

Christian Koch, Eldar Sultanow, Sean Cox

**Divisions by Two in Collatz Sequences:  
A Data Science Approach**

**Second Edition**

Working Paper

Technische Hochschule Nürnberg

05 May 2020

## **Abstract**

The Collatz conjecture is an unsolved number theory problem. We approach the question by examining the divisions by two that are performed within Collatz sequences. Aside from classical mathematical methods, we use techniques of data science. Based on the analysis of 10,000 sequences we show that the number of divisions by two lies within clear boundaries. Building on the results, we develop and prove an equation to calculate the maximum possible number of divisions by two for any given a Collatz sequence. Whenever this maximum is reached, a sequence leads to the result one, as conjectured by Lothar Collatz. Furthermore, we show how many divisions by two are required for a cycle of a specific length. The findings are valuable for further investigations and could form the basis for a comprehensive proof of the conjecture.

*Keywords:* Collatz Conjecture, Divisions by Two, Data Science

**Table of Contents**

1.	Introduction.....	3
2.	Boundaries of $\alpha_i$ .....	6
3.	Analysing $\alpha$ .....	8
4.	Occurrence of Cycles.....	14
5.	Summary.....	20
A.	Appendix.....	21
B.	References.....	23
C.	About the Authors.....	24

# 1. Introduction

## 1.1 The Problem

The Collatz conjecture is a well-known number theory problem and is the subject of numerous publications.<sup>1</sup> Therefore, our description of the topic will be brief. The mathematician Lothar Collatz introduced a function  $g: \mathbb{N} \rightarrow \mathbb{N}$  as follows:

(1.1)

$$g(x) = \begin{cases} 3x + 1, & \text{if } x \equiv 1 \pmod{2} \\ x/2, & \text{if } x \equiv 0 \pmod{2} \end{cases}$$

The conjecture, as treated in this paper, claims that the above function leads to the final result one for every natural starting number, when applied recursively. A series of numbers involved in this process is called a Collatz sequence. With an aim to contribute to a proof of the conjecture, this paper analyses a central aspect of the problem: the divisions by two.<sup>2</sup>

## 1.2 Determining Odd Numbers

Sultanow, Koch and Cox showed that odd numbers of Collatz sequences can be calculated with the following recursive equation:<sup>3</sup>

(1.2)

$$v_{n+1} = 3^n * v_1 * \prod_{i=1}^n \left(1 + \frac{1}{3 * v_i}\right) * \prod_{i=1}^n 2^{-\alpha_i}$$

The variable  $v_1$  denotes the first odd number of the sequence, that is, the starting value. The variable  $v_i$  symbolises the odd number that is the result of a particular iteration.<sup>4</sup> The exponent  $n$  stands for the count of odd numbers that are processed by the algorithm. In the further course of this paper we will call the parameter  $n$  the 'length' of a sequence. The exponent  $\alpha_i$  finally represents the number of divisions

---

<sup>1</sup> An overview is provided by (Lagarias, 2010)

<sup>2</sup> Details on our scientific approach can be found in section A.3.

<sup>3</sup> (Sultanow, Koch, & Cox, 2020, p. 10)

<sup>4</sup> For  $n = 1$  this is the starting value  $v_1$ .

by two that are performed in a specific iteration. Accordingly, the sum of  $\alpha_i$  is the count of divisions by two leading from the starting value  $v_1$  to the outcome  $v_{n+1}$ .<sup>5</sup>

Let us consider the example  $v_1 = 13$  and  $n = 2$ . Applying equation 1.2 yields:<sup>6</sup>

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

Starting with  $v_1 = 33$  for  $n = 3$  we obtain the result:

$$v_{3+1} = 3^3 * 33 * \left(1 + \frac{1}{3 * 33}\right) * \left(1 + \frac{1}{3 * 25}\right) * \left(1 + \frac{1}{3 * 19}\right) * 2^{-5} = 29$$

Improving readability, we denote the factor  $\left(1 + \frac{1}{3 * v_i}\right)$  with the variable  $\beta_i$ .

In addition, we generalise the formula by replacing the factor three with the variable  $k$ . This will be useful for further analysis and leads us to the following generalised version of equation 1.2:

(1.3)

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

$$v_{n+1} = k^n * v_1 * \prod_{i=1}^n \beta_i * \prod_{i=1}^n 2^{-\alpha_i}$$

In order to correctly calculate odd numbers with equation 1.3, we must first define the halting conditions of the algorithm in the next section.

### 1.3 Halting Conditions

Being compliant with the Collatz conjecture, the algorithms 1.2 and 1.3 halt if at least one of the following conditions is fulfilled:

(1.4)

- 1.)  $v_{n+1} = 1$
- 2.)  $v_{n+1} \in \{v_1, v_2, v_3 \dots v_n\}$

<sup>5</sup> For a glossary of notations see section A.4 in the appendix.

