The structure of the $3x + 1$ problem

Alf Kimms

Chair of Logistics and Operations Research, 
Mercator School of Management, 
University of Duisburg–Essen, 
Lotharstr. 65, 47048 Duisburg, Germany

alf.kimms@uni-due.de

Abstract

Paul Erdős said about the $3x + 1$ problem, “Mathematics is not yet ready for such problems”. And he is seemingly right. Although we cannot solve this problem either, we provide some results about its structure. The so-called Collatz graph is iteratively transformed into a sequence of graphs by making use of some hidden structure information. It turns out that the transformation of graphs corresponds to a sequence of sets of numbers. It is shown that if the union of these number sets were equal to the set of integers greater than one, the famous Collatz conjecture would be true.

Keywords: Collatz graph, graph transformation, $3x + 1$ problem, Collatz problem, Syracuse problem, Kakutani’s problem, Haase’s algorithm, Ulam’s problem, Thwaites’ problem, hailstone numbers

Mathematics Subject Classification : 11B83, 05C90

DOI: 10.5614/ejgta.2021.9.1.14

1. Basics of the $3x + 1$ Problem

The so-called $3x + 1$ problem is a problem in number theory that has been around for decades. The exact origin is reported to be obscure and seems to go back to the 1930’s, but for sure it has been known to the community since 1952 (see Lagarias, 1985, and the references therein). Up to now, the problem is considered to be unsolved, although many noble mathematicians have tried to solve it. Depending on who worked on it or where it has been addressed, the problem became known under several different names like, e.g., the Collatz conjecture, the Syracuse problem, Kakutani’s problem, Haase’s algorithm, Ulam’s problem, and the Thwaites problem.
The 3x + 1 problem can be stated in a very simple manner: Given a positive integer $x_0$, consider the following sequence of positive integers:

$$x_{i+1} = \begin{cases} 
\frac{x_i}{2}, & \text{if } x_i \text{ is even,} \\
3x_i + 1, & \text{if } x_i \text{ is odd.}
\end{cases}$$

As an example, consider $x_0 = 9$. We then get the (infinite) sequence given in Table 1. Some authors mention that the sequence of $x_i$-values behaves somewhat “randomly” (e.g. Lagarias, 1985, states that it simulates “random” behavior) which might be a reason for the missing proof. Because of that, the numbers in the sequence are sometimes called hailstone numbers because they seem to behave as erratically as hailstones in a cloud.

| $i$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | ...
|-----|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|
| $x_i$ | 9 | 28 | 14 | 7 | 22 | 11 | 34 | 17 | 52 | 26 | 13 | 40 | 20 | 10 | 5 | 16 | 8 | 4 | 2 | 1 | ...

Table 1. The sequence with initial value $x_0 = 9$

Note, the sequence has reached the value one after a finite number of iterations. Those readers who are not familiar with the problem are invited to try different $x_0$ values to check whether or not the value one is among the numbers in the sequence. Sometimes a little patience is required as for $x_0 = 27$, for example.

If there is a need to emphasize the rules that are applied to construct the sequence, we will use a notation similar to

$$9 \xrightarrow{3x+1} 28 \xrightarrow{x/2} 14 \xrightarrow{x/2} 7 \xrightarrow{3x+1} 22 \xrightarrow{x/2} \ldots \xrightarrow{x/2} 1.$$ 

Often a graph is drawn to illustrate the problem. It is constructed as follows and it is referred to as the Collatz graph: Numbers are represented by nodes. An arc points from a number $x_i$ to a number $x_{i+1}$ if and only if $x_{i+1}$ is the unique number that immediately follows $x_i$ in the sequence defined above. While this is sufficient to draw the graph, one should note that there is a very systematic way to construct the graph.

Since prime factorization is unique, for each number $x$ that is odd there is a sequence of even numbers that precedes $x$. For example, we have $5 \leftarrow 10 \leftarrow 20 \leftarrow 40 \leftarrow \ldots$ in the graph. We call this the $x$-spine and the example just given shows the 5-spine. The number $x$ of the $x$-spine will be called the head of the spine. Since we will modify this graph later on and find out that there are similar substructures in the modified graphs, we will refer to such $x$-spines more precisely as $x$–level $m$–spines, where $m$ denotes the number of modifications that were done meanwhile unless it is clear from the context what level we talk about. Initially, we are at level 0 and $5 \leftarrow 10 \leftarrow 20 \leftarrow 40 \leftarrow \ldots$ would consequently be the 5–level 0–spine. In a similar manner, we call the graphs level $m$–graphs so that the Collatz graph is the level 0–graph.

Furthermore, in the Collatz graph there is an arc from every odd number $x$ to the even number $3x + 1$, i.e., the head $x$ of the $x$–spine is connected to an even number that occurs in another spine (there is one exception: 1 connects to 4 which is in the 1–spine). In our example, 5 connects to 16. We can now arrange the $x$–spines in such a way (see Table 2) that each $x$–spine forms a column, the step $x \to x/2$ (if $x$ is even) means moving vertically upwards, and the step $x \to 3x + 1$ (if $x$
is odd) means moving horizontally leftwards to the next number. Table 2 also illustrates how we proceed through the graph given \( x_0 = 9 \) (compare Table 1).

