Journal of Integer Sequences, Vol. 17 (2014),
Article 14.11.7

On Collatz Words, Sequences, and Trees

Wolfdieter Lang
Karlsruhe
Germany
wolfdieter.lang@partner.kit.edu

Abstract

Motivated by recent work of Trümper, we consider the general Collatz word (up-down pattern) and the sequences following this pattern. We derive recurrences for the first and last sequence entries from repeated application of the general solution of a binary linear inhomogeneous Diophantine equation. We solve these recurrences and also discuss the Collatz tree.

1 Introduction

The Collatz map C for natural numbers maps an odd number m to $3m+1$ and an even number to $m/2$. The Collatz conjecture (see Lagarias [4] for original references) is the claim that every natural number n ends up, after sufficiently many iterations of the map C, in the trivial cycle $(4, 2, 1)$.

Motivated by the work of Trümper [9] we consider a general finite Collatz word on the alphabet $\{u, d\}$, where u (for 'up') indicates application of the map C on an odd number, and d (for 'down') for applying the map C on an even number. The task is to find all sequences which follow this word pattern (to be read from left to right). These sequences are called CS (for Collatz sequence, also for the plural) realizing the CW (for Collatz word, also for the plural) under consideration. This problem was solved by Trümper [9] under the restriction that the first and last sequence entries are odd. Here we do not use this restriction. The solutions are given in terms of recurrence relations for the first and last entries of the CS for a given CW. This involves a repeated application of the general solution in positive integers of the linear inhomogeneous Diophantine equation $ax + by = c$, with $a = 3^m$, $b = 2^n$ and given integer $c = c(m, n)$. Because $\gcd(3, 2) = 1$, one always has a countably infinite number of solutions. This general solution depends on a non-negative integer parameter k. We believe that our solution is more straightforward than the one given by Trümper [9].

2 Collatz words, sequences and the Collatz tree

The Collatz map $C: \mathbb{N} \to \mathbb{N}$, $m \mapsto 3m+1$ if m is odd, $m \mapsto m/2$ if m is even, leads to an increase u (for 'up') or decrease d (for 'down'), respectively. Finite Collatz words over the alphabet $\{u, d\}$ are considered with the restriction that, except for the one letter word u, every u is followed by a d, because $2m+1 \mapsto 2(3m+1)$. This is the reason for also introducing (with Trümper [9]) $s := ud$. Thus s stands for $2m+1 \mapsto 3m+1$. The general finite word is encoded by an $(S+1)-$tuple $\vec{n}_S = [n_0, n_1, \ldots, n_S]$ with $S \in \mathbb{N}$.

$$\begin{aligned} CW(\vec{n}_{S+1}) &= d^{n_0} s d^{n_1-1} s \cdots s d^{n_S-1} \\ &= (d^{n_0} s)(d^{n_1-1} s) \cdots (d^{n_{S-1}-1} s) d^{n_S-1} \end{aligned} \tag{1}$$

with $n_0 \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$, $n_i \in \mathbb{N}$, for $i = 1, 2, \ldots, S$. The number of letters of u (that is of $s = ud$) in the word $CW(\vec{n}_S)$, or $CW(S)$, for short, is S (which is why we have used $\vec{n}_S$ not $\vec{n}_{S+1}$ for the $(S+1)-$tuple), and the number of letters of d is

$$D(S) := \sum_{j=0}^{S} n_j.$$

In the paper of Trümper [9] $n_0 = 0$ (start with an odd number), $y = S$ and $x = D(S)$.

Some special words are not covered by this notation: first the one-letter word u with the Collatz sequence (CS) of length two $CS(u; k) = [2k+1, 2(3k+2)]$, and $CW([n_0]) = d^{n_0}$ with the family of sequences $CS([n_0]; k) = [2^{n_0}k, 2^{n_0-1}k, \ldots, 1k]$ with $k \in \mathbb{N}_0$.

A Collatz sequence CS realizing a word $CW(\vec{n}_S)$ is of length $L = D + S + 1$, and follows the word pattern from the left to the right:

$$CS(\vec{n}_S) = [c_1, c_2, \ldots, c_L].$$

For example, $CW([1,2,1]) = dsds$ with $S = 2$, $D(2) = 2$, and length $L = 7$ with $SC_0(\vec{n}_S) = [2,1,4,2,1,4,2]$ is the first of these sequences (for non-negative integers), the one with smallest starting number c_1. In order to conform with the notation used by Trümper [9], we use $c_1 = M$ for the starting number and $c_L = N$ for the last number.