<sup>6</sup> The result of the first iteration  $v_{1+1}$  equals five.

When the first condition applies, the Collatz conjecture is true for a specific sequence. If the second condition is fulfilled, the sequence has led to a cycle. For every starting value, except  $v_1 = 1$ , the Collatz conjecture is therefore falsified.<sup>7</sup> Let us consider the example  $k = 3$ ,  $v_1 = 13$  and  $n = 2$ . Applying equation 1.3 yields:

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

In the above example the algorithm halts after two iterations because the first condition is fulfilled. If we examine the case  $v_1 = 1$ , we realise that the algorithm finishes after the first iteration, since both halting conditions are true:

$$v_{1+1} = v_1 = 3^1 * 1 * \left(1 + \frac{1}{3 * 1}\right) * 2^{-2} = 1$$

The sequence stops in the example above due to the result being one. Furthermore, the sequence has led to a cycle.

---

<sup>7</sup> This statement refers to the Collatz conjecture in its original form  $3v + 1$ .

## 2. Boundaries of $\alpha_i$

We know that in every iteration of the equations 1.2. and 1.3 at least one division by two is performed. This follows from the constraints of the Collatz problem. Consequently, we can define the minimum of  $\alpha_i$  with the following condition:

$$1 \leq \alpha_i$$

The maximum can be specified in a similarly easy way. According to the halting conditions, defined in the previous section, a Collatz sequence finishes when  $v_{n+1} = 1$ . The maximum of  $\alpha_i$ , hereinafter called  $\widehat{\alpha}_i$ , can hence be defined as:

(2.1)

$$2^{\widehat{\alpha}_i} = k * v_i + 1$$

$$\widehat{\alpha}_i = \log_2(k) + \log_2(v_i) + \log_2(\beta_i)$$

The theorem above builds on the fact that the expression  $2^{\widehat{\alpha}_i}$  must equal the next even number  $(k * v_i + 1)$  in order to lead to  $v_{n+1} = 1$ . Being greater, the result  $v_{n+1}$  would be less than one. The second step inverts the exponentiation of  $\widehat{\alpha}_i$  by taking the binary logarithm. Appropriately, we replace the operation ‘plus one’ by  $\beta_i$ . For a better understanding of the above term, let us consider the example  $k = 3$  and  $v_1 = 5$ . In this case theorem 2.1 results in:

$$\alpha_1 = \widehat{\alpha}_1 = 4 = \log_2(3) + \log_2(5) + \log_2\left(1 + \frac{1}{3 * 5}\right)$$

If a sequence reaches the maximum  $\widehat{\alpha}_i$ , it finishes with one, thus verifying the Collatz conjecture. If we could prove that every odd number finally leads to this maximum for  $k = 3$ , the Collatz problem would be solved. Summarising, we can define the following boundaries for  $\alpha_i$ :

(2.2)

$$1 \leq \alpha_i \leq \log_2(k) + \log_2(v_i) + \log_2(\beta_i)$$

Before we continue, we validate theorem 2.2 empirically. We will do so at various points in this paper to avoid obvious errors in our mathematical reasoning. The basis for the validation is a sample of 10,000 Collatz sequences. The data set comprises

information about sequences for the odd starting numbers  $v_1 \in \{1, 3, 5, \dots, 3999\}$  and  $k \in \{1, 3, 5, 7, 9\}$ . Since we do not know that all generated sequences halt, we limited the number of iterations per sequence to  $n = 100$ . For further details on the data set see A.1.

Unsurprisingly, we found that all values of  $\alpha_i$  in the sample are compliant with theorem 2.2.<sup>8</sup> In the next section we move on to more sophisticated considerations and study the properties of  $\prod_{i=1}^n 2^{\alpha_i}$ .

---

<sup>8</sup> Source: Own empirical analysis, see A.1 for details.



### 3. Analysing $\alpha$

#### 3.1 Boundaries of $\alpha$

In equations 1.2 and 1.3, the expression  $\prod_{i=1}^n 2^{\alpha_i}$  represents the divisions by two performed by the algorithms. The number of divisions by two can be determined with the following formula and will be symbolised by  $\alpha$ :

(3.1)

$$\alpha = \sum_{i=1}^n \alpha_i$$

Based on theorem 2.2 we can define the minimum of  $\alpha$  as follows:

$$n \leq \alpha$$

Since we carry out at least one division by two in every iteration of theorem 1.2 and theorem 1.3, the minimum of  $\alpha$  equals the sequence's length. The maximum value is harder to determine. In the first step we derive it empirically from the data set mentioned in the previous section. Based on the observed data we formulate the hypothesis that the maximum of  $\alpha$  can be calculated with the following equation:

(3.2)

$$\hat{\alpha} = \lfloor n * \log_2(k) + \log_2(v_1) \rfloor + 1$$

$$\alpha \leq \hat{\alpha}$$

The hypothesis holds for all Collatz sequences in the empirical data set.<sup>9</sup> If a Collatz sequence reaches the above formulated maximum, it finishes with one, as conjectured by Lothar Collatz. Let us, for example, consider the case  $v_1 = 13$ ,  $n = 2$  and  $k = 3$ . Applying theorems 3.2 and 1.3 leads to:

$$\hat{\alpha} = \lfloor 2 * \log_2(3) + \log_2(13) \rfloor + 1 = 7$$

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

---

<sup>9</sup> Source: Own empirical analysis, see A.1 for details.

Throughout the next sections we will formulate a proof of the developed hypothesis step by step.

### 3.2 Proving $\hat{\alpha}$ for $k=1$

First, we examine the case  $k = 1$ , where theorem 3.2 can be simplified as follows:

(3.3)

$$\hat{\alpha} = \lfloor n * \log_2(1) + \log_2(v_1) \rfloor + 1 = \lfloor \log_2(v_1) \rfloor + 1$$

In order to prove theorem 3.2, we have to demonstrate that the number of divisions by two,  $\alpha$ , is less than or equal to the maximum  $\hat{\alpha}$ . This can be achieved by analysing the binary representation of Collatz numbers.<sup>10</sup> Let us consider the case  $v_1 = 25$  and  $k = 1$  in the decimal system. Applying theorem 1.3 leads to the sequence shown in the following table.

<b>n</b>	<b>Variable</b>	<b>Decimal</b>	<b>Log2</b>	<b>Binary</b>	<b>Binary Length</b>	$\alpha_i$	$\alpha$	<b>Operation</b>
<b>1</b>	$v_1$	25	4.64	11001 <sub>2</sub>	5			+1
	$v_1 + 1$	26	4.70	11010 <sub>2</sub>	5	1	1	* 2 <sup>-1</sup>
<b>2</b>	$v_2$	13	3.70	1101 <sub>2</sub>	4			+1
	$v_2 + 1$	14	3.81	1110 <sub>2</sub>	4	1	2	* 2 <sup>-1</sup>
<b>3</b>	$v_3$	7	2.81	111 <sub>2</sub>	3			+1
	$v_3 + 1$	8	3.00	1000 <sub>2</sub>	4	3	5	* 2 <sup>-3</sup>
<b>4</b>	$v_4$	1	1.00	1 <sub>2</sub>	1			

Table 1: Binary representation of a Collatz sequence for  $k=1$

The sequence presented in Table 1 starts with the decimal number  $v_1 = 25$  at  $n = 1$ . Subsequently it comprises the odd numbers  $v_2 = 13$ ,  $v_3 = 7$  and finally  $v_4 = 1$ . In the binary system the sequence starts accordingly with  $v_1 = 11001_2$ .

<sup>10</sup> To avoid confusion between decimal and binary numbers, we will label binary numbers with a subscripted 2.

The binary length of the starting number  $len(v_1)$  equals five.<sup>11</sup> This observation is crucial for our proof. For clarification, it is important to note that the length of a binary number can be calculated with the following equation:<sup>12</sup>

(3.4)

$$len(v_i) = \lfloor \log_2(v_i) \rfloor + 1$$

For example, consider the case  $v_i = 13$  in decimal, rendered in binary that means  $v_i = 1101_2$ . Here, the equation 3.4 leads to the following result:

$$len(13) = len(1101_2) = \lfloor \log_2(13) \rfloor + 1 = 4$$

The comparison of equation 3.4 with theorem 3.3 makes it clear that they are identical. This raises the question why the maximum number of divisions by two of a Collatz sequence corresponds to the binary length of  $v_1$ .<sup>13</sup> To answer this, we take a closer look at the mechanics of a Collatz sequence in the binary system.

We start with  $v_1 = 11001_2$  in the above example. Adding one, we obtain the even number  $(v_1 + 1) = 11010_2$ . The binary length of  $v_1$  equals the binary length of  $(v_1 + 1)$ , which is five. Due to the trailing zero we immediately realise that  $(v_1 + 1)$  is even. A division by two can be performed in the binary system by deleting the trailing zero. The result is  $v_2 = 1101_2$ . Adding one again, leads to the next even number  $(v_2 + 1) = 1110_2$ . Deleting the trailing zero once more, results in  $v_3 = 111_2$ .

