrant further investigation, but its statistical insignificance does not impact the overall validity of the proof regarding the Collatz Conjecture.

References

David Barina. Convergence verification of the collatz problem. *The Journal of Supercomputing*, 77(2):2001–2009, May 2020. doi: 10.1007/s11227-020-03368-x. URL `https://link.springer.com/article/10.1007/s11227-020-03368-x`. This paper discusses computational methods for verifying the Collatz conjecture, with verification up to 2^{68} reported elsewhere by Eric Roosendaal around the same period.

Jeffrey C. Lagarias. The 3x+1 problem: An overview, 2021. URL `https://arxiv.org/abs/2111.02635`.