Table 2. An illustration of (parts of) the Collatz graph (the level 0–graph)

As of May 2020 it has been verified that at least for all positive integer starting values up to \( 2^{68} \) we do indeed reach the value one (see the web sites http://sweet.ua.pt/tos/3x+1.html maintained by Thomas Oliveira e Silva and http://www.ericr.nl/wondrous/ maintained by Eric Roosendaal, respectively, for recent computational results). This leads to the so-called Collatz conjecture:

For any positive integer \( x_0 \), the corresponding sequence of \( x_i \)-values will reach the value one in a finite number of steps.

No accepted proof for this conjecture is available yet.

Trying to grasp the nature of the problem, one can draw a graph that illustrates the fixed point iteration given a starting value \( x_0 \). Figure 1, for example, shows what is going on when we start with \( x_0 = 9 \).

There exists a cycle \( 4 \rightarrow 2 \rightarrow 1 \rightarrow 4 \rightarrow \ldots \) which contains the value one. But it is not clear whether or not other cycles exist which may not contain the value one and which may lead to infinite looping. Also, it might be that a sequence diverges in the sense that an infinite number of different numbers is enumerated without getting to the value one.

Nevertheless, many insightful results have been gained. Jeffrey Lagarias eagerly collected papers on that problem and published surveys (Lagarias, 1985) and an edited volume of important papers (Lagarias, 2010) which makes him become a historian of that problem. Because his collection of literature is extremely comprehensive, we do not review here the many papers that exist and refer to the reviews published by Lagarias instead (updated versions are available online: Lagarias, 2011 and 2012). The book by Wirsching (1998) also contains a lot of material. Chamberland (2003) provides an overview of major trends. The work that comes closest to our paper is by Andaloro (2002) who investigates the connectivity of the Collatz graph.

With this paper, we contribute to the problem by revealing some properties of the structure of the problem which, to the best of our knowledge, have not been used before. The paper is
Figure 1. The fixed point iteration with $x_0 = 9$
organized as follows: Section 2 shows a systematic way to reduce the $3x + 1$ problem to a problem with a reduced set of numbers. While not only the original problem can be represented as a graph (the Collatz graph), the reduced problem can be represented by a graph, too. We show that there is a very systematic way of transforming one graph into the other. In Section 3, we will demonstrate that this systematic process can be iterated to get a sequence of graphs. While doing so, the set of numbers that are contained in these graphs is reduced from one level to the next. This allows us to reformulate the Collatz conjecture based on the numbers that are eliminated. Section 4 demonstrates how the eliminated numbers can be computed in a systematic manner. A final Section 5 concludes the paper and points to some conjectures that we have and that may inspire future work.

2. Sequences of Odds

By studying the numbers in the Collatz graph carefully, we can derive some properties that relate certain numbers to each other. This will be done in this section and we will show that the sequence of $x_i$–values can be replaced by a sequence of $y_j$–values where most of the $y_j$–numbers are odd and only the starting value $y_0$ might be even.

Case 1. Let us begin with assuming $x_i$ being an even number. Since there is a unique prime factorization, an even $x_i$ is of the form $x_i = 2^k P$ with $k \geq 1$ and $P$ odd. Thus, the sequence would evolve from $x_i$ to $x_{i+k} = P$ by repeatedly divide the incumbent number by 2:

$$x_i = 2^k P \xrightarrow{x/2} x_{i+1} \xrightarrow{x/2} \ldots \xrightarrow{x/2} x_{i+k} = P.$$  

That means that we can consider a sequence $y_j = x_i$ and $y_{j+1} = P$ in order to find out whether or not we can eventually reach the value one. As an example, consider $x_i = 40 = 2^3 \cdot 5$. We will jump to 5 then (see Table 2) and Case 1 is an obvious shortcut.

Case 2. Let us now consider an odd number $x$. In the Collatz graph $x$ connects to the even number $3x + 1$ which belongs to some spine. In that spine $3x + 1$ is preceded by $6x + 2$ which in turn is preceded by $12x + 4$. It is important to note that there cannot be any odd integer number $x'$ which connects to $6x + 2$ in the Collatz graph. This is easy to see by contradiction. Assume there is an odd integer $x'$ such that $3x' + 1 = 6x + 2$ holds. This implies $x' = 2x + \frac{1}{3}$ which contradicts to the assumption that $x'$ is an integer. On the other hand there exists an odd integer $x''$ so that $3x'' + 1 = 12x + 4$. The value of $x''$ is $x'' = 4x + 1$. We can turn this round to observe that if we face an odd integer $x_i$ such that $x_i - 1$ can be divided by 4 giving a result that is an odd integer, we can jump to $(x_i - 1)/4$ instead. The reason is that both, $x_i$ and $(x_i - 1)/4$, are connected to the very same spine so that for the task of finding out whether or not we reach the value one in the sequence, a starting point $(x_i - 1)/4$ is equivalent to $x_i$ (if $(x_i - 1)/4$ is an odd integer). This allows us to define $y_j = x_i$ and $y_{j+1} = (y_j - 1)/4$ in such circumstances. Figure 2 illustrates this case. Consider the example $x_i = 21$ which is connected to the 1–spine. Instead of 21, we will consider $(21 - 1)/4 = 5$ which is connected to the 1–spine, too (see Table 2).