Up to this point we have performed two divisions by two. The parameter  $\alpha$  therefore equals two. The case  $v_3 = 111_2$  is very important for our proof. Adding one to  $v_3 = 111_2$  leads to an overflow of the binary number. As a result, we obtain the even number  $(v_3 + 1) = 1000_2$ , which is a power of two and equals  $2^3$  in decimal. Knowing that every power of two in a Collatz sequence directly leads to the terminal value  $v_{n+1} = 1$ , we can deduce that the sequence ends after the third iteration.

---

<sup>11</sup> With binary length we mean the count of digits of a binary number.

<sup>12</sup> (Sedgewick & Wayne, 2011, p. 185)

<sup>13</sup> The statement is only true for  $k = 1$ .

The binary length  $len(v_3) = 3$  increases to  $len(v_3 + 1) = 4$  in the final step. This situation only occurs once in a Collatz sequence for  $k = 1$ . Whenever adding one to a number  $v_n$  causes an overflow of its binary representation, the result  $(v_n + 1)$  will be a power of two. The binary length will, in this scenario, increase from  $len(v_n)$  to  $len(v_n) + 1$ . The sequence will consequently halt. For all other cases the following condition applies:<sup>14</sup>

(3.5)

$$len(v_n) = len(v_n + 1) > len(v_{n+1})$$

Only the final iteration increases the length of the binary number. In any other case the binary length decreases from  $v_n$  to  $v_{n+1}$ .

Let us now reflect what this implies for the maximum  $\hat{\alpha}$ . We know that the binary length of the starting value  $v_1$  can be calculated with equation 3.4. In order to reach the final result  $v_{n+1} = 1$ , starting at  $v_1$ , we have to perform the following number of divisions by two:

(3.6)

$$\alpha = \hat{\alpha} = len(v_1) + 1 - 1 = \lfloor \log_2(v_1) \rfloor + 1$$

The equation builds on the binary length of the starting value  $len(v_1)$ . We add one to respect the binary overflow in the final iteration. Furthermore, we subtract the binary length of the final result  $v_{n+1} = len(v_{n+1}) = 1$ . No value of  $\alpha$  can possibly exceed this maximum, since  $\hat{\alpha}$  directly leads to the terminal value  $v_{n+1} = 1$ , halting the sequence.<sup>15</sup> The above equation thus proves theorem 3.2 for  $k = 1$ . In the next section we will explain why this argumentation is in principle valid for all  $k$ .

### 3.3 Proving $\hat{\alpha}$ for $k > 1$

Let us now examine the case  $k = 3$ , which is most interesting because it relates to the original Collatz conjecture. The first question we need to address is whether or not the principles discussed in the previous paragraph are transferable to this form

<sup>14</sup> The condition is only true for  $k=1$ .

<sup>15</sup> The following notebook can be used to validate the proof experimentally:  
<https://github.com/c4ristian/collatz/blob/v1.0/notebooks/lambda.ipynb>

of the problem. To find an answer, we analyse a sequence, starting with  $v_1 = 17$  and  $k = 3$ . The results are displayed in the following table.

n	Variable	Decimal	Log2	Binary	Binary Length	$\alpha_i$	$\alpha$	Operation
1	$v_1$	17	4.09	10001 <sub>2</sub>	5			* 3
	$3v_1$	51	5.67	110011 <sub>2</sub>	6			+1
	$3v_1 + 1$	52	5.70	110100 <sub>2</sub>	6	2	2	* 2 <sup>-2</sup>
2	$v_2$	13	3.70	1101 <sub>2</sub>	4			* 3
	$3v_2$	39	5.29	100111 <sub>2</sub>	6			+1
	$3v_2 + 1$	40	5.32	101000 <sub>2</sub>	6	3	5	* 2 <sup>-3</sup>
3	$v_3$	5	2.32	101 <sub>2</sub>	3			* 3
	$3v_3$	15	3.91	1111 <sub>2</sub>	4			+1
	$3v_3 + 1$	16	4.00	10000 <sub>2</sub>	5	4	9	* 2 <sup>-4</sup>
4	$v_4$	1	1.00	1 <sub>2</sub>	1			NA

Table 2: Binary representation of a Collatz sequence for  $k=3$

The example presented in Table 2 reveals that in comparison to the previous case  $k = 1$ , the algorithm performs an additional operation, which is the multiplication with three. This operation leads to a growth of the binary length when comparing  $v_n$  to  $3v_n$ . The result of the operation can be calculated as follows:

$$(3.7)$$

$$\text{len}(3 * v_n) = \lfloor \log_2(3) + \log_2(v_n) \rfloor + 1$$

In determining the maximum  $\hat{\alpha}$  for  $k = 3$ , we have to take the additional binary growth into account. With regard to the operation +1 we can utilise the same arguments as in the previous section. Whenever adding one leads to an overflow in the binary representation of  $3v_n$ , the result will be a power of two, halting the sequence. The length of  $(3v_n + 1)$  will, in this case, increase by one in contrast

to  $3v_n$ . This can happen only once in a Collatz sequence, since the resulting power of two will lead to a termination.