However, in the paper [9] M and N are restricted to be odd, which is not the case here. Later one can get the words with odd starting number M by choosing $n_0 = 0$. In order to also have N odd, one has to pick only the odd numbers from $SC_k([0, n_1, \ldots, n_S])$, for $k \in \mathbb{N}_0$.

In the Trümper [9] article the monoid of Collatz words, with the unit element $e = \textit{empty word}$ is treated. This is not considered in this work. Also the connection to the $3m-1$ problem is not pursued here.

The Collatz tree CT is an infinite (incomplete) ternary tree, starting with the root, the number 8 on top at level $l = 0$. Three branches, labeled L, V and R can be present. If a node (vertex) has label $n \equiv 4 \pmod 6$ the out-degree is 2 with a left edge (branch) labeled L ending in a node with label $\frac{n-1}{3}$ and a right edge (label R) ending in the node labeled $2n$.

In the other cases, $n \equiv 0, 1, 2, 3, 5 \pmod 6$, with out-degree 1, a vertical edge (label V) ends in the node labeled $2n$. The root labeled 8 stands for the trivial cycle 8 repeat$(4, 2, 1)$. See Figure 1 for CT_7 with only the first eight levels.

It may seem that this tree is left-right symmetric (disregarding the node labels), but this is no longer the case starting at level $l = 12$. At level $l = 10$ the mod 6 structure of the left and right part of CT, also taking into account the node labels, is broken for the first time, but the node labels 4 (mod 6) are still symmetric. At the next level $l = 11$ the left-right symmetry concerning the labels 4 (mod 6) is also broken, leading at level $l = 12$ to a symmetry breaking in the branch structure of the left and right part of CT. Thus at level $l = 12$ the number of nodes becomes odd for the first time: 15 nodes on the left side versus 14 nodes on the right one. See rows $l + 3$ of A127824 for the node labels of the first levels, and A005186$(l + 3)$ for the number of nodes. The number of 4 (mod 6) nodes at level l is given in A176866$(l + 4)$.

A CS is determined uniquely from its starting number M. Therefore no number can appear twice in CT, except for the numbers $1, 2, 4$ of the (hidden) trivial cycle. The Collatz conjecture is that every natural number appears in CT at some level ($1, 2$, and 4 are hidden in the root 8). A formula for $l = l(n)$ would prove the conjecture.

Reading CT from bottom to top, beginning with some number M at a certain level l, recording the edge labels up to level $l = 0$, leads to a certain L, V, R-sequence. For example, $M = 40$ at level $l = 5$ generates the length 5 sequence $[V, R, V, L, V]$. This is related to the CS starting with $M = 40$, namely $[40, 20, 10, 5, 16, 8]$, one of the realizations of the CW $dddud = d^3s$, with $S = 1$ and $\vec{n}_1 = [3, 1]$. (Later we shall see that this is the realizations with the third smallest starting number, the smaller once are 8 and 24). One has to map V and R to d and L to u. This shows that the map from a L, V, R-sequence to a CW is not one to one. The numbers $n \equiv 4 \pmod 6$ except 4 (see A016957) appear exactly in two distinct CS. For example, $64 \equiv 4 \pmod 6$ shows up in all CS starting at any vertex which descends from the bifurcation at 64, e.g., $21, 128$; $42, 256$; $84, 85, 512$; etc. The CS starting with 512 and ending in 64 corresponds to the same CW ddd as the CS starting with 84 and ending in 64.

3 Solution of a certain linear inhomogeneous Diophantine equation

For the derivation of the recurrence relations for the start and end numbers M and N of Collatz sequences (CS) with prescribed up-down pattern (realizing a given CW) we need the general solution of the following linear and inhomogeneous Diophantine equation.

$$D(m, n; c): \qquad 3^m x - 2^n y = c(m, n), \quad m \in \mathbb{N}_0, \ n \in \mathbb{N}_0, \ c(m, n) \in \mathbb{Z} \ . \tag{2}$$

It is well known (e.g., Niven et al. [6, pp. 212–214]) how to solve the equation $ax + by = c$ for integers a, b (not 0) and c provided $g = \gcd(a, b)$ divides c for integers x and y; otherwise