In Case 3 and 4, finally, assume $x_i$ is an odd number, i.e. $x_i = 2n + 1$ with $n$ being a non–negative integer, and $(x_i - 1)/4$ is not an odd integer. In such a case, the next number in the
The structure of the $3x + 1$ problem

Alf Kimms

sequence is defined to be $x_{i+1} = 3x_i + 1$ and it is clear that $x_{i+1}$ is even because $x_{i+1} = 3(2n + 1) + 1 = 6n + 4$. Depending on $n$ two cases exist.

Case 3. $n$ is even. For example, consider $x_i = 9 = 2 \cdot 4 + 1$ ($n = 4$ is even). If $n$ is even it can be written as $n = 2n'$ with $n'$ being a non–negative integer. Thus, we have $x_i = 4n' + 1$. Recall that we have assumed that $(x_i - 1)/4$ is not odd. Consequently, $n'$ must be even. It turns out that $x_{i+1} = 6n + 4 = 12n' + 4$. $x_{i+1}$ can now be divided by 4 to get $x_{i+3} = 3n' + 1$. Since $n'$ is even, $x_{i+3}$ is odd and we can define $y_j = x_i$ and $y_{j+1} = (3y_j + 1)/4$. Compare Table 2 to check that we can take a shortcut and jump from 9 to $(3 \cdot 9 + 1)/4 = 7$. To repeat how this relates to the definition of the sequence, note that the following subsequence was considered:

\[ x_i \xrightarrow{3x+1} x_{i+1} \xrightarrow{x/2} x_{i+2} \xrightarrow{x/2} x_{i+3}. \]

Case 4. $n$ is odd. As an example, one may consider $x_i = 11 = 2 \cdot 5 + 1$ ($n = 5$ is odd). In that case $n$ is of the form $n = 2n' + 1$ with $n'$ being a non–negative integer and $x_i = 4n' + 3$. Applying the definition of the sequence we get $x_{i+1} = 6(2n' + 1) + 4 = 12n' + 10$. And $x_{i+2} = 6n' + 5$ turns out to be an odd number. With this observation, we can define $y_j = x_i$ and $y_{j+1} = (3y_j + 1)/2$. Again, we can refer to Table 2 as an example to see that once we have reached 11 we can go to $(3 \cdot 11 + 1)/2 = 17$ next. The following subsequence illustrates what we did:

\[ x_i \xrightarrow{3x+1} x_{i+1} \xrightarrow{x/2} x_{i+2}. \]
The structure of the $3x + 1$ problem  |  Alf Kimms

In summary we have proven that, given a positive integer value $y_0 = x_0$, it is sufficient to investigate the following sequence of positive integers in order to find out if the $3x + 1$ problem eventually reaches the value one in all cases:

$$y_{j+1} = \begin{cases} 
\frac{y_j}{2^k}, & \text{if } y_j \text{ is even and of the form } 2^kP (P \text{ odd}), \\
\frac{(y_j - 1)}{4}, & \text{if } y_j \text{ and } (y_j - 1)/4 \text{ are odd}, \\
\frac{(3y_j + 1)}{4}, & \text{if } y_j \text{ and } (3y_j + 1)/4 \text{ are odd and } (y_j - 1)/4 \text{ is no odd integer}, \\
\frac{(3y_j + 1)}{2}, & \text{if } y_j \text{ and } (3y_j + 1)/2 \text{ are odd and } (y_j - 1)/4 \text{ is no odd integer}.
\end{cases}$$

The question is, given any positive integer $y_0 = x_0$, does the sequence of $y_j$–values eventually reach the value one? Appendix Appendix A provides details for the odd numbers $y_j$ from 1 to 767.

Since only $y_0$ might be even, we can focus on odd numbers. The possible sequences of $y_j$–values can be represented as an infinite graph, the level 1–graph, as shown in Table 3. For each odd number $y$ for which $(y - 1)/4$ is not an odd integer, there is a set of odd numbers which result from calculating $4y + 1$ repeatedly (compare Figure 2). $9 \leftarrow 37 \leftarrow 149 \leftarrow \ldots$ is an example. These numbers are connected in monotonic order. In Table 3 you will see these numbers as columns to form a level 1–spine. The example $9 \leftarrow 37 \leftarrow 149 \leftarrow \ldots$ is the 9–level 1–spine. Each odd number $y$ where $(y - 1)/4$ is not an odd integer (such a number is the head of a level 1–spine; see Table 3) is connected to either $(3y + 1)/4$ (if this is an odd number) or $(3y + 1)/2$ (in Figure 3 you have to move horizontally to the next number in such cases). As an example, for $y_0 = 9$ we get the following sequence:

$$9 \xrightarrow{(3y+1)/4} 7 \xrightarrow{(3y+1)/2} 11 \xrightarrow{(3y+1)/2} 17 \xrightarrow{(3y+1)/4} 13 \xrightarrow{(y-1)/4} 3 \xrightarrow{(3y+1)/2} 5 \xrightarrow{(y-1)/4} 1.$$  