In order to prove our hypothesis, we have to adjust equation 3.6 by considering the additional binary growth that is caused by the multiplications with three. Therefore, we obtain the following formula:

(3.8)

$$\alpha = \hat{\alpha} = \lfloor n * \log_2(3) + \log_2(v_1) \rfloor + 1$$

The above term proves theorem 3.2 for the case  $k = 3$ . A closer look makes clear that it is not only valid for  $k = 3$ , but for all  $k$ . In conclusion, we can define the following boundaries for the number of divisions by two in a Collatz sequence:

(3.9)

$$n \leq \alpha \leq \hat{\alpha}$$

If one shows that every sequence finally leads to  $\hat{\alpha}$ , that means to a binary overflow of  $(3v_n + 1)$ , the Collatz problem would be solved. In the following we will discuss the consequences of our findings for the occurrence of cycles and further confirm our line of reasoning.

## 4. Occurrence of Cycles

### 4.1 Definition

A promising possibility to falsify the Collatz conjecture in its original form is a cycle. We have found such a counterexample if the following halting condition from section 1.3 is fulfilled:

$$v_{n+1} \in \{v_1, v_2, v_3 \dots v_n\}$$

The single known cycle for  $k = 3$  is the trivial one starting with  $v_1 = 1$ :

$$v_1 = 1 = v_{1+1} = 3 * 1 * \left(1 + \frac{1}{3 * 1}\right) * 2^{-2}$$

The Collatz conjecture claims that the above example is the only possibility of a cycle for  $k = 3$ . Based on theorem 1.3 we derive the following condition for the occurrence of a cycle within a Collatz sequence:<sup>16</sup>

(4.1)

$$2^\alpha = k^n * \prod_{i=1}^n \beta_i$$

For the convenience of the reader, the expression  $\prod_{i=1}^n \beta_i$  will be referred to as  $\beta$  subsequently. Showing that equation 4.1 is true for  $k = 3$  would partially prove the Collatz conjecture. Yet there would still remain the possibility of an eternally growing sequence. This makes theorem 3.2 particularly interesting.

A major difficulty in analysing cycles in Collatz sequences is that there seems to be just one example. This is, however, not true for our generalised form of the problem. Let us consider the example  $k = 5$  and  $v_1 = 13$ . Applying theorem 1.3 leads to a cycle of the length  $n = 3$ :

$$13 = 5^3 * 13 * \left(1 + \frac{1}{5 * 13}\right) * \left(1 + \frac{1}{5 * 33}\right) * \left(1 + \frac{1}{5 * 83}\right) * 2^{-7}$$

<sup>16</sup> (Sultanow, Koch, & Cox, 2020, p. 11)

Setting  $k = 5$  and  $v_1 = 13$  in equation 4.1, we obtain the following result after three iterations:

$$128 = 5^3 * \left(1 + \frac{1}{5 * 13}\right) * \left(1 + \frac{1}{5 * 33}\right) * \left(1 + \frac{1}{5 * 83}\right)$$

To determine the number of divisions by two, which can lead to a cycle, we need to investigate the parameter  $\beta$  more thoroughly.

## 4.2 Analysing $\beta$

The starting point of our analysis of  $\beta$  is theorem 3.2. The formula can be used to calculate the maximum possible divisions by two of a Collatz sequence:

$$\hat{\alpha} = \lfloor n * \log_2(k) + \log_2(v_1) \rfloor + 1$$

In section 3 we showed that  $\hat{\alpha}$  relates to the binary length of the starting value  $v_1$ . Furthermore, the equation accounts for the binary growth, caused by the  $n$ -fold multiplication with  $k$  as well as the final overflow, triggered by the operation  $+1$ . If a sequence reaches  $\hat{\alpha}$ , it halts at the terminal value  $v_{n+1} = 1$ .

