Contents

1. Introduction

It is well known that the inverted Collatz sequence can be represented as a graph or a tree. Similarly, it is acknowledged that in order to prove the Collatz conjecture, one must demonstrate that this tree covers all odd natural numbers. A structured reachability analysis is hitherto not available. This paper investigates the problem from a graph theory perspective. We define a tree that consists of nodes labeled with Collatz sequence numbers. This tree will be transformed into a sub-tree that only contains odd labeled nodes. The analysis of the tree will provide new insights into the structure of Collatz sequences. We prove that cycles can only occur within a sequence under restricted conditions. Furthermore, we describe the constraints that must be met to reach the root node of the tree. These findings could form the basis for a future prove of the Collatz conjecture.

1.1 Motivation

The Collatz conjecture is a number theoretical problem, which has puzzled countless researchers using myriad approaches. Presently, there are scarcely any methodologies to describe and treat the problem from the perspective of the Algebraic Theory of Graphs. Such an approach is promising with respect to facilitating the comprehension of the Collatz sequence's "mechanics".

The current gap in research forms the motivation behind the present contribution. The authors are convinced that exploring the Collatz conjecture in an algebraic manner, relying on the findings and fundamentals of Graph Theory, will contribute to a simplification of the problem as a whole.

1.2 Related Research

The following literature study is largely based on one given by a similar earlier essay [1] which deals with the Collatz conjecture from the vantage of automata theory.

The Collatz conjecture is one of the unsolved "Million Buck Problems" [2]. When Lothar Collatz began his professorship in Hamburg in 1952, he mentioned this problem to his colleague Helmut Hasse. From 1976 to 1980, Collatz wrote several letters but missed referencing that he first proposed the problem in 1937. He introduced a function $g : \mathbb{N} \to \mathbb{N}$ as follows:

$$g(x) = \begin{cases} 3x+1 & 2 \nmid x \\ x/2 & \text{otherwise} \end{cases} \tag{1.1}$$

This function is surjective, but it is not injective (for example $g(3) = g(20)$) and thus is not reversible.

In his book "The Ultimate Challenge: The 3x+1 Problem" [3], along with his annotated bibliographies [4], [5] and other manuscripts like an earlier paper from 1985 [6], Lagarias reseached and put together different approaches from various authors intended to describe and solve the Collatz conjecture.

For the integers up to 2,367,363,789,863,971,985,761 the conjecture holds valid. For instance, see the computation history given by Kahermanes [7] that provides a timeline of the results which have already been achieved.

Inverting the Collatz sequence and constructing a Collatz tree is an approach that has been carried out by many researchers. It is well known that inverse sequences [8] arise from all functions $h \in H$, which can be composed of the two mappings $q, r : \mathbb{N} \to \mathbb{N}$ with $q : m \mapsto 2m$ and $r : m \mapsto (m-1)/3$:

$$H = \{h : \mathbb{N} \to \mathbb{N} \mid h = r^{(j)} \circ q^{(i)} \circ \ldots, i, j, h(1) \in \mathbb{N}\}$$

An argumentation that the Collatz Conjecture cannot be formally proved can be found in the work of Craig Alan Feinstein [9], who presents the position that any proof of the Collatz conjecture must have an infinite number of lines and thus no formal proof is possible. However, this statement will not be acknowledged in depth within this study.

Treating Collatz sequences in a binary system can be performed as well. For example, Ethan Akin [10] handles the Collatz sequence with natural numbers written in base 2 (using the Ring $\mathbb{Z}_2$ of two-adic integers), because divisions by 2 are easier to deal with in this method. He uses a shift map σ on $\mathbb{Z}_2$ and a map τ:

$$\sigma(x) = \begin{cases} (x-1)/2 & 2 \nmid x \\ x/2 & \text{otherwise} \end{cases} \qquad \tau(x) = \begin{cases} (3x+1)/2 & 2 \nmid x \\ x/2 & \text{otherwise} \end{cases}$$

The shift map's fundamental property is $\sigma(x)_i = x_{i+1}$, noting that $\sigma(x)_i$ is the i-th digit of $\sigma(x)$. This property can easily be comprehended by an example $x = 5 = 1010000\ldots = x_0 x_1 x_2 \ldots$, containing $\sigma(x) = 2 = 0100000\ldots$

Akin then defines a transformation $Q : \mathbb{Z}_2 \to \mathbb{Z}_2$ by $Q(x)_i = \tau^i(x)_0$ for non-negative integers i which means $Q(x)_i$ is zero if $\tau^i(x)$ is even and then it is one in any other instance. This transformation is a bijective map that defines a conjugacy between τ and σ: $Q \circ \tau = \sigma \circ Q$ and it is equivalent to the map denoted Q_∞ by Lagarias [6] and it is the inverse of the map Φ introduced by Bernstein [11]. Q can be described as follows: Let x be a 2-adic integer. The transformation result $Q(x)$ is a 2-adic integer y, so that $y_n = \tau^{(n)}(x)_0$. This means, the first bit y_0 is the parity of $x = \tau^{(0)}(x)$, which is one, if x is odd and otherwise zero. The next bit y_1 is the parity of $\tau^{(1)}(x)$, and the bit after next y_2 is parity of $\tau \circ \tau(x)$ and so on. The conjugancy $Q \circ \tau = \sigma \circ Q$ can be demonstrated by transforming the expression as follows: $(\sigma \circ Q(x))_i = Q(x)_{i+1} = \tau^{(i+1)}(x)_0 = \tau^{(i)}(\tau(x))_0 = Q(\tau(x))_i$

A simulation of the Collatz function by Turing machines has been presented by Michel [12]. He introduces Turing machines that simulate the iteration of the Collatz function, where he considers them having 3 states and 4 symbols. Michel examines both turing machines, those that never halt and those that halt on the final loop.

A function-theoretic approach to this problem has been provided by Berg and Meinardus [13], [14] as well as Gerhard Opfer [15], who consistently relies on the Berg's and Meinardus' idea. Opfer tries to prove the Collatz conjecture by determining the kernel intersection of two linear operators U, V that act on complex-valued functions. First he determined the kernel of V, and then he attempted to prove that its image by U is empty. Benne de Weger [16] contradicted Opfer's attempted proof.

Reachability Considerations based on a Collatz tree exist as well. It is well known that the inverted Collatz sequence can be represented as a graph; to be more specific, they can be depicted as a tree [17], [18]. It is acknowledged that in order to prove the Collatz conjecture, one needs to demonstrate that this tree covers all odd natural numbers.

The Stopping Time theory has been introduced by Terras [19], it has been taken up and continued, inter alia, by Silva [20] and Idowu [21]. Terras introduces another notation of the Collatz function $T(n) = (3^{X(n)}n + X(n))/2$, where $X(n) = 1$ when n is odd and $X(n) = 0$ when n is even, and defined the stopping time of n, denoted by $\chi(n)$, as the least positive k for which $T^{(k)}(n) < n$, if it exists, or otherwise it reaches infinity. Let L_i be a set of natural numbers, it is observable that the stopping time exhibits the regularity $\chi(n) = i$ for all n fulfilling $n \equiv l(mod 2^i)$, $l \in L_i$, $L_1 = \{4\}$, $L_2 = \{5\}$, $L_4 = \{3\}$, $L_5 = \{11, 23\}$, $L_7 = \{7, 15, 59\}$ and so on. As i increases, the sets L_i, including their elements, become significantly larger. Sets L_i are empty when $i \equiv l(mod 19)$ for $l = 3, 6, 9, 11, 14, 17, 19$. Additionally, the largest element of a non-empty set L_i is always less than 2^i.

Dynamical systems provide a wide basis for examining the Collatz sequence as well [22]. A dynamical system [23, p. 464] is a triple (M, G, Φ) for a set M, a group $(G, +)$ and a map $\Phi : M \times G \to M$ for which $\Phi(\cdot, 0) = id_M(\cdot)$ firstly applies and secondly $\Phi(\Phi(m, s), t) = \Phi(m, s+t)$ for all $m \in M$, $s, t \in G$. The set M is called phase space. Terence Tao [24] considers orbits of the dynamical system generated by the Collatz map (an orbit is a subset of the phase space). He proved that almost all of these orbits attain almost bounded values. To achieve this, he advanced the results of Allouche [25] and Korec [26]. Their main idea was to prove that the set of positive integers with finite stopping time has a density one, in this case the term density refers to the concept of *natural density* (also known as *asymptotic density*). It measures how large a subset of the set of natural numbers is. The natural density of a set $M \subseteq \mathbb{N}$ is defined as:

$$\lim_{n\to\infty} \frac{\#\{m \in M : m < n\}}{n}$$

In this context, the authors used the Collatz map as the map Φ. They proved that the set $\{x \in \mathbb{N} : (\exists t \in \mathbb{N})(\Phi(x, t) < x)\}$ has a natural density one.

Many other approaches exist as well. From an algebraic perspective Trümper [27] analyzes the Collatz problem in the light of an Infinite Free Semigroup. Kohl [28] generalized the problem by introducing residue class-wise affine mappings, in short rcwa mappings. A polynomial analogue of the Collatz Conjecture has been provided by Hicks et al. [29] [30] and there are also stochastical, statistical and Markov chain-based and permutation-based approaches to proving this elusive theory.

2. The Collatz Tree

2.1 The Connection between Groups and Graphs

Let (a_k) be a numerical sequence with $a_k = g^{(k)}(m)$, then a reversion produces an infinite number of sequences of reversely-written Collatz members [8].