Table 3. The sequences of odd numbers can be represented by a graph, the level 1–graph

It might be worth to note that hidden in the definition of the sequence of $y_j$–values there is a systematic pattern for the sequence of odd numbers. Remember that each odd number $y_j$ can be represented as $y_j = 2n + 1$ where $n$ is a non–negative integer. The rule that determines what number comes next in the sequence is clearly defined above. So, let us first look at those odd numbers $y_j$ for which $(y_j - 1)/4$ is an odd integer (see Table 4). Let $n_a$ be a counter of these numbers in increasing order starting with zero. The counter $n_a$ relates to $n$ in the following way: $n = 4n_a + 2$. These numbers have the form $y_j = 8n_a + 5$ and the odd number that follows in the sequence is $y_{j+1} = 2n_a + 1$. The distance between $y_j$ and $y_{j+1}$ in terms of $n$ is $\Delta n = -(3n_a + 2)$.

In a similar manner we can study the structure of those odd numbers $y_j$ for which the subsequent number $y_{j+1}$ is calculated by the rule $(3y_j + 1)/4$ (see Table 5). Let $n_b$ be a counter of these numbers. The relation to $n$ is $n = 4n_b$. We have $y_j = 8n_b + 1$ and the subsequent number is
The structure of the 3x + 1 problem  |  Alf Kimms

| $y_j$ | 5 13 21 29 37 45 53 ... |
| $n_a$ | 0 1 2 3 4 5 6 ... |
| $n$ | 2 6 10 14 18 22 26 ... |
| $y_{j+1}$ | 1 3 5 7 9 11 13 ... |

Table 4. Odd numbers $y_j$ where the next number in the sequence is $(y_j - 1)/4$

$y_{j+1} = y_j - 2n_b$. The distance between these two numbers is $\Delta n = -n_b$.

| $y_j$ | 1 9 17 25 33 41 49 ... |
| $n_b$ | 0 1 2 3 4 5 6 ... |
| $n$ | 0 4 8 12 16 20 24 ... |
| $y_{j+1}$ | 1 7 13 19 25 31 37 ... |

Table 5. Odd numbers $y_j$ where the next number in the sequence is $(3y_j + 1)/4$

Finally, we look at odd numbers $y_j$ where the follower $y_{j+1}$ is defined to be $(3y_j + 1)/2$ (see Table 6). Let $n_c$ be a counter for those numbers so that we have $n = 2n_c + 1$. $y_j$ is of the form $y_j = 4n_c + 3$ and $y_{j+1} = y_j + 2(n_c + 1)$. The distance is $\Delta n = n_c + 1$.

| $y_j$ | 3 7 11 15 19 23 27 ... |
| $n_c$ | 0 1 2 3 4 5 6 ... |
| $n$ | 1 3 5 7 9 11 13 ... |
| $y_{j+1}$ | 5 11 17 23 29 35 41 ... |

Table 6. Odd numbers $y_j$ where the next number in the sequence is $(3y_j + 1)/2$

Figure 3 illustrates the relationships between the odd numbers and indicates which number follows a particular number.

3. Level $m$–Graphs and the Collatz Conjecture

In this section we will carefully study what the transformation of the level 0–graph to the level 1–graph, which was described in the previous section, really does. This process can then be iterated so that we will learn something about the problem that allows us to reformulate the Collatz conjecture.

To start with, let us have a look at the graph at a level $m$. It consists of level $m$–spines where the head of each spine is connected to some other spine. Figure 4 illustrates the structure of the level $m$–graph (compare Table 2 for an illustration of the level 0–graph). Note that it will turn out to be convenient to consider several copies of the number one in the graph (in the level 0–graph, for instance, we have one node $x^1 = 1$ for the number one because $x = 2$ results in $x^1 = 1$, we have a second node for the number one because $x^2 = 1$ yields $x = 4$ which means that the node
The structure of the $3x + 1$ problem  |  Alf Kimms

Figure 3. An illustration of the internal structure of the sequence of odds

for $x^2 = 1$ connects to the $x^1$–spine, we have a third node for the number one because $x^3 = 1$ connects to the $x^2 = 1$–spine, and so on).

Figure 4. Sketch of the level $m$–graph

There are certain relationships among the numbers in the level $m$–graph that can be described by distinct functions (see Figure 5) where the domains and the codomains are subsets of the set $\mathbb{N} := \{1, 2, 3, \ldots\}$ of positive integers: First of all, there is a function $s_m : \mathbb{N} \rightarrow \mathbb{N}$ that is applied whenever we move along a spine from one number to another. A second function $g_m : \mathbb{N} \rightarrow \mathbb{N}$ describes what happens when we jump from the head of a spine to the spine this head number is connected to. The head nodes of all the spines which are connected to a common spine can be related to each other by a function $f_m : \mathbb{N} \rightarrow \mathbb{N}$. And finally, we can derive a function $h_m : \mathbb{N} \rightarrow \mathbb{N}$ that describes the relationship between the last node that connects to a particular spine to the head of that spine.