In order to learn more about the parameter  $\beta$ , we take a look at the relation between theorems 3.2 and 1.3. We examine the situation in which theorem 1.3 leads to the final result one. Consequently, we set  $v_{n+1} = 1$  and  $\alpha = \hat{\alpha}$ :

(4.2)

$$1 = k^n * v_1 * \prod_{i=1}^n \beta_i * 2^{-\hat{\alpha}}$$

$$1 = k^n * v_1 * \beta * 2^{-\hat{\alpha}}$$

$$2^{\hat{\alpha}} = k^n * v_1 * \beta$$

$$\hat{\alpha} = n * \log_2(k) + \log_2(v_1) + \log_2(\beta)$$

$$\lfloor n * \log_2(k) + \log_2(v_1) \rfloor + 1 = n * \log_2(k) + \log_2(v_1) + \log_2(\beta)$$

$$\log_2(\beta) = -n * \log_2(k) - \log_2(v_1) + \lfloor n * \log_2(k) + \log_2(v_1) \rfloor + 1$$



For a better understanding of the above term, let us examine two examples. We begin with the border case where  $k = 1$  and  $v_1 = 1$ . Here, equation 4.2 leads to:

$$\log_2(\beta) = 1 = -n * \log_2(1) - \log_2(1) + [n * \log_2(1) + \log_2(1)] + 1$$

$$\beta = 2$$

Moreover, we study the example where  $k = 5$ ,  $v_1 = 19$  and  $n = 2$ . Equation 4.2 in this case results in:

$$\log_2(\beta) = 0.1082 = -2 * \log_2(5) - \log_2(19) + [2 * \log_2(5) + \log_2(19)] + 1$$

$$\beta = 1.0780$$

Based on equation 4.2 and the fact that  $\beta$  must always be greater than one, we define the following boundaries of  $\beta$ :

(4.3)

$$0 < \log_2(\beta) \leq 1$$

$$1 < \beta \leq 2$$

In the next section we will discuss the implications of theorem 4.3 on the occurrence of cycles.

### 4.3 Analysing $\bar{\alpha}$

How many divisions by two can lead to a cycle within a Collatz sequence? We can derive an equation for this number, subsequently called  $\bar{\alpha}$ , on the basis of theorems 1.3 and 4.3. Therefore, we examine the case in which theorem 1.3 leads to a cycle by setting  $v_{n+1} = v_1$ :

(4.4)

$$v_1 = k^n * v_1 * \beta * 2^{-\bar{\alpha}}$$

$$2^{\bar{\alpha}} = k^n * \beta$$

$$\bar{\alpha} = n * \log_2(k) + \log_2(\beta)$$

$$\bar{\alpha} = [n * \log_2(k)] + 1$$

The last transformation above is applied since  $\bar{\alpha}$  is a whole number.<sup>17</sup> Now that it is clear that  $1 < \beta \leq 2$ , we truncate the fractional part of  $n * \log_2(k)$  and add one to the result. In a Collatz sequence a cycle can only occur if the number of divisions by two equals  $\bar{\alpha}$ . Conversely, this does not imply that reaching  $\bar{\alpha}$  inevitably leads to a cycle. The following example demonstrates this. Let us consider the case where  $k = 3$ ,  $v_1 = 83$  and  $n = 3$ . Here, theorems 4.4 and 1.3 yield the following result:

$$\bar{\alpha} = 5 = \lfloor 3 * \log_2(3) \rfloor + 1$$

$$71 = 3^3 * 83 * \left(1 + \frac{1}{3 * 83}\right) * \left(1 + \frac{1}{3 * 125}\right) * \left(1 + \frac{1}{3 * 47}\right) * 2^{-5}$$

Before we continue, we will empirically validate theorem 4.4. Our tool is a linear search performed by a Python script. For details on the program see A.2. With the script we searched and evaluated cycles in Collatz sequences for the odd starting numbers  $v_1 \in \{1, 3, 5, \dots, 9999\}$  and  $k \in \{1, 3, 5, \dots, 999\}$ . In order to restrict the runtime of the program we limited the length of the investigated cycles to  $n = 100$ . The results of our empirical validation are shown in the following table.

$k$	$v_1 \cdot v_n$	$n$	$\alpha$	$\bar{\alpha}$
1	(1)	1	1	1
3	(1)	1	2	2
5	(1, 3)	2	5	5
5	(13, 33, 83)	3	7	7
5	(27, 17, 43)	3	7	7
7	(1)	1	3	3
15	(1)	1	4	4
31	(1)	1	5	5
63	(1)	1	6	6
127	(1)	1	7	7
181	(27, 611)	2	15	15
181	(35, 99)	2	15	15
255	(1)	1	8	8
511	(1)	1	9	9

Table 3: Cycles in Collatz sequences

<sup>17</sup> Accordingly, the sum of the mantissas of  $n * \log_2(k) + \log_2(\beta)$  must be one.