Let S be a set containing two elements q and r, which are bijective functions over $\mathbb{Q}$:

$$\begin{aligned} q(x) &= 2x \\ r(x) &= \tfrac{1}{3}(x-1) \end{aligned} \tag{2.1}$$

Let a binary operation be the right-to-left composition of functions $q \circ r$, where $q \circ r(x) = q(r(x))$. Composing functions is an associative operation. All compositions of the bijections q and r and their inverses q^{-1} and r^{-1} are again bijective. The set, whose elements are all these compositions, is closed under that operation. It forms a free group F of rank 2 with respect to the free generating set S, where the group's binary operation $\circ$ is the function composition and the group's identity element is the identity function $id_{\mathbb{Q}} = e$. We call e an *empty string*. F consists of all expressions (strings) that can be concatenated from the generators q and r. The corresponding Cayley graph $Cay(F,S) = G$ is a regular tree whose vertices have four neighbors [31, p. 66]. A tree is called *regular* or *homogeneous* when every vertex has the same degree, in this case, $d(v) = 4$ for every vertex v in G. The Cayley graph's set of vertices is $V(G) = F$, and its set of edges is $E(G) = \left\{\{f, f \circ s\} \mid f \in F, s \in \left(S \cup S^{-1}\right) \setminus \{e\}\right\}$ [31, p. 57]. More precisely, the vertices are *labeled* by the elements (strings) of F.

In conformance with graph-theoretical precepts [32], [33], [34] we specify a subgraph H of G as a triple $(V(H), E(H), \psi_H)$ consisting of a set $V(H)$ of vertices, a set $E(H)$ of edges, and an incidence function ψ_H. The latter is, in our case, the restriction $\psi_G|_{E(H)}$ of the Cayley graph's incidence function to the set of edges that only join vertices, which are labeled by a string over alphabet $\{r, q\}$ without the inverses: $E(H) = \{\{f, f \circ s\} \mid f \in F, s \in S \setminus \{e\}\}$.

This subgraph corresponds to the monoid S^*, which is freely generated by S follows related thoughts [27] that examine the Collatz problem in terms of a free semigroup on the set S^{-1} of inverse generators. Note that this semigroup is not to be confused with an *inverse semigroup* "in which every element has a unique inverse" [35, p. 26], [31, p. 22].

Let $Y^X = \{f \mid f \text{ is a map } X \to Y\}$ be the set of functions, which in category theory is referred to as the *exponential object* for any sets X, Y. The evaluation function $ev : Y^X \times X \to Y$ sends the pair (f, x) to $f(x)$. For a detailed description of this concept, see [36, p. 127], [37, p. 155], [38, p. 54] and [39, p. 188]. We define the evaluation function $ev_{S^*} : S^* \times \{1\} \to \mathbb{Q}$ that evaluates an element of S^*, id est a composition of q and r, for the given input value 1.

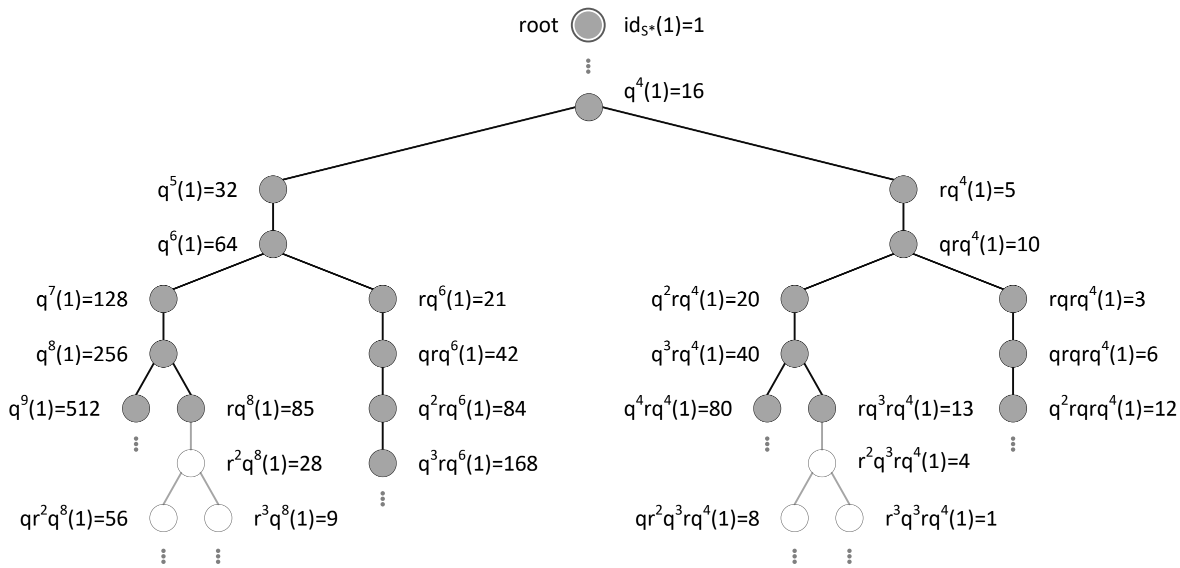

Figure 2.1: Small section of H_T with darkly highlighted subtree H_U

Furthermore we define the corestriction $ev^0_{S^*}$ of ev_{S^*} to $\mathbb{N}$. Since a corestricion of a function resricts the function's codomain [40, p. 3], the function $ev^0_{S^*}$ operates on a subset $T \subset S^*$ that contains only those compositions of q and r, which return a natural number when inputting the value 1.

The set T forms not a monoid under function composition, for example $ev_{S^*}(qrq^4, 1) = 10$ and $ev_{S^*}(rq^6, 1) = 21$, but the composition qrq^4rq^6 does not lie in T, because the evaluation $ev_{S^*}(qrq^4rq^6, 1)$ yields a value outside the codomain $\mathbb{N}$. However, each element of this set labels a vertex of a tree $H_T \subset H$, which is a proper subtree of H.

Let $U \subset T$ be a subset of T, which does not contain a reduced word with two or more successive characters r. The corresponding tree $H_U \subset H_T$ reflects Collatz sequences as demonstrated in figure 2.1.

*When talking about trees having a root ("rooted trees"), another important concept should be explained: the **level of a vertex** or often called **depth of a vertex** is the length of the path from the root to this vertex [41, p. 804]. In other words, it is the vertex's distance (the number of edges in the path) from the root. The **height of a vertex** is its level plus one $level(v) + 1 = height(v)$, see [42, p. 169].*

2.2 Defining the Tree

The starting point for specifying our tree is H_U. Due to its significance, we first concertize H_U by the definition 2.1 below, which establishes four essential characteristics.

Definition 2.1 The graph H_U possess the following key properties:

- ***H_U* is a directed graph (digraph):** Fundamentally, when we consider the more general case, an undirected graph as a triple (V, E, ψ), the incidence function maps an edge to an arbitary vertex pair $\psi : E \to \{X \subseteq V : |X| = 2\}$. In a digraph, the set $V \times V$ represents ordered vertex pairs. Accordingly the incidence function is more specifically defined, namely as a mapping of the edges to that set $\psi : E \to \{(v, w) \in V \times V : v \neq w\}$, see [43, p. 15].

- ***H_U* is a rooted tree:** According to Rosen [41, p. 747], a rooted tree is "a tree in which one vertex has been designated as the root and every edge is directed away from the root." Peculiarly, this definition considers the directionality as an inherent part of rooted trees. Unlike Mehlhorn and Sanders [44, p. 52], for example, who distinguish between an undirected and directed rooted tree.

  Note: As long as we do not stipulate that vertices may collapse, it is absolutely guaranteed that the graph is a tree.

- ***H_U* is an out-tree:** There is exactly one path from the root to every other node [44, p. 52], which means that edge directions go from parents to children [45, p. 108]. This property is implied in Rosen's definition for a rooted tree as well by saying "every edge is directed away from the root." An out-tree is sometimes designated as *out-arborescence* [45, p. 108].

- ***H_U* is a labeled tree:** For defining a labeled graph, Ehrig et al. [46, p. 23] use a label alphabet consisting of a vertex label set and an edge label set. Since we only label the vertices, in our case the specification of a vertex label set L_V together with the vertex label function $l_V : V \to L_V$ is sufficient. Originally, we said vertex labels are strings over the alphabet $S = \{q, r\}$, through which the free monoid S^* is generated. We illustrate labeling H_U by defining $l_{V(H_U)}(v) = ev^0_{S^*}(l_{V(G)}(\iota(v)), 1)$, whereby $\iota : V(H_U) \hookrightarrow V(G)$ is the inclusion map [47, p. 142] from the set of vertices of H_U to the set of vertices from the previously defined Cayley graph G.

We define a tree H_C by taking the tree H_U as a basis and for every vertex $v \in V(H_U)$ satisfying $2 \mid l_{V(H_U)}(v)$, we contract the incoming edge. We attach the label of the parent of v to the new vertex, which results by replacing (merging) the two overlapping vertices that the contracted edge used to connect. Visually, we obtain H_C by contracting all edges in H_U that have an even-labeled target vertex, which (due to contraction) gets "merged into its parent." Edge contraction is occasionally referred to as *collapsing an edge*. For more details and examples on edge contraction, one can see Voloshin [48, p. 27] and Loehr [49].

The tree H_C is a *minor of H_U*, since it can be obtained from H_U "by a sequence of any vertex deletions, edge deletions and edge contractions" [48, p. 32]. The sequence of contracting the edges between adjacent (in our case even-labeled) vertices is called *path contraction*.

A small section of the tree H_C is shown in figure 2.2. Other definitions of the same tree exist, see for example Conrow [50] or Bauer [51, p. 379].