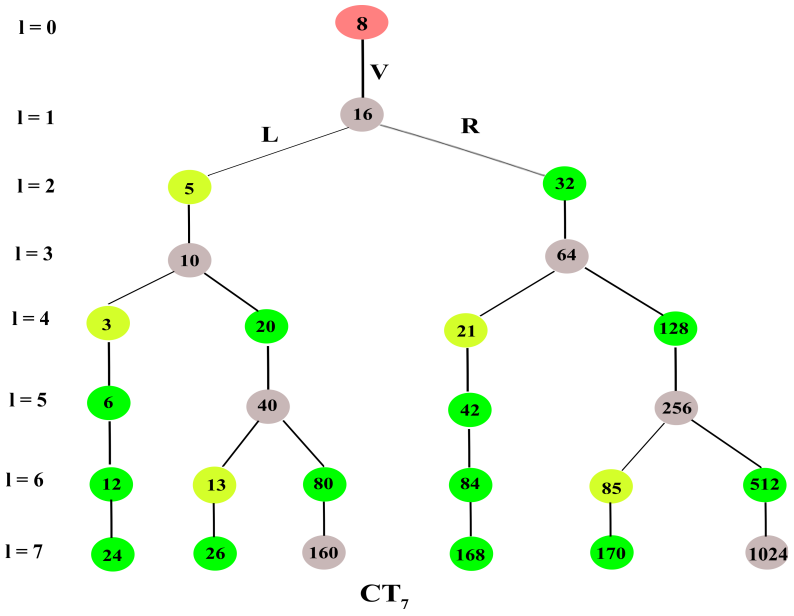

Figure 1: The Collatz Tree $\mathbf{CT_7}$

there is no solution. One finds a sequence of solutions parameterized by $t \in \mathbb{Z}$. Then one has to restrict the t range to obtain all positive solutions. The procedure is to first find a special solution (x_0, y_0) of the equation with $c = g$. Then the general solution is $(x = \frac{c}{g}x_0 + \frac{b}{g}t, y = \frac{c}{g}y_0 - \frac{a}{g}t)$ with $t \in Z$. The proof is found in [6]. For our problem $g = \gcd(3^m, 2^n) = 1$ for non-negative m, n, and this g divides any $c(m, n)$.

Lemma 1. *Solution of* $D(m, n; c)$

(a) *A special positive integer solution of* $D(m, n; 1)$ *is*

$$\begin{aligned} y_0(m,n) &= \left(\frac{3^m+1}{2}\right)^{n+3^{m-1}} \pmod{3^m}\ , \\ x_0(m,n) &= \frac{1+2^n y_0(m,n)}{3^m}\ . \end{aligned} \tag{3}$$

(b) *The general solution with positive* x *and* y *is*

$$\begin{aligned} x(m,n) &= c(m,n)x_0(m,n) + 2^n t_{\min}(m,n;\mathrm{sgn}(c)) + 2^n k\ , \\ y_0(m,n) &= c(m,n)y_0(m,n) + 3^m t_{\min}(m,n;\mathrm{sgn}(c)) + 3^m k\ , \end{aligned} \tag{4}$$

with $k \in \mathbb{N}_0$, *and*

$$t_{\min}(m,n;\mathrm{sgn}(c)) = \begin{cases} \left\lceil |\, c(m,n)\, | \frac{x_0(m,n)}{2^n} \right\rceil, & \text{if } c < 0; \\ \left\lceil -c(m,n) \frac{y_0(m,n)}{3^m} \right\rceil, & \text{if } c \geq 0. \end{cases} \tag{5}$$

For the proof we use the following lemma:

Lemma 2. *$A(n,m) := \binom{n-1}{m-1}\frac{\gcd(m,n)}{m}$ is a positive integer for $m = 1, 2, \ldots, n,\ n \in \mathbb{N}$.*

Proof. (due to *Peter Bala*, see A107711, history, February 28 2014)

This is the triangle A107711 with A(0,0) =1. By a rearrangement of factors one also has $A(n,m) = \binom{n}{m}\frac{\gcd(n,m)}{n}$. Use $\gcd(n,m)\,\mathrm{lcm}(m,n) = nm$ (e.g., [2, Theorem 2.2.2, pp. 15–16], where the uniqueness of the lcm is also shown). Now

$$A(n,m) = \frac{a(n,m)}{\mathrm{lcm}(n,m)}$$

with

$$a(n,m) = \binom{n}{m}m,$$

a positive integer because the binomial is a combinatorial number. Now $m|a(n,m)$ and $n|a(n,m)$ because

$$a(n,m) = n\binom{n-1}{m-1}$$

by a rearrangement. Hence $a(n,m) = k_1 m = k_2 n$, i.e., $a(n,m)$ is a common multiple of n and m (call it $\mathrm{cm}(n,m)$). Now $\mathrm{lcm}(n,m)|a(n,m)$ because $\mathrm{lcm}(n,m)$ is the (unique) lowest $\mathrm{cm}(n,m)$. Therefore $\frac{a(n,m)}{\mathrm{lcm}(n,m)} \in \mathbb{N}$, since only natural numbers are in the game. □

Now to the proof of Lemma 1.