As one can see in Table 3, we found several cycles for our generalised form of the Collatz problem. All of which comply with theorem 4.4.<sup>18</sup>

#### 4.4 Binary growth

As we have emphasised at several points in this paper, theorem 3.2 builds on the binary length of the starting value  $len(v_1)$ . Furthermore, it accounts for the maximum binary growth, henceforth denoted with  $\widehat{\Lambda}$ . We define binary growth as the total number of digits by which the binary length of  $v_1$  increases in a sequence.<sup>19</sup> In order to reach the final result  $v_{n+1} = len(v_{n+1}) = 1$ , we have to subtract  $\widehat{\alpha}$  from the sum of the binary length of  $v_1$  and the binary growth:

(4.5)

$$1 = len(v_1) + \widehat{\Lambda} - \widehat{\alpha}$$

$$\widehat{\Lambda} = \widehat{\alpha} + 1 - len(v_1)$$

$$\widehat{\Lambda} = [n * \log_2(k) + \log_2(v_1)] + 1 + 1 - [\log_2(v_1)] - 1$$

$$\widehat{\Lambda} = [n * \log_2(k) + \log_2(v_1)] + 1 - [\log_2(v_1)]$$

$$[n * \log_2(k)] + 1 \leq \widehat{\Lambda} \leq [n * \log_2(k)] + 2$$

In the final step the above equation is condensed by subtracting the starting value  $v_1$ . As a result, we obtain a range for  $\widehat{\Lambda}$ . The reason is a possible overflow which can be instigated by the expression  $n * \log_2(k) + \log_2(v_1)$ . Let us examine two examples to illustrate this. Starting with the case  $k = 3$ ,  $v_1 = 13$  and  $n = 2$  we find that the result is equal to the lower limit of  $\widehat{\Lambda}$ :

$$\widehat{\Lambda} = [2 * \log_2(3) + \log_2(13)] + 1 - [\log_2(13)] = [2 * \log_2(3)] + 1 = 4$$

Setting  $v_1 = 7$ ,  $k = 3$  and  $n = 5$  leads to the upper limit of the variable:

$$\widehat{\Lambda} = [5 * \log_2(3) + \log_2(7)] + 1 - [\log_2(7)] = [5 * \log_2(3)] + 2 = 9$$

The parameter  $\widehat{\Lambda}$  represents the maximum binary growth of a Collatz sequence. In other words, the binary growth of a sequence cannot exceed  $\widehat{\Lambda}$ , even if we would

<sup>18</sup> Own empirical analysis, see A.2 for details.

<sup>19</sup> This means that  $\widehat{\Lambda}$  does not account for the divisions by two that reduce the binary length of  $v_1$ .

not perform any divisions by two. Examining theorem 4.5, it is not surprising that we find the following relation to theorem 4.4:

$$\bar{\alpha} = [n * \log_2(k)] + 1 \leq \widehat{\Lambda}$$

As we know, a cycle occurs in a Collatz sequence when the condition  $v_1 = v_{n+1}$  is fulfilled. The binary length of the starting number  $v_1$ , must therefore grow exactly as much as it is reduced by the divisions by two. Thus, for a cycle to occur, the number of divisions by two has to be equal to the binary growth.

One might argue that this reasoning is erroneous since a sequence does not necessarily reach the maximum binary growth. We build on theorem 1.3 to show that our arguments are valid. By setting  $v_{n+1}=v_1$  we examine the case where the growth of the binary length of a sequence is neutralised by the divisions by two:

$$v_1 = k^n * v_1 * \beta * 2^{-\alpha}$$

$$2^\alpha = k^n * \beta$$

$$\alpha = n * \log_2(k) + \log_2(\beta)$$

Knowing that  $1 < \beta \leq 2$ , we derive the following limits for the binary growth of a cycle, subsequently called  $\bar{\Lambda}$ :

(4.6)

$$n * \log_2(k) < \bar{\Lambda} \leq [n * \log_2(k)] + 1$$

The binary growth of every Collatz sequence that leads to a cycle must lie within these boundaries. Due to the fact that  $\bar{\alpha}$  is a whole number, it is obvious that it must equal the maximum on the right side of the expression. For all other cases a cycle is impossible.

## 5. Summary

In our paper we have shed light on a central aspect of the Collatz conjecture: the divisions by two. We analysed the problem in its original form  $3v + 1$  as well as in the generalised variant  $kv + 1$ . Based on mathematical reasoning and empirical studies we derived and proved theorems on the occurrence of cycles and the termination of sequences. Our reasoning primarily builds on the binary representation of Collatz numbers and the underlying operations. Theorem 4.4 determines the number of divisions by two that can lead to a cycle. The theorem is based on the simple truth that a cycle can only occur if the binary growth of a sequence is exactly neutralised by the divisions by two. Theorem 3.2. determines the maximum number of divisions by two that can be performed in a sequence. If one could show that every starting number finally leads to this maximum, the Collatz problem would be solved. We are convinced that a profound study of the binary mechanics of Collatz sequences will lead to this proof.