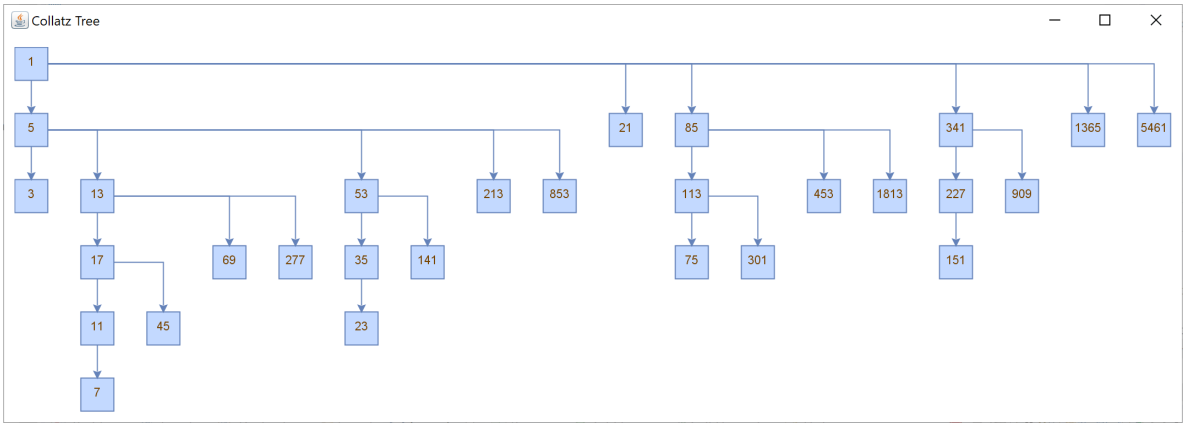

Figure 2.2: Small section of H_C (displaying the trivial cycle is waived)

2.3 Relationship of successive nodes in H_C

Let v_1 and v_{n+1} be two vertices of H_C, where v_1 is reachable from v_{n+1} with $level(v_1) - level(v_{n+1}) = n$. Hence, a path $(v_{n+1}, \ldots, v_1)$ exists between these two vertices. Theorem 2.1 specifies the following relationship between v_1 and v_{n+1}.

Theorem 2.1 $l_{V(H_C)}(v_{n+1}) = 3^n l_{V(H_C)}(v_1) \prod_{i=1}^{n} \left(1 + \frac{1}{3 l_{V(H_C)}(v_i)}\right) 2^{-\alpha_i}$. In order to simplify readability, we waive writing down the vertex label function and put it shortly: $v_{n+1} = 3^n v_1 \prod_{i=1}^{n} \left(1 + \frac{1}{3v_i}\right) 2^{-\alpha_i}$. The value $\alpha_i \in \mathbb{N}$ is the number of edges which have been contracted between v_i and v_{i+1} in H_U.

In order to demonstrate the construction produced by theorem 2.1 in an illustrative fashion, example 2.1 runs through a concrete path in H_C.

Example 2.1 For example, the two vertices $v_1 = 45$ and $v_{1+3} = v_4 = 5$ are connected via the path $(5, 13, 17, 45)$, see figure 2.2. Furthermore, one can retrace in figure 2.3 the uncontracted path between these two nodes within H_U. When applied to this example, theorem 2.1 produces the following:

$$5 = v_{1+3} = 3^3 * 45 * \left(1 + \frac{1}{3*45}\right) * 2^{-3} * \left(1 + \frac{1}{3*17}\right) * 2^{-2} * \left(1 + \frac{1}{3*13}\right) * 2^{-3}$$

Proof. This relationship of successive nodes can simply be proven inductively. For the base case, we set $n = 1$ and retrieve

$$v_{1+1} = 3v_1 \left(1 + \frac{1}{3v_1}\right) 2^{-\alpha_1} = (3v_1 + 1)\, 2^{-\alpha_1} = v_2$$

The path from v_2 to v_1 can conformly be expressed by a string $rq \cdots q$ of S^*, because of $v_1 =$

$r \circ q^{\alpha_1}(v_2)$. We set $n = n+1$ for the step case, which leads to

$$\begin{aligned} v_{n+2} &= 3^{n+1} v_1 \prod_{i=1}^{n+1} \left(1 + \tfrac{1}{3v_i}\right) 2^{-\alpha_i} \\ &= 3^{n+1} v_1 \left(1 + \tfrac{1}{3v_{n+1}}\right) 2^{-\alpha_{n+1}} \prod_{i=1}^{n} \left(1 + \tfrac{1}{3v_i}\right) 2^{-\alpha_i} \\ &= 3\left(1 + \tfrac{1}{3v_{n+1}}\right) 2^{-\alpha_{n+1}} 3^n v_1 \prod_{i=1}^{n} \left(1 + \tfrac{1}{3v_i}\right) 2^{-\alpha_i} \\ &= 3\left(1 + \tfrac{1}{3v_{n+1}}\right) 2^{-\alpha_{n+1}} v_{n+1} \\ &= (3v_{n+1} + 1)\, 2^{-\alpha_{n+1}} \end{aligned}$$

In this case the path from v_{n+2} to v_{n+1} is conformly expressable by a string $rq \cdots q$ of S^* too, since $v_{n+1} = r \circ q^{\alpha_{n+1}}(v_{n+2})$. □

Even though the tree may theoretically contain two or more identically labeled vertices, it is essential to emphasize that we only consider such paths $(v_{n+1}, \ldots, v_1)$ whose vertices are all labeled differently. Later in section 3.1, we even require that identically labeled nodes are one and the same. In order to correctly determine successive nodes using Theorem 2.1, we must consider the halting conditions. These are specified in Definition 2.2.

Definition 2.2 When determining successive nodes starting at v_1 according to Theorem 2.1, we halt if one of the following two conditions is fulfilled:

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

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

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

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

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

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

Theorem 2.1 can be used for specifying the condition of a cycle as follows:

$$\begin{aligned} v_1 &= 3^n v_1 \prod_{i=1}^{n} \left(1 + \tfrac{1}{3v_i}\right) 2^{-\alpha_i} \\ 2^{\alpha_1 + \cdots + \alpha_n} &= \prod_{i=1}^{n} \left(3 + \tfrac{1}{v_i}\right) \end{aligned} \tag{2.2}$$

A similar condition has been formulated by Hercher [52]. Taking a first look at equation 2.2, we are able to recognize the trivial cycle for $n = 1$. One might easily come to the false conclusion that the term only results in a natural number for this trivial cylce, since we are multiplying fractions. The following counterexample, starting at $v_1 = 31$, disproves this assumption:

$$20480 = \left(3+\frac{1}{31}\right)\left(3+\frac{1}{47}\right)\left(3+\frac{1}{71}\right)\left(3+\frac{1}{107}\right)\left(3+\frac{1}{161}\right)\left(3+\frac{1}{121}\right)\left(3+\frac{1}{91}\right)\left(3+\frac{1}{137}\right)\left(3+\frac{1}{103}\right)$$

According to OESIS [53], the integer $v_1 = 31$ is called *self-contained*. The term self-contained is based on the fact that the node $v_{n+1} = v_{10} = 155$ is divisible by the starting node $v_1 = 31$. Moreover, v_{10} results from applying one and the same function (in this case the Collatz function) using v_1 as input, see also Guy [54, p. 332]. For such a case equation 2.2 leads to a natural number, but not necessarily to a cycle. A cycle only occurs if the term results in a power of two. One example is the trivial cycle. We find another case when we choose the factor 5 instead of 3:

$$128 = 2^7 = \left(5+\tfrac{1}{13}\right)\left(5+\tfrac{1}{33}\right)\left(5+\tfrac{1}{83}\right)$$

The above example shows that non-trivial cycles can be found if we generalize the Collatz conjecture by replacing the factor 3 with the variable k. We study this generalized form and the occurance of cycles in chapter 3.1. A detailed elaboration of the divisibility and a deeper understanding of the tree H_C needs to be performed in order to get towards any proof of the Collatz conjecture.

2.4 Relationship of sibling nodes in H_C

In a rooted tree, vertices which have the same parent are called "siblings" [36, p. 702], [41, p. 747]. Sibling vertices accordingly have the same level.

Let w be a vertex, from which a path exists to the vertex v_1. Let v_2 be the immediate right-sibling of v_1, then $l_{V(H_C)}(v_2) = 4 * l_{V(H_C)}(v_1) + 1$. This fact has been expressed differently by Kak [18] as follows: "If an odd number a leads to another odd number (after several applications of the Collatz transformation) b, then $4a+1$ also leads to b."

Applied to our approach, consider w as the parent of v_1 and v_2. Suppose, in H_U, a path consisting of $n+1$ edges goes from w to v_1. Then we can straightforwardly show that n edges in H_U have been contracted between both nodes w and v_1 and $n+2$ edges between w and v_2 (for simplicity we again omit writing the label function):

$$v_1 = \tfrac{w*2^n-1}{3}$$
$$v_2 = \tfrac{w*2^{n+2}-1}{3} = 4*v_1+1$$

For example, $n = 3$ edges in H_U have been contracted between $w = 5$ and $v_1 = 13$ and $n+2 = 5$ edges between w and $v_2 = 53$, whereby in H_C, the vertex v_2 is the right-sibling of v_1 and these two sibling vertices are immediate children of w.