Proof. **(a)** $x_0(m,n) = \frac{1+2^n y_0(m,n)}{3^m}$ is a solution of $D(m,n;1)$ for any $y_0(m,n)$. The given $y_0(m,n)$ is a positive integer $\in \{1,2,\ldots,3^m-1\}$ $m \in \mathbb{N}$ and $y_0(0,n) = 1$ for $n \in \mathbb{N}_0$. One has to prove that $x_0(m,n)$ is a positive integer. This can be done by showing that $1 + 3^n y_0(n,m) \equiv 0 \pmod{3^m}$ for $m \in \mathbb{N}$. One first observes that $\frac{3^m+1}{2} \equiv \frac{1}{2} \pmod{3^m}$, because obviously $2\frac{3^m+1}{2} \equiv 1 \pmod{3^m}$ (2 is a unit in the ring $\mathbb{Z}_{3^m}$). For $m = 0$ one has $x_0(0,n) = 1 + 2^n$, $n \in \mathbb{N}_0$, which is positive. In the following $m \in \mathbb{N}$.

$$1 + 2^n\left(\frac{3^m+1}{2}\right)^{n+3^{m-1}} \equiv 1 + 2^n\left(\frac{1}{2}\right)^{n+3^{m-1}} \equiv 1 + \left(\frac{1}{2}\right)^{3^{m-1}} \pmod{3^m}\,. \qquad (6)$$

Now we show that $L(m) := \left(\frac{3^m+1}{2}\right)^{3^{m-1}} \equiv 0 \pmod{3^m}$ by using $\frac{3^m+1}{2} = 3k(m) - 1$ with

$k(m) := \frac{3^{m-1}+1}{2}$, a positive integer. The binomial theorem leads with $a(m) = 3^{m-1}$ to

$$\begin{aligned}
(3k(m)-1)^{a(m)} &= \sum_{j=0}^{a(m)-1} \binom{a(m)}{j} (-1)^j (3k(m))^{a(m)-j} \\
&= 3^m \Sigma_1(m) + \Sigma_2(m), \text{with} \qquad (7) \\
\Sigma_1(m) &= \sum_{j=0}^{a(m)-m} (-1)^j \binom{a(m)}{j} k(m)^{a(m)-j} 3^{a(m)-m-j}, \text{ and} \qquad (8) \\
\Sigma_2(m) &= \sum_{j=a(m)-m+1}^{a(m)-1} (-1)^j \binom{a(m)}{j} (3k(m))^{a(m)-j} \ . \qquad (9)
\end{aligned}$$

Now $\Sigma_1(m)$ is an integer because of $a(m) - j \geq a(m) - m - j \geq 0$ and the integer binomial, hence $L(m) \equiv \Sigma_2 \pmod{3^m}$. Rewriting Σ_2 with $j' = j - a(m) + m - 1$, and also using the symmetry of the binomial, one has

$$\begin{aligned}
\Sigma_2(m) &= \sum_{j=0}^{m-2} \binom{a(m)}{j} (-1)^{j-m} (3k(m))^{m-1-j} \\
&= \sum_{j=1}^{m-1} (-1)^{j+1} \binom{a(m)}{j} (3k(m))^j = 3^m \widehat{\Sigma}_2(m) \text{ with} \qquad (10) \\
\widehat{\Sigma}_2(m) &= \sum_{j=1}^{m-1} (-1)^{1+j} \binom{a(m)}{j} k(m)^j 3^{j-m} \\
&= \sum_{j=1}^{m-1} (-1)^{1+j} k(m)^j \binom{a(m)-1}{j-1} \frac{1}{j} 3^{j-1} \ . \qquad (11)
\end{aligned}$$