For the level 0–graph $s_0$ is defined to be

$$s_0 : \mathbb{N} \rightarrow \mathbb{N}, s_0(x) = x/2,$$

and the domain of $s_0$ is the set of even integers, i.e. $\text{dom}(s_0) = \mathbb{N}^{\text{even}} := \{x \in \mathbb{N} | x \text{ is even}\}$. The
The structure of the $3x + 1$ problem  |  Alf Kimms

Function $g_0$ is defined as 

$$g_0 : \mathbb{N} \rightarrow \mathbb{N}, g_0(x) = 3x + 1,$$

the domain is the set of odd integers, i.e. $\text{dom}(g_0) = \mathbb{N} \setminus \text{dom}(s_0) = \mathbb{N}^{\text{odd}} := \{x \in \mathbb{N} \mid x \text{ is odd}\}$, and the codomain is the set of even integers, i.e. $\text{codom}(g_0) = \mathbb{N}^{\text{even}}$. By construction, a composition of $s_m$ and $g_m$ defines $f_m$ and $h_m$. To be precise,

$$f_m(x) = g_m^{-1} \circ s_m^{k_x} \circ g_m(x)$$

for a proper value $k_x \geq 1$ that may or may not depend on the argument $x$ ($g_m^{-1}$ denotes the inverse of $g_m$). The previous section taught us (compare Case 2 and Figure 2) that

$$f_0(x) = (x - 1)/4, \text{ if } (x - 1)/4 \in \mathbb{N}^{\text{odd}}.$$ 

Regarding $h_m$ we have

$$h_m(x) = s_m^{k_x} \circ g_m(x)$$

where $s_m^k = s_m \circ \ldots \circ s_m$ and $k_x \geq 0$ is appropriately chosen possibly depending on the argument $x$. For $m = 0$ we have already derived in the previous section (see Cases 3 and 4) that

$$h_0(x) = \begin{cases} 
(3x + 1)/4, & \text{if } (3x + 1)/4 \in \mathbb{N}^{\text{odd}} \text{ and } x \notin \text{dom}(f_0), \\
(3x + 1)/2, & \text{if } (3x + 1)/2 \in \mathbb{N}^{\text{odd}} \text{ and } x \notin \text{dom}(f_0).
\end{cases}$$

Note, one is a fixed point of $h_0$, i.e. $h_0(1) = 1$.

Up to here we have just recalled what we know about the Collatz graph (i.e. the level 0–graph). From that we have derived the sequences of odds in the previous section using among other things
the simple insight that each even number can be reduced to an odd number by applying the function \(s_0\) repeatedly (i.e. by moving along the spine to a smaller number). In other words, those numbers which belong to \(\text{dom}(s_0)\) could have been eliminated from further consideration. Let \(\mathcal{N}_m\) be the set of numbers that are considered at level \(m\). Initially, we have \(\mathcal{N}_0 := \mathcal{N}\). At level 1 we have \(\mathcal{N}_1 := \mathcal{N}_0 \setminus \text{dom}(s_0) = \mathcal{N}^{\text{odd}}\).

The remaining numbers (i.e. the numbers in \(\mathcal{N}_1\)) can be partially ordered and arranged in a graph (see Table 3). The way this graph is constructed can be generalized as follows: Let \(\mathcal{N}_m\) be the set of numbers that are considered at level \(m\). Given the level \(m\)–graph (and, thus, the functions \(s_m, g_m, f_m,\) and \(h_m\)) the level \((m + 1)\)–graph consists of level \((m + 1)\)–spines which are made out of those numbers that are linked by the function \(f_m\). The set \(\mathcal{N}_{m+1}\) therefore is \(\mathcal{N}_{m+1} := \mathcal{N}_m \setminus \text{dom}(s_m)\). The number to which the head of such a level \((m + 1)\)–spine is connected to in the level \((m + 1)\)–graph is determined by the function \(h_m\). Numbers to which the function \(s_m\) is applied do not show up any more in the level \((m + 1)\)–graph. Again, we refer to Tables 2 and 3 where the level 0–graph and the level 1–graph are illustrated, respectively. Additionally, Figure 6 highlights some key numbers to emphasize the underlying idea of the graph transformation. To wrap up, the resulting functions for the level \((m + 1)\)–graph are:

\[
\begin{align*}
    s_{m+1} &= f_m \\
    g_{m+1} &= h_m
\end{align*}
\]

By construction, there is a 1–spine in the level \((m + 1)\)–graph (and in all graphs with a lower level). The functions \(f_{m+1}\) and \(h_{m+1}\) can then be derived using the basic cooking rule from above, i.e.

\[
f_{m+1}(x) = g_{m+1}^{-1} \circ s_{m+1}^k \circ g_{m+1}(x) \quad \text{with } k \geq 1
\]

and

\[
h_{m+1}(x) = s_{m+1}^k \circ g_{m+1}(x) \quad \text{with } k \geq 0,
\]

while taking into account that \(\text{dom}(f_{m+1}) \subset \mathcal{N}_{m+2} = \mathcal{N}_{m+1} \setminus \text{dom}(s_{m+1}), \text{codom}(f_{m+1}) = \mathcal{N}_{m+2} \subset \mathcal{N}_{m+1}, \text{dom}(h_{m+1}) = \mathcal{N}_{m+1} \setminus (\text{dom}(s_{m+1}) \cup \text{dom}(f_{m+1})),\) and \(\text{codom}(h_{m+1}) = \mathcal{N}_{m+2} \subset \mathcal{N}_{m+1}\).