2.5 A vertex's n-fold left-child and right-sibling in H_C

Referring to the "left-child, right-sibling representation" of rooted trees [55, p. 246], the function *left-child* : $V \to V$ returns the leftmost child of a vertex v. Nesting this function n times

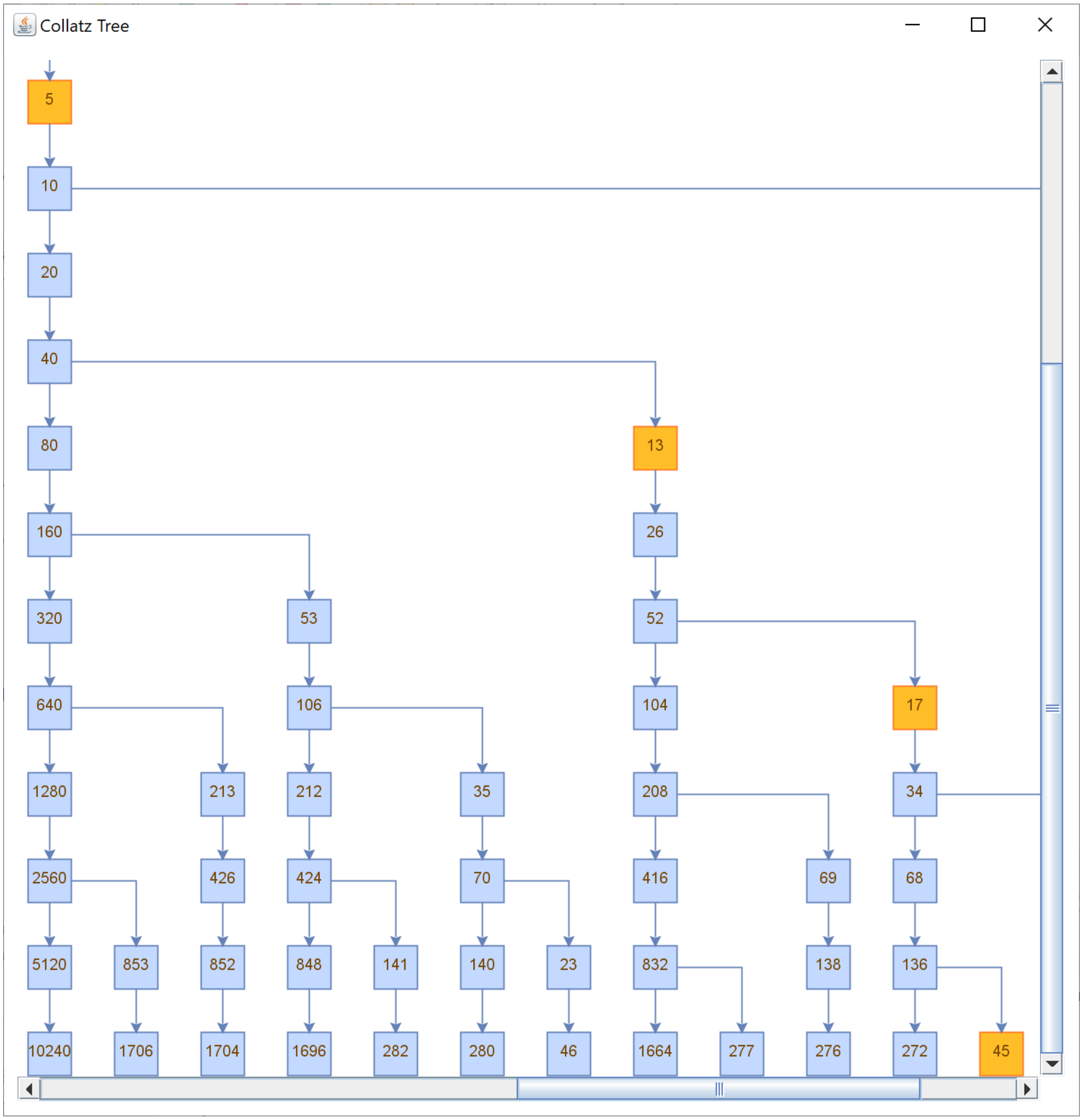

Figure 2.3: Section of H_U containing the path from 5 to 45

leads to the definition of a vertex's n-fold left-child, which is given by $\textit{left-child}^n(v)$. As shown in figure 2.2, for example $\textit{left-child}^3(13) = 7$.

The function $\textit{right-sibling} : V \to V$ points to the sibling of a vertex v immediately to its right [55, p. 246]. If this function is nested n times, we get a vertex's n-fold right-sibling defined by $\textit{right-sibling}^n(v)$. One example is $\textit{right-sibling}^2(113) = 1813$ which has been demonstrated in figure 2.2 too.

Let w be a vertex in H_C and v_0 the left-child of w. The n-fold right-sibling of v_0 can be calculated as follows:

$$v_n = \textit{right-sibling}^n(v_0) = \frac{1}{3} * \left(w * 2^{2*n+\pi_3(w \bmod 3)} - 1\right) \tag{2.3}$$

The function π_3 is the self-inverse permutation (involution):

$$\pi_3 = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \tag{2.4}$$

We consider permutations of the set $\{1,2\}$ and not of $\{0,1,2\}$, due to the fact that $w \bmod 3$ cannot be zero. A node w in H_C, which is labeled by an integer divisible by 3 is a leaf; and therefore such node has no left-child, more specifically it has no children at all.

When setting $n = 0$, we trivially retrieve the vertex's w left-child:

$$v_0 = \textit{left-child}(w) = \tfrac{1}{3} * \left(w * 2^{\pi_3(w \bmod 3)} - 1\right)$$

Example 2.2 Let us refer to figure 2.2 again and pick out $w = 5$. Then the vertex's w left-child is $v_0 = 3$ and the threefold right-sibling $v_3 = 213$:

$$v_0 = \tfrac{1}{3} * \left(5 * 2^{\pi_3(5 \bmod 3)} - 1\right) = 3$$
$$v_3 = \tfrac{1}{3} * \left(5 * 2^{2*3+\pi_3(5 \bmod 3)} - 1\right) = 213$$

2.6 Left-child and right-sibling in the $5x+1$ variant of H_C

In the following we take a look at the $5x+1$ variant of H_C. We name this graph $H_{C,5}$ and must note that it is not a tree and moreover that not all of its vertices are reachable from the root. We define the permutation π_5 as follows:

$$\pi_5 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 1 & 2 \end{pmatrix}$$

Next, by letting w be a vertex in $H_{C,5}$ and v_0 the left-child of w we obtain the n-fold right-sibling of v_0 by the function that is slightly different to the one defined by 2.3:

$$v_n = \textit{right-sibling}^n(v_0) = \frac{1}{5} * \left(w * 2^{4*n+\pi_5(w \bmod 5)} - 1\right) \tag{2.5}$$

Analogous to 2.4 only permutations on the set without zero $\{1,2,3,4\}$ need to be considered, since $w \bmod 5$ cannot be zero. Otherwise, if $w \equiv 0 (\bmod 5)$ which means that w were labeled by an integer divisible by 5, then the node w has no successor in $H_{C,5}$.

By setting $n = 0$, the function (above given by 2.5) returns the left child of w:

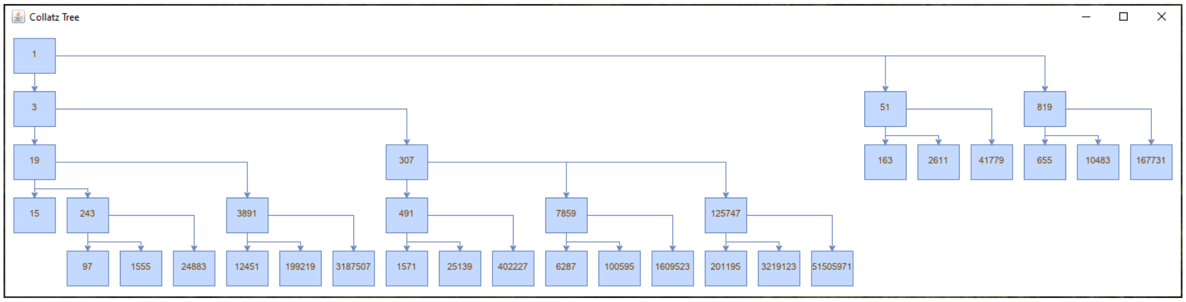

Figure 2.4: Section of the graph $H_{C,5}$ starting at its root (without branches that reflect a subsequence containing the trivial cycle)

$$v_0 = \textit{left-child}(w) = \frac{1}{5} * \left(w * 2^{\pi_5(w \bmod 5)} - 1\right)$$

Figure 2.4 illustrates a small section of $H_{C,5}$ starting at its root. The particularly interesting thing about the graph $H_{C,5}$ is that it contains three cycles, the trivial cycle starting from the root $1,3$ and two non-trivial cycles $43,17,27$ and $83,33,13$. To be precise, three cycles are known (as it will become apparent later in section 3.2), and on the basis of present knowledge it cannot be ruled out with any certainty that other cycles exist.

3. Cycles in the Collatz Tree

3.1 A remark about cycles

In graph theory, a path of length $n \geq 1$ that starts and ends at the same vertex is called a circuit. A circuit, in which no vertex is repeated with the sole exception that the initial vertex is the terminal vertex, is called a cycle. A cycle of length n is referred to as an n-cycle. For these definitions, we rely on [41, p. 599], [56, p. 35] and [57, p. 445]. Furthermore, we call a cycle originating from the root a trivial cycle.

In order for the cycles to become graphically visible, we now require that in a graph H two vertices v_1 and v_2 are one and the same if the label of both nodes are identical: $l_{V(H)}(v_1) = l_{V(H)}(v_2) \rightarrow v_1 = v_2$. As a consequence, there is no guarantee that the graph precisely refers to the algebraic structure of a free monoid anymore. A free monoid requires that each of its elements can be written in one and only one way.

When different nodes collapse on one, the graph is no longer necessarily a tree. Let us point to the monoid S*, which we introduced in section 2.1. Take for example four of its elements, the empty string e, the strings qqr, $qqrqqr$, and $qqrqqrqqr$. These elements lie as well within the subset $U \subset T \subset S^*$, and they are represented by nodes of the tree H_U that all have the same label $1 = ev_{S^*}(qqr, 1) = ev_{S^*}(qqrqqr, 1) = ev_{S^*}(qqrqqrqqr, 1)$. These nodes are one and the same, the root of H_U. Visually, then in H_U a directed edge goes from the vertex labeled with 4 back to the root node. Analogically, in H_C a loop connects the root to itself, since due to the path contraction even labeled nodes do not exist in H_C. The aforementioned example reflects the trivial cycle of the Collatz sequence.

Figure 3.1 depicts a section of $H_{C,5}$, which includes the 3-cycle $43, 17, 27$. Because of the two non-trivial cycles $43, 17, 27$ and $83, 33, 13$, in $H_{C,5}$ there does not exist a path between the root and the vertex 43 and between the root and the vertex 83. Hence, $H_{C,5}$ is said to be a disconnected graph. Generally, a graph is called a disconnected graph if it is impossible to walk (along its edges) from any vertex to any other [56, pp. 46-47].