## A. Appendix

### A.1 Data Set

This empirical data set was used to derive and validate theorems 2.2 and 3.2. The sample was generated with a Python script and comprises information about sequences for the odd starting numbers  $v_1 \in \{1, 3, 5, \dots, 3999\}$  and  $k \in \{1, 3, 5, 7, 9\}$ .<sup>20</sup> Since we do not know that all generated sequences halt, we limited the number of iterations per sequence to  $n = 100$ . In total, the sample contains 651,159 Collatz numbers, which are not necessarily distinct. This is due to the fact that different starting numbers can lead to the same subsequent values. For example, both starting values,  $v_1 = 13$  and  $v_1 = 53$  result in the number five.

### A.2 Cycle Finder

This Python script was used to validate theorem 4.4.<sup>21</sup> The program performs a linear search for the odd starting numbers  $v_1 \in \{1, 3, 5, \dots, 9999\}$  and  $k \in \{1, 3, 5, \dots, 999\}$ . To restrict the runtime of the script, we limited the length of the investigated cycles to  $n = 100$ . Furthermore, the results are not persisted. In order to reproduce our findings, the program must be executed again.

### A.3 Scientific Approach

The contents published in this paper have been achieved with an interdisciplinary approach. Not surprising, we applied classic mathematical theory and reasoning. Since we are convinced that the Collatz problem cannot be solved with classical maths alone, we additionally used techniques of data science. We combined the two fields in different ways. On one hand, we analysed sequences and related features empirically, in order to derive new formulas and theorems. On the other hand, we used data science to validate our proofs. As suggested by Karl Popper, we tried to falsify them with counterexamples. In the course of our work, we have learned that the combination of the two fields leads to a very efficient working mode. This might be the topic of another paper.

---

<sup>20</sup> [https://github.com/c4ristian/collatz/blob/v1.0/run\\_alpha\\_export.py](https://github.com/c4ristian/collatz/blob/v1.0/run_alpha_export.py)

<sup>21</sup> [https://github.com/c4ristian/collatz/blob/v1.0/run\\_cycle\\_finder.py](https://github.com/c4ristian/collatz/blob/v1.0/run_cycle_finder.py)

## A.4 Glossary of Notations

Notation	Description
$v_1$	First odd number of a Collatz sequence, also referred to as starting value
$v_i$	Odd number that is the result of a particular iteration. In the first iteration this is the starting value $v_1$
$k$	Factor that is multiplied with odd numbers; equals three in the original Collatz conjecture
$n$	Count of odd numbers in a sequence, also referred to as the sequence's length
$\beta_i$	Symbolises the term $\left(1 + \frac{1}{k*v_i}\right)$
$\beta$	Product of all $\beta_i$
$\alpha_i$	Represents the number of divisions by two that are performed in a specific iteration
$\alpha$	Number of divisions by two that leads from the starting value $v_1$ to the result $v_{n+1}$
$\hat{\alpha}_i$	Maximum possible number of divisions by two in a specific iteration
$\hat{\alpha}$	Maximum possible number of divisions by two in a Collatz sequence
$\bar{\alpha}$	Number of divisions by two that is required for a cycle
$\hat{\Lambda}$	Maximum binary growth of a Collatz sequence
$\bar{\Lambda}$	Binary growth that is required for a cycle

## **B. References**

Lagarias, J. C. (2010). *The Ultimate Challenge*. Providence: RI: American Mathematical Society.

Sedgewick, R., & Wayne, K. (2011). *Algorithms*. New York: Addison-Wesley Professional.

Sultanow, E., Koch, C., & Cox, S. (2020). *Collatz Sequences in the Light of Graph Theory*. Potsdam: Universität Potsdam.  
doi:<https://doi.org/10.25932/publishup-44325>



## C. About the Authors

**Christian Koch** is a Data Architect at TeamBank AG and lecturer at the Institute of Technology (Technische Hochschule Georg Simon Ohm) in Nuremberg. He began his career as an IT-Consultant and has since developed analytical systems for various European banks and public institutions. His areas of expertise include data architecture, data science and machine learning.

[christian.koch@th-nuernberg.de](mailto:christian.koch@th-nuernberg.de)

**Eldar Sultanow** is an Architect at Capgemini. In 2015, he completed his doctoral studies at the Chair of Business Information Systems and Electronic Government at the University of Potsdam. He likes mathematics, computer science and scuba diving.

[eldar.sultanow@capgemini.com](mailto:eldar.sultanow@capgemini.com)

**Sean Cox** is an Analyst at RatPac-Dune Entertainment. He is an expert in mathematics of finance and econometrics. Previously he worked for The Blackstone Group as mathematician and analyst. In his spare time he pursues outdoor activities.

[sean.cox@ratpacent.com](mailto:sean.cox@ratpacent.com)