For \(m = 1\) we get

\[
\begin{align*}
    s_1(x) &= f_0(x) = (x - 1)/4, \text{ if } (x - 1)/4 \in \mathcal{N}_1 = \mathcal{N}^{\text{odd}} \\
    g_1(x) &= h_0(x) = \begin{cases} 
        (3x + 1)/4, & \text{if } (3x + 1)/4 \in \mathcal{N}_1 = \mathcal{N}^{\text{odd}} \text{ and } x \not\in \text{dom}(s_1), \\
        (3x + 1)/2, & \text{if } (3x + 1)/2 \in \mathcal{N}_1 = \mathcal{N}^{\text{odd}} \text{ and } x \not\in \text{dom}(s_1).
    \end{cases}
\end{align*}
\]

where

\[
\text{dom}(s_1) = \{x \in \mathcal{N}_1 = \mathcal{N}^{\text{odd}} | (x - 1)/4 \in \mathcal{N}_1 = \mathcal{N}^{\text{odd}}\},
\]

= \{5, 13, 21, 29, 37, 45, 53, 61, 69, 77, 85, 93, 101, \ldots\}.

We can iterate this procedure to construct a sequence of level \(m\)–graphs. The fundamental lesson to learn is that if we can prove that all numbers included in the level \((m + 1)\)–graph reduce to the value one by applying the functions \(s_{m+1}\) and \(g_{m+1}\), we can then be sure that all numbers in the set \(\bigcup_{i=0}^{m} \text{dom}(s_i)\) eventually lead to the value one as well. That means to prove the Collatz
Figure 6. Transformation of the level 0–graph into the level 1–graph
conjecture it is sufficient to show that for every positive integer \( x_0 \) greater than one there is a level \( m \) such that \( x_0 \in \text{dom}(s_m) \). In mathematical terms the Collatz conjecture can be reformulated as

\[
\forall x_0 \in \mathbb{N} \setminus \{1\}: \exists m \geq 0 : x_0 \in \text{dom}(s_m).
\]

### 4. The Domains of the Spine Functions

What remains to do is to determine the domains of the spine functions \( s_m \). We do this level by level. For \( m = 0 \) and \( m = 1 \) we do have some results already.

**Level \( m = 0 \) (see above):**

\[
\text{dom}(s_0) = \mathbb{N}^{\text{even}} = \{2, 4, 6, 8, 10, 12, 14, \ldots\}.
\]

**Level \( m = 1 \) (see above):**

\[
\text{dom}(s_0) = \{ x \in \mathbb{N}_1 = \mathbb{N}^{\text{odd}} | (x - 1)/4 \in \mathbb{N}_1 = \mathbb{N}^{\text{odd}} \},
\]

\[
= \{5, 13, 21, 29, 37, 45, 53, 61, 69, 77, 85, 93, 101, \ldots\}.
\]

For higher levels, we will need \( f_1 \). We know that \( f_1 \) is of the form

\[
 f_1(x) = g_1^{-1} \circ s_1^{k_1} \circ g_1(x)
\]

where

\[
 s_1(x) = f_0(x) = (x - 1)/4, \text{ if } (x - 1)/4 \in \mathbb{N}_1 = \mathbb{N}^{\text{odd}}
\]

\[
 g_1(x) = h_0(x) = \left\{ \begin{array}{ll}
 (3x + 1)/4, & \text{if } (3x + 1)/4 \in \mathbb{N}_1 = \mathbb{N}^{\text{odd}} \text{ and } x \not\in \text{dom}(s_1), \\
 (3x + 1)/2, & \text{if } (3x + 1)/2 \in \mathbb{N}_1 = \mathbb{N}^{\text{odd}} \text{ and } x \not\in \text{dom}(s_1).
\end{array} \right.
\]

is known from above. This is equivalent to

\[
 x = g_1^{-1} \circ s_1^{-k_1} \circ g_1(f_1(x)).
\]

It is crucial to take into account that all intermediate results must belong to \( \mathbb{N}_1 \), i.e. we must have \( g_1(f_1(x)) \in \mathbb{N}_1 \), \( s_1^{-1}(g_1(f_1(x))) \) and so on (or, which is equivalent, \( g_1(x) \in \mathbb{N}_1 \), \( s_1(g_1(x)) \) and so on).