The following considerations focus on non-trivial cycles, and therefore on cycles that do not originate from the root, but cause the graph to be a disconnected graph. Utilizing the example of the graph $H_{C,5}$ we are able to deduct from the cycle $43, 17, 27$ the simple and self-evident equality $\textit{left-child}^3(43) = 43$:

$$\textit{left-child}(43) = \tfrac{1}{5} * \left(43 * 2^1 - 1\right) = 17$$

$$\textit{left-child}(17) = \tfrac{1}{5} * \left(17 * 2^3 - 1\right) = 27$$

$$\textit{left-child}(27) = \tfrac{1}{5} * \left(27 * 2^3 - 1\right) = 43$$

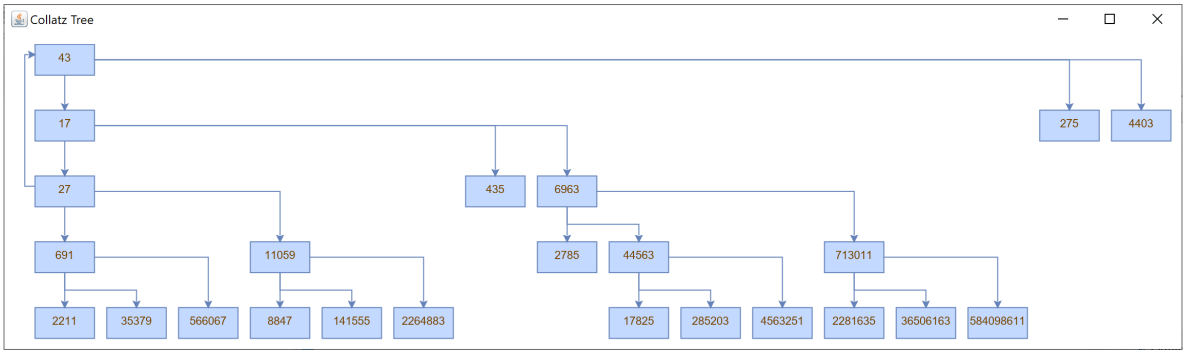

Figure 3.1: Section of $H_{C,5}$ including the 3-cycle $43, 17, 27$

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

3.2 Which variants of H_C have non-trivial cycles?

Let us refer to a $kx+1$ variant of H_C as $H_{C,k}$. By having introduced and proven theorem 2.1 we already started an assertion about the reachability of successive nodes in H_C. This reachability relationship can be generalized for any graph $H_{C,k}$ as follows:

$$v_{n+1} = k^n v_1 \prod_{i=1}^{n} \left(1 + \frac{1}{kv_i}\right) 2^{-\alpha_i} \tag{3.1}$$

This generalization leads to the condition for an existence of an n-cycle in any $kx+1$ variant of H_C, which looks analogous to the condition given by equation 2.2 that specifies H_C has a cycle:

$$2^\alpha = \prod_{i=1}^{n} \left(k + \frac{1}{v_i}\right) \tag{3.2}$$

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

$$\alpha \geq n \tag{3.3}$$

Using incremental search, one can calculate cycles through trial and error. Table 3.1 lists all empirically discovered cycles having a length up to 100 that appear in $kx+1$ variants of H_C for $k \in [1, 1000]$. Within each of these variants, the cycles have been searched at potential starting nodes v_1 with a label between 1 and 1000. Note that the cycles in table 3.1 are written in reverse order, i.e. in the order which corresponds to the Collatz sequence. To obtain the cycles in terms of graph theory referring to the graph H_C, read them from right to left.

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

Table 3.1: Known n-cycles in $kx+1$ variants of H_C for $k \leq 1000$, $n \leq 100$

Based on the results shown in table 3.1 we state the following theorem 3.1 that renders more precisely the prerequisite for cycles that may occur in variants of H_C.

Theorem 3.1 An n-cycle can only exist in a graph $H_{C,k}$, that means in a $kx+1$ variant of H_C, if the following equation holds:

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

The key of theorem 3.1 consists in the claim that, in order for an n-cycle to occur, the exponent α has to be $\bar{\alpha} = \lfloor n \log_2 k \rfloor + 1$. We approach a proof by expressing formally that $\bar{\alpha}$ is not allowed to be smaller and it is not allowed to be greater than $\lfloor n \log_2 k \rfloor + 1$, in other words we indicate a lower and an upper limit for $\bar{\alpha}$ as follows:

$$\bar{\alpha} > \lfloor n \log_2 k \rfloor \tag{3.4}$$
$$\bar{\alpha} < \lfloor n \log_2 k \rfloor + 2 \tag{3.5}$$

The validity of the first part (3.4), which specifies $\lfloor n \log_2 k \rfloor + 1$ as the lower limit for $\bar{\alpha}$, can be demonstrated in a fairly simple way: Our starting point is equation 3.1, which describes the relationship of successive vertices in $H_{C,k}$. Having a cycle, requires us to consider the first and the last vertex being one and the same $v_{n+1} = v_1$. Setting a smaller exponent $\bar{\alpha} = \lfloor n \log_2 k \rfloor$

into equation 3.1 results in the inequality $v_{n+1} > v_1$, which is in any case a true statement:

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

The validity of the second part (3.5) is not so trivial to prove. Analogous to the above-shown proof of the cylce-alpha's lower limit, we again refer to equation 3.1 as our starting point and we need to show that v_{n+1} is smaller than v_1 if $\alpha = \lfloor\ nlog_2 k\rfloor + 2$:

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

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

$$n\log_2 k - \lfloor n\log_2 k\rfloor < 2 - log_2\left(\prod_{i=1}^{n}\left(1+\frac{1}{kv_i}\right)\right) \tag{3.6}$$

A product $\prod(1+a_n)$ with positive terms a_n is convergent if the series $\sum a_n$ converges, see Knopp [58, p. 220]. Thus, to verify whether the product in condition 3.6 is converging towards a limiting value, it is sufficient to examine the following sum:

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

The sum of reciprocal vertices depending only from v_1 is given in appendix A.1.

3.3 Existence of a solitary cycle for $k = 1$

As per theorem 3.1, for $k = 1$, the only possible alpha for a cycle is 1:

$$\bar{\alpha} = \lfloor n\log_2 1\rfloor + 1 = 1$$

In accordance with the condition $\alpha \geq n$ stated by 3.3 it is clear that between two successive vertices at least one edge has been contracted or respectively one division by two took place. This is the reason why, if theorem 3.1 is true, a cycle can only occur for $n = 1$. Based on equation 3.2 we can show that this is the case for the trivial cycle, starting at the root $v_1 = 1$:

$$2^{\bar{\alpha}} = 2^{\lfloor 1\log_2 1\rfloor+1} = 2^1 = \left(1+\frac{1}{v_1}\right) = \left(1+\frac{1}{1}\right)$$

Since no other value of v_1 results in a natural number, no other cycle for $n = 1$ is possible. In order to prove theorem 3.1 for $k = 1$, we now have to show that condition 3.6 is true.

3.4 Verifying cycle-alpha's upper limit for the $1x+1$ variant of H_C

We prove that theorem 3.1 is true for $k = 1$ using the so-called Engel expansion, which we will explore more closely in appendix A.2. Setting $b = 2$ and $k = 1$ into equation A.3 leads to the formula that calculates the node v_{n+1} for a sequence, in which we divide by 2 only once per iteration:

$$v_{n+1} = \frac{v_1 + 2^n - 1}{2^n} \tag{3.7}$$

Example 3.1 Let us consider the sequence $v1 = 17$, $v_2 = 9$, $v_3 = 5$, $v_4 = 3$. Setting $v_1 = 17$ and $n = 3$ results in:

$$v_{3+1} = v_4 = \frac{17 + 2^3 - 1}{2^3} = 3$$

Equation 3.7 represents the (hypothetical) case in which a sequence progresses to the highest possible successive node for a specific starting node v_1. Actually, the sequence decreases in any case except $v_1 = 1$ and $n = 1$. We can show that setting $v_1 = 1$ and $n = 1$ results in the trivial cycle:

$$v_1 = 1 = v_2 = \frac{1 + 2^1 - 1}{2^1}$$

The equation above, complies to (and verifies) theorem 3.1, since $1 = n = \alpha = \bar{\alpha}$:

$$\bar{\alpha} = \lfloor n * log_2 1 \rfloor + 1 = 1$$

The condition 3.3, namely the inequality $\alpha \geq n$, can be used to prove that no other α than $\bar{\alpha}$ leads to a cycle. To show this, we set $v_{n+1} = v_1$:

$$v_1 = \frac{v_1 + 2^n - 1}{2^n} = \frac{v_1}{2^n} - \frac{1}{2^n} + 1 = \frac{v_1 - 1}{2^n} + 1$$

The above term is only true for $v_1 = 1$ and $n = \alpha = \bar{\alpha} = 1$. Any higher value for v_1, n or α leads to a result less than v_1. Therefore, a cycle is not possible for $\alpha \neq 1$ and theorem 3.1 is true for $k = 1$. A cycle can only occur for the case $v_1 = 1$ and $\alpha = \bar{\alpha} = n = 1$. For any other case the following condition applies:

$$v_1 > \frac{v_1 - 1}{2^n} + 1$$

Knowing that theorem 3.1 is true, we can revisit condition 3.6 determining the upper limit of $\bar{\alpha}$. We set $k = 1$ into this condition and obtain:

$$n \log_2 1 - \lfloor n \log_2 1 \rfloor < 2 - log_2\left(\prod_{i=1}^{n}\left(1 + \frac{1}{1v_i}\right)\right) \tag{3.8}$$

The above given inequality gets simplified to a condition which is true and proves that the product in condition 3.6 is always less than four:

$$4 > \prod_{i=1}^{n}\left(1 + \frac{1}{v_i}\right)$$

3.5 Verifying cycle-alpha's upper limit for the $H_{C,k>1}$

In order to prove the upper limit of $\bar{\alpha}$ for $k = 3$, we have to show that condition 3.6 is true. For a improved readability we denote the factor $(1 + \frac{1}{3v_i})$ with β_i:

$$n \log_2 3 - \lfloor n \log_2 3 \rfloor < 2 - log_2 \prod_{i=1}^{n} \beta_i$$

Having a look at the above term makes clear that to prove the inequality, we have to show that the following condition is true:

$$\prod_{i=1}^{n} \beta_i < 2 \tag{3.9}$$

We formulate a proof for condition 3.9 based on theorem 2.1:

$$v_{n+1} = 3^n v_1 \prod_{i=1}^{n} \beta_i \prod_{i=1}^{n} 2^{-\alpha_i} \tag{3.10}$$

The variable β_{n+1} can be calculated with the following equation:

$$\beta_{n+1} = 1 + \frac{1}{3v_{n+1}} \tag{3.11}$$

Example 3.2 Setting $v_1 = 5$ and $n = 1$ leads to:

$$\beta_{1+1} = 1 + \frac{1}{3v_{1+1}} = 1 + \frac{1}{3 \cdot 5} = 1.0\overline{6}$$

When we replace v_{n+1} in equation equation 3.11 with theorem 2.1, we obtain the following formula:

$$\beta_{n+1} = 1 + \frac{1}{3 \cdot 3^n v_1 \prod_{i=1}^{n} \beta_i \prod_{i=1}^{n} 2^{-\alpha_i}} = 1 + \frac{\prod_{i=1}^{n} 2^{\alpha_i}}{3^{n+1} v_1 \prod_{i=1}^{n} \beta_i} \tag{3.12}$$

Example 3.3 Let us consider $v_1 = 13$ and $n = 1$. In this case β_{n+1} equals $1.0\overline{6}$:

$$\beta_{1+1} = 1 + \frac{\prod_{i=1}^{1} 2^{\alpha_i}}{3^{1+1} v_1 \prod_{i=1}^{1} \beta_i}$$

$$\beta_{1+1} = 1 + \frac{2^3}{3^2 \cdot 13 \cdot 1.0256} = 1 + \frac{1}{3 \cdot 5} = 1.0\overline{6}$$

We now assume that the product $\prod_{i=1}^{n+1} \beta_i$ reaches the value 2 in the next iteration, which would violate the inequality 3.9:

$$2 = \prod_{i=1}^{n+1} \beta_i = \beta_{n+1} \cdot \prod_{i=1}^{n} \beta_i \tag{3.13}$$

Replacing β_{n+1} in assumption 3.13 with equation 3.12 leads to:

$$
\begin{aligned}
&2 = \left(1 + \frac{\prod_{i=1}^{n} 2^{\alpha_i}}{3^{n+1} v_1 \prod_{i=1}^{n} \beta_i}\right) \cdot \prod_{i=1}^{n} \beta_i \\
&2 = \prod_{i=1}^{n} \beta_i + \frac{\prod_{i=1}^{n} 2^{\alpha_i}}{3^{n+1} v_1} \\
&2 - \prod_{i=1}^{n} \beta_i = \frac{\prod_{i=1}^{n} 2^{\alpha_i}}{3^{n+1} v_1} \\
&3^{n+1} v_1 \cdot \left(2 - \prod_{i=1}^{n} \beta_i\right) = \prod_{i=1}^{n} 2^{\alpha_i} \\
&\prod_{i=1}^{n} 2^{\alpha_i} = 3^{n+1} v_1 \left(2 - \prod_{i=1}^{n} \beta_i\right)
\end{aligned}
\tag{3.14}
$$

We finally insert equation 3.14 into equation 3.10 and we increment the vertex's index by one:

$$
\begin{aligned}
v_{n+2} &= 3^{n+1} v_1 \prod_{i=1}^{n+1} \beta_i \prod_{i=1}^{n+1} 2^{-\alpha_i} \\
v_{n+2} &= 3^{n+1} v_1 \prod_{i=1}^{n+1} \beta_i \left(\prod_{i=1}^{n} 2^{-\alpha_i}\right) 2^{-\alpha_{n+1}} \\
v_{n+2} &= \frac{3^{n+1} v_1 \prod_{i=1}^{n+1} \beta_i}{\left(\prod_{i=1}^{n} 2^{\alpha_i}\right) 2^{\alpha_{n+1}}} = \frac{3^{n+1} \cdot v_1 \cdot 2}{3^{n+1} v_1 \left(2 - \prod_{i=1}^{n} \beta_i\right) \cdot 2^{\alpha_{i+1}}} \\
v_{n+2} &= \frac{2}{\left(2 - \prod_{i=1}^{n} \beta_i\right) \cdot 2^{\alpha_{i+1}}}
\end{aligned}
$$

Knowing that $1 < \prod_{i=1}^{n} \beta_i < 2$ and $2^{\alpha_{i+1}} \geq 2$ leads to the following true statement:

$$
v_{n+2} < 1 \tag{3.15}
$$

The statement 3.15 shows that $\prod_{i=1}^{n+1} \geq 2$ is impossible, since it would result in a final node $v_{n+2} < 1$. We therefore have proven theorem 3.1 for $k = 3$ by contradiction. Having a look at the proof makes clear that it is not only valid for $k = 3$ but for all $k > 1$. It is also applicable for $k = 1$, except for the case $v_1 = 1$. Here β_1 equals 2 immediately in the first iteration. This is the reason why we had to rely on another proof for $k = 1$.

4. Reachability of the Tree's Root Node

4.1 The neccesary alpha for reaching the root

In the previous section we have shown how many divisions by two lead to a cycle in the Collatz tree. We now study the case in which a Collatz sequence reaches the root node $v_{n+1} = 1$. Our proof builds on theorem 2.1. As in the last chapter we replace $(1 + \frac{1}{3v_i})$ with the variable β_i:

$$v_{n+1} = 3^n v_1 \prod_{i=1}^{n} \beta_i \prod_{i=1}^{n} 2^{-\alpha_i}$$

Setting $v_{n+1} = 1$ leads to:

$$\begin{aligned} 1 &= 3^n v_1 \textstyle\prod_{i=1}^{n} \beta_i \prod_{i=1}^{n} 2^{-\alpha_i} \\ \textstyle\prod_{i=1}^{n} 2^{\alpha_i} &= 3^n v_1 \textstyle\prod_{i=1}^{n} \beta_i \end{aligned} \tag{4.1}$$

Equation 4.1 defines the maximum possible value of α for a given Collatz sequence. When a Collatz sequence reaches this alpha value, it finishes at the root node. The number of divisions by two required for this is referred to as $\hat{\alpha}$ subsequently:

$$\begin{aligned} 2^{\hat{\alpha}} &= 3^n v_1 \textstyle\prod_{i=1}^{n} \beta_i \\ \hat{\alpha} &= n log_2 3 + log_2 v_1 + log_2 \textstyle\prod_{i=1}^{n} \beta_i \end{aligned}$$

In the previous chapter we proved $1 < \prod_{i=1}^{n} \beta_i < 2$. We use this knowledge to further restrict $\hat{\alpha}$ in theorem 4.1.

Theorem 4.1 The maximum possible number of divisions by two in a Collatz sequence can be calculated as follows:

$$\hat{\alpha} = \lfloor n \cdot log_2 3 + log_2 v_1 \rfloor + 1$$

If a Collatz sequence reaches $\hat{\alpha}$, it ends with the result $v_{n+1} = 1$.

Since $\hat{\alpha}$ is a whole number, we truncate the fractional part. Knowing that $1 < \prod_{i=1}^{n} \beta_i < 2$ we add one to the result.

Example 4.1 Setting $v_{n+1} = 13$ and $n = 2$ leads to:

$$v_{2+1} = 3^2 \cdot 13 \cdot \left(1 + \frac{1}{3 \cdot 13}\right) \cdot \left(1 + \frac{1}{3 \cdot 5}\right) \cdot 2^{\lfloor 2log_3 + log_2 13 \rfloor + 1} = 1$$

Building on $\hat{\alpha}$ we define the following restrictions on the alpha of a Collatz sequence:

$$n \leq \alpha \leq \hat{\alpha} \tag{4.2}$$

Condition 4.2 is not only valid for $k = 3$, but for all k. Similar to $\bar{\alpha}$, the variable $\hat{\alpha}$ could form the basis for a proof of the Collatz conjecture. As $\bar{\alpha}$ teaches us about cycles in the Collatz tree, $\hat{\alpha}$ leads us the way to its root node. If one shows that each Collatz sequence finally reaches $\hat{\alpha}$, the problem is solved as a whole. This is, however, not in the scope of this paper. It could be the foundation for a future work.

5. Conclusion and Outlook

5.1 Summary

We defined an algebraic graph structure that expresses the Collatz sequences in the form of a tree. Next, the vertex reachability properties were unveiled by examining the relationship between successive nodes in H_C. Moreover, we dealt with graphs that represent other variants of Collatz sequences, for instance $5x+1$ or $181x+1$. The interesting part of both variants just mentioned is that for these sequences the existence of cycles is known. With regard to a proof of the Collatz conjecture, theorems 3.1 and 4.1 seem promising. They serve as the basis for further investigations of the problem.

5.2 Further Research

In subsequent studies, the properties of vertices in H_C might be elaborated upon more closely by taking into account a vertex's label as well as its properties. In addition, future steps may include a detailed analysis of theorems 3.1 and 4.1.