Two cases can occur:

1. \( f_1(x) \xrightarrow{g_1} (3f_1(x) + 1)/4 \xrightarrow{s_1^{-1}} 4((3f_1(x) + 1)/4) + 1 = 3f_1(x) + 2 \xrightarrow{g_1^{-1}} (2(3f_1(x) + 2) - 1)/3 = 2f_1(x) + 1 \). That is,

\[
 f_1(x) = (x - 1)/2.
\]

2. \( f_1(x) \xrightarrow{g_1} (3f_1(x) + 1)/2 \xrightarrow{s_1^{-1}} 4((3f_1(x) + 1)/2) + 1 = 6f_1(x) + 3 \xrightarrow{s_1^{-1}} 4(6f_1(x) + 3) + 1 = 24f_1(x) + 13 \xrightarrow{g_1^{-1}} (4(24f_1(x) + 13) - 1)/3 = 32f_1(x) + 17 \). That is,

\[
 f_1(x) = (x - 17)/32.
\]
The function \( h_1 \) will be needed on higher levels, too, because \( g_2 = h_1 \) will be used. From the deduction of \( f_1 \) we can conclude the definition of \( h_1 \):

\[
h_1(x) = \begin{cases} 
g_1(x), & \text{if } g_1(x) \notin \text{dom}(s_1), \\
(3x + 1)/4 - 1, & \text{if } g_1(x) \in \text{dom}(s_1). \end{cases}
\]

Note that the number one is a fixed point of \( h_1 \), i.e. \( h_1(1) = 1 \).

Level \( m = 2 \):

Since \( s_2(x) = f_1(x) \) with \( x \in \mathbb{N}_2 \) and \( s_2(x) \in \mathbb{N}_2 \), we can determine \( \text{dom}(s_2) \) as follows:

\[
\text{dom}(s_2) = \{ x \in \mathbb{N}_2 | (3x + 1)/2 \in \mathbb{N}_1 \text{ and } (3x + 1)/2 - 1)/4 \in \mathbb{N}_1 \text{ and } (x - 1)/2 \in \mathbb{N}_2 \}
\]

\[
\cup \{ x \in \mathbb{N}_2 | (3x + 1)/4 \in \mathbb{N}_1 \text{ and } (3x + 1)/4 - 1)/4 \in \mathbb{N}_1 \text{ and } ((3x + 1)/4 - 1)/4 - 1)/4 \in \mathbb{N}_1 \text{ and } (x - 17)/32 \in \mathbb{N}_2 \}
\]

\[
= \{ 3, 7, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, \ldots \}
\]

\[
\cup \{ 49, 113, 177, 241, 305, 369, 433, 497, \ldots \}.
\]

Note that we have listed odd numbers \( x \in \mathbb{N}_2 \) where \((x - 1)/2) \text{ or } (x - 17)/32 \), respectively, is odd and we then canceled out all those \( x \)-values which do not belong to \( \text{dom}(s_2) \) to emphasize that the conditions under which a number \( x \) belongs to \( \text{dom}(s_2) \) is a bit more complex than one might expect. Some examples which show why certain numbers do not belong to \( \text{dom}(s_2) \) are given here: \( 11 \notin \text{dom}(s_2) \) because \((11 - 1)/2 = 5 \notin \text{dom}(s_1) \notin \mathbb{N}_2 \). \( 7 \notin \text{dom}(s_2) \) because \( 7 \xrightarrow{g_1} 11 \xrightarrow{s_1} (11 - 1)/4 \notin \mathbb{N}_1 \). \( 49 \notin \text{dom}(s_2) \) because \( 49 \xrightarrow{g_1} 37 \xrightarrow{s_1} 9 \xrightarrow{s_1} 2 \notin \mathbb{N}_1 \).

Further levels can be investigated in a similar manner. We stop here because we assume that there is in infinite sequence of levels to consider and the inductive construction of the domains of the functions \( s_m \) will not terminate.

5. Conclusion

The so-called Collatz conjecture states that the \( 3x + 1 \) problem generates a sequence of numbers that will reach the value one in a finite number of steps. Many people including us believe that this is true, but no proof is available yet. As many others before, we also failed to provide a complete proof and the \( 3x + 1 \) problem remains to be a mystery.

We studied the so-called Collatz graph. We suggested to transform this graph into another graph which is transformed into just another graph and so on. Each transformation reduces the numbers to be considered. We have shown the following result: if the union set of the numbers that are eliminated during the course of transformation equals the set of positive integers greater than one then the Collatz conjecture is true. Unfortunately, we assume that the sequence of graphs, which are constructed in this way, is infinite and this idea does not lead to a constructive proof. This defines an open question to be tackled during future work: (1) We conjecture that the number \( m^* \) of levels to consider is infinite. If that was false we have a proof of the Collatz conjecture readily at hand: We simply need to check if \( \text{dom}(s_0) \cup \text{dom}(s_1) \cup \ldots \cup \text{dom}(s_{m^*}) = \mathbb{N} \setminus \{1\} \).
However, it is interesting to mention that the numbers that are eliminated in each transformation step have a very systematic pattern. At level 0 we eliminate all even numbers, i.e. all integers of the form $2 + 2^n$ where $n \geq 0$ is an integer. At level 1 the eliminated numbers are $5 + 2^n$ with $n \geq 0$ being a non–negative integer. At level 2 all numbers of the form $3 + 2^4n$ and $113 + 2^7n$ are eliminated ($n \geq 0$ and integer). It might be interesting to investigate this further during future work and to work on a second open question: (2) We conjecture that the numbers which are eliminated do always have the form $p + 2^k n$ (with integer values $n \geq 0$) where $p$ is an integer (odd, for levels $m \geq 1$) and $k \geq 1$ is an integer. If that was true the question to prove the Collatz conjecture is whether or not all odd numbers can be written in that form using the specific values of $p$ and $k$ that we get from the different levels of the graph transformation process.