A. Appendix

A.1 Sum of reciprocal vertices

One condition deduced from theorem 2.1 is the product condition 3.6, which specifies the validity of the cycle-alpha's upper limit. This condition requires the sum $\frac{1}{kv_1} + \frac{1}{kv_2} + \frac{1}{kv_3} + \ldots$ to be limited. In order to formulate this sum independently from the successive vertices $v_2, v_3, \ldots$, we substitute these as follows:

$$\begin{aligned}
v_1 &= v_1 \\
v_2 &= \frac{kv_1 + 1}{2^{\alpha_1}} \\
v_3 &= \frac{k^2 v_1 + k + 2^{\alpha_1}}{2^{\alpha_1+\alpha_2}} \\
v_4 &= \frac{k^3 v_1 + k^2 + k \cdot 2^{\alpha_1} + 2^{\alpha_1+\alpha_2}}{2^{\alpha_1+\alpha_2+\alpha_3}} \\
&\vdots \\
v_{n+1} &= \frac{k^n v_1 + \sum_{j=1}^{n} k^{j-1} 2^{\alpha_1+\ldots+\alpha_n - \sum_{l>n-j} \alpha_l}}{2^{\alpha_1+\ldots+\alpha_n}}
\end{aligned}$$

The sum of the reciprocal vertices can be expressed as a term that depends from v_1 and from the number of contracted edges, id est the number of dvisions by two, between two successive vertices $\alpha_1, \alpha_2, \alpha_3, \ldots$:

$$\sum_{i=1}^{n+1} \frac{1}{kv_i} = \frac{1}{k}\left(\frac{1}{v_1} + \sum_{i=1}^{n} \frac{1}{v_{i+1}}\right) = \frac{1}{k}\left(\frac{1}{v_1} + \sum_{i=1}^{n} \frac{2^{\alpha_1+\ldots+\alpha_i}}{k^i v_1 + \sum_{j=1}^{i} k^{j-1} 2^{\alpha_1+\ldots+\alpha_n - \sum_{l>i-j} \alpha_l}}\right)$$

A.2 Which sequence is a worst case?

Regarding worst case scenarios, two cases must be distinguished:

- The vertex v_{n+1} becomes a maximum. This kind of worst case we used for proving cycle-alpha's upper limit in $H_{C,1}$ with section 3.4.
- The product in condition 3.6 and consequently the sum of reciprocal vertices, formulated in A.1, becomes a maximum.

Trying to find a worst case that maximizes the product in condition 3.6 means to search for a sequence of odd numbers that rises as high as possible. One could try the ascending sequence of odd integers $v_i = 2i - 1$ (beginning at $v_1 = 1$), but will find that for this case the product will not converge against a limit value. This sequence (beginning at $v_1 = 1$) allow us to transform the product contained in condition 3.6 into a limit analyzable function using the Pochhammer's symbol (sometimes referred to as the *rising factorial* or *shifted factorial*), which is denoted by $(x)_n$ and defined as follows [59], [60, p. 679] and [61, p. 1005]:

$$(x)_n = x(x+1)(x+2)\cdots(x+n-1) = \prod_{i=0}^{n-1}(x+i) = \prod_{i=1}^{n}(x+i-1) = \frac{\Gamma(x+n)}{\Gamma(x)}$$

Setting $v_i = 2i - 1$ into the product expressed by condition 3.6 and setting $x = \frac{k+1}{2k}$ into Pochhammer's symbol $(x)_n$ interestingly makes it possible for us to perform the following transformation:

$$\prod_{i=1}^{n}\left(1+\frac{1}{kv_i}\right) = \frac{\prod_{i=1}^{n}(kv_i+1)}{\prod_{i=1}^{n}kv_i} = \frac{\prod_{i=1}^{n}(k(2i-1)+1)}{k^n\prod_{i=1}^{n}(2i-1)} = \frac{2^{2n}n!}{(2n)!}\cdot\frac{\Gamma\left(\frac{k+1+2kn}{2k}\right)}{\Gamma\left(\frac{k+1}{2k}\right)} \tag{A.1}$$

Example A.1 One simple example that is easy to recalculate may be provided by choosing $k = 3$ and $n = 4$:

$$\left(1+\frac{1}{3*1}\right)\left(1+\frac{1}{3*3}\right)\left(1+\frac{1}{3*5}\right)\left(1+\frac{1}{3*7}\right) = 1,6555 = \frac{2^8*4!}{8!}\cdot\frac{\Gamma(\frac{14}{3})}{\Gamma(\frac{4}{6})}$$

The product in the numerator in equation A.1 will be transformed into a form that allows us to use the Pochhammer's symbol:

$$\prod_{i=1}^{n}((2i-1)k+1) = 2^nk^n\prod_{i=1}^{n}\frac{(2i-1)k+1}{2k} = 2^nk^n\prod_{i=1}^{n}\frac{k+1+2ki-2k}{2k} = 2^nk^n\prod_{i=1}^{n}\left(\frac{k+1}{2k}+i-1\right)$$

This product can be written now as $2^nk^n(x)_n$, whwereby $x = \frac{k+1}{2k}$:

$$\prod_{i=1}^{n}((2i-1)k+1) = 2^nk^n\frac{\Gamma\left(\frac{k+1+2kn}{2k}\right)}{\Gamma\left(\frac{k+1}{2k}\right)}$$

The product in the denominator in equation A.1 can be transformed as follows:

$$\prod_{i=1}^{n}kv_i = k^n\prod_{i=1}^{n}v_i = k^n\prod_{i=1}^{n}(2i-1) = k^n\frac{(2n)!}{2^nn!}$$

This product is divergent, it does not converge to a limiting value. Thankfully, the ascending sequence of natural odd numbers overshoots the worst-case scenario. According to this scenario we would not have contracted a single edge between two successive nodes.

A.3 Engel expansions maximize the v_{n+1}

A sequence $v_{n+1}, v_n, \dots, v_2, v_1$ describing a path in $H_{C,3}$ from v_{n+1} down to v_1 allows at most one division by 2 between two successive nodes. Dividing only once between two successive nodes, maximizes the v_{n+1}, but it does not maximize the product contained in condition 3.6. Such a sequence forms the following ascending continued fraction (cf. also [62, p. 11]):

$$v_{n+1} = \frac{3\frac{3\frac{3v_1+1}{2}+1}{2}+1}{2}\cdots = \frac{3^n v_1 + \sum_{i=0}^{n-1} 3^i 2^{n-1-i}}{2^n} = \frac{3^n(v_1+1)-2^n}{2^n} \tag{A.2}$$

The sum of the products of the powers of three and two, contained within the above term, can be simplified to the difference $3^n - 2^n$ by converting the sum expression into the form $(x-1)(1+x+x^2+\cdots+x^{n-2}+x^{n-1}) = x^n - 1$ as follows:

$$\frac{2^n}{2^n}(3-2)\sum_{i=0}^{n-1} 3^i 2^{n-1-i} = \frac{2^n}{\cancel{2^{n-1}}}\cdot\frac{3-2}{2}\sum_{i=0}^{n-1} 3^i 2^{\cancel{n-1}-i} = 2^n\left(\frac{3}{2}-1\right)\sum_{i=0}^{n-1}\left(\frac{3}{2}\right)^i = 2^n\left(\left(\frac{3}{2}\right)^n - 1\right)$$

Example A.2 A concrete example for such a sequence is $v_1 = 31$, $v_2 = 47$, $v_3 = 71$, $v_4 = 107$, $v_5 = 161$. And, to follow that example, we can calculate the label of the vertex v_5 in a straightforward way:

$$v_5 = v_{n+1} = \frac{3^4(31+1)-2^4}{2^4} = 161$$

Ascending variants of a continued fraction, such as used in equation A.2, shall not be confused with continued fractions as treated for example in [63], [64], [65]. These ascending continued fractions correspond to the so-called "Engel Expansions" [66].

As illustrated below, we can formulate the ascending continued fractions in a generalized fashion, whereas the analogy to A.2 is given by $b_1 = b_2 = b_3 = b_4 = 2$ and $a_1 = 3^0$, $a_2 = 3^1$, $a_3 = 3^2$ and $a_4 = 3^3 + 3^4 v_1$:

$$\frac{a_1 + \frac{a_2 + \frac{a_3 + \frac{a_4}{b_4}}{b_3}}{b_2}}{b_1}\cdots = \frac{a_1}{b_1} + \frac{a_2}{b_1 b_2} + \frac{a_3}{b_1 b_2 b_3} + \frac{a_4}{b_1 b_2 b_3 b_4} + \cdots$$

The generalized form of equation A.2 may be used to compute any of the above-named ascending continued fraction that has $a_i = k^{i-1}$, $b_i = b$ for $i \in \mathbb{N}$ and $a_n = k^{n-1} + k^n v_1$:

$$v_{n+1} = \frac{k^n(kv_1 - bv_1 + 1) - b^n}{b^n(k-b)} \tag{A.3}$$

Literature

[1] E. Sultanow, D. Volkov, and S. Cox. "Introducing a Finite State Machine for Processing Collatz Sequences". In: *International Journal of Pure Mathematical Sciences* 19 (2017), pp. 10–19.

[2] S. W. Williams. "Million Buck Problems". In: *National Association of Mathematicians Newsletter* 31.2 (2000), pp. 1–3.

[3] J. C. Lagarias. *The Ultimate Challenge: The 3x+1 Problem*. Providence, RI: American Mathematical Society, 2010. ISBN: 978-0821849408.

[4] J. C. Lagarias. "The 3x + 1 Problem: An Annotated Bibliography (1963-1999)". In: *ArXiv Mathematics e-prints* (2011). eprint: math/0309224v13.

[5] J. C. Lagarias. "The 3x + 1 Problem: An Annotated Bibliography, II (2000-2009)". In: *ArXiv Mathematics e-prints* (2012). eprint: math/0608208v6.

[6] J. C. Lagarias. "The 3x + 1 Problem and Its Generalizations". In: *The American Mathematical Monthly* 92.1 (1985), pp. 3–23.

[7] S. Kahermanes. *Collatz Conjecture*. Tech. rep. Math 301 Term Paper. San Francisco State University, 2011.

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

[9] C. A. Feinstein. "The Collatz 3n+1 Conjecture is Unprovable". In: *Global Journal of Science Frontier Research Mathematics and Decision Sciences* 12.8 (2012), pp. 13–15.