It should be noted by the way that the latter discussion somehow reminds of so–called obstinate numbers (see Pickover, 2005, Chapter 2, page 62): In 1848 Alphonse Armand Charles Georges Marie (a.k.a. Prince de Polignac) conjectured that every odd number greater than one is of the form $p + 2^k$ (with $p$ being a prime number and $k > 0$). This conjecture turned out to be false (127, for instance, is a counterexample).

References

[1] P.J. Andaloro, The $3x + 1$ problem and directed graphs, *Fibonacci Quarterly* 40 (2002), 43–54.

[2] M. Chamberland, *An Update on the $3x + 1$ Problem*, Working Paper, Department of Mathematics and Computer Science, Grinnell College (2003), http://www.math.grin.edu/~chamberl/papers/3x_survey_eng.pdf

[3] J.C. Lagarias, The $3x + 1$ problem and its generalizations, *The American Mathematical Monthly* 92 (1985), 3–23.

[4] J.C. Lagarias (ed.), *The ultimate Challenge: The $3x + 1$ Problem*, American Mathematical Society (2010)

[5] J.C. Lagarias, The $3x+1$ Problem: An Annotated Bibliography, (1963–1999), V13, arXiv.org (2011), arXiv:math/0309224

[6] J.C. Lagarias, The $3x + 1$ Problem: An Annotated Bibliography, II (2000–2009), V6, arXiv.org (2012), arXiv:math/0608208

[7] C.A. Pickover, *A Passion for Mathematics: Numbers, Puzzles, Madness, Religion, and the Quest for Reality*, Hoboken, Wiley (2005)

[8] G.J. Wirsching, *The Dynamical System Generated by the $3n + 1$ Function*, Heidelberg, Springer (1998)
Appendix A. The Sequences of Odds in Detail from 1 to 767

For the odd numbers $y_j$ from 1 to 767 we provide here the successors in the sequence of $y_j$–values. Tables A.7 and A.8 provide the number $y_j$, the rule that has to be applied, and the successor $y_{j+1}$. In addition the tables contain the number $\Delta j$ of steps that are required to reach the value one, some authors call $\Delta j$ the total stopping time. For instance, given $y_j = 11$, we would apply the rule $(3y + 1)/2$ next to get 17. If we proceeded, we would see that we reach the value one $\Delta j = 5$ steps later, i.e. $y_{j+5} = 1$. 
| $y_i \Delta 1 \rangle$ | $\rule{8pt}{0.4pt} (3y_i+1)/4 \langle y_{i+1}$ |
|-----------------|-----------------|
| 1               | 0 [0] (3y+1)/4  |
| 3               | 2 [2] (3y+1)/2  |
| 5               | 1 [1] (y-1)/4   |
| 7               | 6 [6] (3y+1)/2  |
| 9               | 7 [7] (3y+1)/4  |
| 11              | 5 [5] (3y+1)/2  |
| 13              | 3 [3] (y-1)/4   |
| 15              | 7 [7] (3y+1)/2  |
| 17              | 4 [4] (3y+1)/4  |
| 19              | 8 [8] (3y+1)/2  |
| 21              | 2 [2] (3y+1)/2  |
| 23              | 6 [6] (3y+1)/2  |
| 25              | 9 [9] (3y+1)/4  |
| 27              | 48 [48] (3y+1)/2 |
| 29              | 7 [7] (y-1)/4   |
| 31              | 46 [46] (3y+1)/2 |
| 33              | 10 [10] (3y+1)/4 |
| 35              | 5 [5] (3y+1)/2  |
| 37              | 8 [8] (y-1)/4   |
| 39              | 14 [14] (3y+1)/2 |
| 41              | 47 [47] (3y+1)/2 |
| 43              | 11 [11] (3y+1)/2 |
| 45              | 6 [6] (y-1)/4   |
| 47              | 45 [45] (3y+1)/2 |
| 49              | 9 [9] (3y+1)/4  |
| 51              | 10 [10] (3y+1)/2 |
| 53              | 4 [4] (y-1)/4   |
| 55              | 49 [49] (3y+1)/2 |
| 57              | 12 [12] (3y+1)/4 |
| 59              | 13 [13] (3y+1)/2 |
| 61              | 8 [8] (y-1)/4   |
| 63              | 47 [47] (3y+1)/2 |

Table A.7. Sequences of odds in greater detail (1 to 383)
Table A.8. Sequences of odds in greater detail (385 to 767)