[10] E. Akin. "Why is the 3x + 1 Problem Hard?" In: *Chapel Hill Ergodic Theory Workshops*. Ed. by I. Assani. Vol. 356. Contemporary Mathematics. Providence, RI: American Mathematical Society, 2004, pp. 1–20. DOI: http://dx.doi.org/10.1090/conm/364.

[11] D. J. Bernstein and J. C. Lagarias. "The 3x + 1 Conjugacy Map". In: *Canadian Journal of Mathematics* 48 (1996), pp. 1154–1169.

[12] P. Michel. "Simulation of the Collatz 3x + 1 function by Turing machines". In: *ArXiv Mathematics e-prints* (2014). eprint: 1409.7322v1.

[13] L. Berg and G. Meinardus. "Functional Equations Connected With The Collatz Problem". In: *Results in Mathematics* 25.1 (1994), pp. 1–12. DOI: 10.1007/BF03323136.

[14] L. Berg and G. Meinardus. "The 3n+1 Collatz Problem and Functional Equations". In: *Rostocker Mathematisches Kolloquium*. Vol. 48. Rostock, Germany: University of Rostock, 1995, pp. 11–18.

[15] G. Opfer. "An analytic approach to the Collatz 3n + 1 Problem". In: *Hamburger Beiträge zur Angewandten Mathematik* 2011-09 (2011).

[16] B. de Weger. *Comments on Opfer's alleged proof of the 3n + 1 Conjecture*. Tech. rep. Eindhoven University of Technology, 2011.

[17] S. Andrei and C. Masalagiu. "About the Collatz conjecture". In: *Acta Informatica* 35.2 (1998), pp. 167–179. DOI: 10.1007/s002360050117.

[18] S. Kak. *Digit Characteristics in the Collatz 3n+1 Iterations*. Tech. rep. Oklahoma State University, 2014. URL: https://subhask.okstate.edu/sites/default/files/collatz4.pdf.

[19] R. Terras. "A stopping time problem on the positive integers". In: *Acta Arithmetica* 30.3 (1976), pp. 241–252.

[20] T. Oliveira e Silva. "Maximum Excursion and Stopping Time Record-Holders for the 3x + 1 Problem: Computational Results". In: *Mathematics of Computation* 68.225 (1999), pp. 371–384.

[21] M. A. Idowu. "A Novel Theoretical Framework Formulated for Information Discovery from Number System and Collatz Conjecture Data". In: *Procedia Computer Science* 61 (2015), pp. 105–111.

[22] G. J. Wirsching. *The Dynamical System Generated by the 3n+1 Function*. Springer, 1998. DOI: 10.1007/BFb0095985.

[23] G. Walz, ed. *Lexikon der Mathematik*. 2nd ed. Vol. 1. Springer, 2017. DOI: 10.1007/978-3-662-53498-4.

[24] T. Tao. "Almost all orbits of the collatz map attain almost bounded values". In: *ArXiv Mathematics e-prints* (2019). eprint: 1909.03562.

[25] J.-P. Allouche. "Sur la conjecture de "Syracuse-Kakutani-Collatz"". In: *Séminaire de Théorie des Nombres de Bordeaux* (1978), pp. 1–15. ISSN: 09895558. URL: https://www.jstor.org/stable/44166344.

[26] I. Korec. "A density estimate for the 3x+1 problem". In: *Mathematica Slovaca* 44.1 (1994), pp. 85–89. URL: http://dml.cz/dmlcz/133225.

[27] M. Trümper. "The Collatz Problem in the Light of an Infinite Free Semigroup". In: *Chinese Journal of Mathematics* (2014), pp. 105–111. DOI: http://dx.doi.org/10.1155/2014/756917.

[28] S. Kohl. "On conjugates of Collatz-type mappings". In: *International Journal of Number Theory* 4.1 (2008), pp. 117–120. DOI: http://dx.doi.org/10.1142/S1793042108001237.

[29] K. Hicks et al. "A Polynomial Analogue of the 3n + 1 Problem". In: *The American Mathematical Monthly* 115.7 (2008), pp. 615–622.

[30] B. Snapp and M. Tracy. "The Collatz Problem and Analogues". In: *Journal of Integer Sequences* 11.4 (2008).

[31] C. Löh. *Geometric Group Theory: An Introduction*. Springer, 2017. DOI: https://doi.org/10.1007/978-3-319-72254-2.

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

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

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

[35] J. Almeida. "Profinite semigroups and applications". In: *Structural Theory of Automata, Semigroups, and Universal Algebra*. Ed. by V. B. Kudryavtsev and I. G. Rosenberg. Vol. 207. NATO Science Series II: Mathematics, Physics and Chemistry. Dordrecht, Netherlands: Springer, 2005, pp. 1–45.

[36] R. Johnsonbaugh. *Discrete Mathematics*. 8th ed. Pearson, 2017. ISBN: 0-321-96468-3.

[37] S. Mac Lane and G. Birkhoff. *Algebra*. 3rd ed. AMS Chelsea Publishing, 1999. ISBN: 0821816462.

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

[39] R. Angot-Pellissier. "The Relation Between Logic, Set Theory and Topos Theory as It Is Used by Alain Badiou". In: *The Road to Universal Logic: Festschrift for the 50th Birthday of Jean-Yves Beziau*. Ed. by A. Koslow and A. Buchsbaum. Vol. 2. Birkhäuser, 2015, pp. 181–200. DOI: 10.1007/978-3-319-15368-1.

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

[41] K. H. Rosen. *Discrete Mathematics and Its Applications*. 7th ed. McGraw-Hill, 2011. ISBN: 978-0-07-338309-5.

[42] D. Makinson. *Sets, Logic and Maths for Computing*. 2nd ed. Springer, 2012. DOI: 10.1007/978-1-4471-2500-6.

[43] B. Korte and J. Vygen. *Combinatorial Optimization: Theory and Algorithms*. 6th ed. Springer, 2018. DOI: https://doi.org/10.1007/978-3-662-56039-6.

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

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

[46] H. Ehrig et al. *Fundamentals of Algebraic Graph Transformation*. Springer, 2006. DOI: 10.1007/3-540-31188-2.

[47] L. N. Childs. *A Concrete Introduction to Higher Algebra*. 3rd ed. Springer, 2006. DOI: 10.1007/978-0-387-74725-5.

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

[49] N. A. Loehr. *Combinatorics*. 2nd ed. CRC Press, 2017. ISBN: 978-1-4987-8025-4.

[50] K. Conrow. *The Structure of the Collatz Graph; A Recursive Production of the Predecessor Tree; Proof of the Collatz 3x+1 Conjecture*. 2010. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.423.3396.

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

[52] C. Hercher. "Über die Länge nicht-trivialer Collatz-Zyklen". In: *Die Wurzel* (Hefte 6 und 7 2018).

[53] The OEIS Foundation. *Self-contained numbers: odd numbers n whose Collatz sequence contains a higher multiple of n*. 2020. URL: https://oeis.org/A005184.

[54] R. K. Guy. *Unsolved Problems in Number Theory*. 3rd ed. Springer, 2004. ISBN: 978-1-4419-1928-1. DOI: 10.1007/978-0-387-26677-0.

[55] T. H. Cormen et al. *Introduction to Algorithms*. 3rd ed. The MIT Press, 2009. ISBN: 978-0-262-03384-8.

[56] A. Benjamin, G. Chartrand, and P. Zhang. *The Fascinating World of Graph Theory*. Princeton University Press, 2015. ISBN: 978-0-691-16381-9.

[57] G. Chartrand and P. Zhang. *Discrete Mathematics*. Waveland Press, Inc., 2011. ISBN: 978-1-57766-730-8.

[58] K. Knopp. *Theorie und Anwendung der Unendlichen Reihen*. 2nd ed. Springer, 1924. ISBN: 978-3-662-41730-0. DOI: 10.1007/978-3-662-41871-0.

[59] D. Zwillinger and S. Kokoska. *Standard Probability and Statistics Tables and Formulae*. Chapman & Hall/CRC, 2000. ISBN: 1-58488-059-7.

[60] Y. A. Brychkov. *Handbook of Special Functions: Derivatives, Integrals, Series and Other Formulas*. Chapman & Hall/CRC, 2008. ISBN: 978-1-58488-956-4.

[61] M. Trott. *The Mathematica GuideBook for Symbolics*. Springer Science+Business Media, Inc., 2006. ISBN: 978-0387-95020-4.

[62] T.M.M. Laarhoven. "The 3n + 1 conjecture". Eindhoven University of Technology, July 2009.

[63] C. G. Moore. *Introduction to Continued Fractions*. National Council of Teachers of Mathematics, 1964.

[64] D. Hensley. *Continued Fractions*. World Scientific Publishing, 2006. ISBN: 981-256-477-2.

[65] J. Borwe et al. *Neverending Fractions: An Introduction to Continued Fractions*. Cambridge University Press, 2014. ISBN: 978-0-521-18649-0.

[66] C. Kraaikamp and J. Wu. "On a New Continued Fraction Expansion with Non-Decreasing Partial Quotients". In: *Monatshefte für Mathematik* 143 (4 2004), pp. 285–298. DOI: 10.1007/s00605-004-0246-3.

About Our Approach

The results published in this paper have been achieved with an interdisciplinary approach. Not suprising, we applied classic mathematical theory and reasoning. Since we are convinced that the Collatz problem cannot be solved with classical maths alone, we furthermore used techniques and tools of modern data science. We combined the two fields in different ways. Firstly, we analyzed Collatz sequences and related features empirically, to derive new formulas and theorems. On the other hand, we used data science to challenge 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, however. The interested reader can find the source code of our Python scripts at

https://github.com/c4ristian/collatz

and

https://github.com/Sultanow/collatz/tree/master/Python

About Us

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

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. Christian likes programming, data science and swimming.

✉ christian.koch@th-nuernberg.de

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