In the last step a rearrangement of the binomial has been applied, remembering that $a(m) = 3^{m-1}$. It remains to be shown that $A_{m,j} := 3^{j-1}\binom{3^{m-1}-1}{j-1}\frac{1}{j}$ is a (positive) integer for $j = 1, 2, \ldots, m-1$. Here Lemma 2 comes to help. Consider there $A(3^{m-1}, j)$ for $j = 1, 2, \ldots, m-1$ ($m = 0$ has been treated separately above), which is a positive integer. If $3 \nmid j$ then $3^{j-1}A(3^{m-1}j) = A_{m,j}$, hence a positive integer. If $j = 3^k J$, with $k \in \mathbb{N}$ the largest power of 3 dividing j then $\gcd(3, J) = 1$, and $j = 3^k J \leq m - 1 < 3^{m-1}$ and $\gcd(3^{m-1}3^k J) = 3^q$ with $q = \min(k, m-1)$. □

Proof. **(b)** The general integer solution of Eq. (2) is then (see [6, pp. 212–214]; note that there $b > 0$, here $b < 0$, and we have changed $t \mapsto -t$)

$$\begin{aligned}
x = \hat{x}(m, n; t) &= c(m, n) x_0(m, n) + 2^n t, \\
y = \hat{y}(m, n; t) &= c(m, n) y_0(m, n) + 3^m t, \ t \in \mathbb{Z}. \qquad (12)
\end{aligned}$$

In order to find all positive solutions for x and y one has to restrict the range of t, depending on $\operatorname{sgn} c$. If $c(m,n) \geq 0$ then, because x_0 and y_0 are positive and

$$\frac{x_0(m,n)}{2^n} = \frac{y_0(m,n)}{3^m} + \frac{1}{2^n 3^m},$$

we have $t > -\frac{c(m,n)x_0(m,n)}{2^n}$ and $t > -\frac{c(m,n)y_0(m,n)}{3^m}$, i.e.,

$$\begin{aligned} t &\geq \left\lceil \max\left(-\frac{c(m,n)x_0(m,n)}{2^n}, -\frac{c(m,n)y_0(m,n)}{3^m} \right) \right\rceil \\ &= \left\lceil -c(m,n) \min\left(\frac{x_0(m,n)}{2^n}, \frac{y_0(m,n)}{3^m} \right) \right\rceil \\ &= \left\lceil -c(m,n) \frac{y_0(m,n)}{3^m} \right\rceil \\ &= t_{\min}(m,n;+). \end{aligned}$$

If $c(m,n) < 0$ then

$$\begin{aligned} t &\geq \left\lceil |\, c(m,n)\,| \max\left(\frac{x_0(m,n)}{2^n}, \frac{y_0(m,n)}{3^m} \right) \right\rceil \\ &= \left\lceil |\, c(m,n)\,| \frac{x_0(m,n)}{2^n} \right\rceil \\ &= t_{\min}(m,n;-). \end{aligned}$$

Thus with $t = t_{\min}(m,n;\operatorname{sgn}(c)) + k$, with $k \in \mathbb{N}_0$ one has the desired result. Note that $(x_0(m,n), y_0(m,n))$ is the smallest positive solution of the equation $D(m,n;1)$, Eq. (2), because, for $c(m,n) = 1$, $t_{\min}(m,n;+) = \left\lceil -\frac{y_0(m,n)}{3^m} \right\rceil$, but with $y_0(m,n) \in \{1,2,\ldots,3^m-1\}$ this is 0. □

A proposition on the periodicity of the solution $y_0(m,n)$ follows.

Proposition 3. *Periodicity of $y_0(m,n)$ in n*

(a) *The sequence $y_0(m,n)$ is periodic in n with primitive period length $L_0 = \varphi(3^m)$, for $m \in \mathbb{N}_0$ with Euler's totient function $\varphi(n) =$ A000010(n), where $\varphi(1) := 1$.*

(b) *The sequence $x_0(m, n = L_0(m)) = q(m)x_0(m,n) - r(m)$, $m \in \mathbb{N}_0$, with $q(m) := 2^{\varphi(3^m)}$ and $r(m) := \frac{2^{\varphi(3^m)}-1}{3^m}$. See A152007.*

(c) *The set $Y_0(m) := \{y_0(m,n) | n = 0,1,\ldots,\varphi(3^m)-1\}$, is, for $m \in \mathbb{N}_0$, a representation of the set $RRS(3^m)$, the smallest positive restricted residue system modulo 3^m. See Apostol [1, p. 113] for the definition. The multiplicative group modulo 3^m, called $\mathbb{Z}_{3^m}^{\times} = (\mathbb{Z}/3^m\mathbb{Z})^{\times}$ is congruent to the cyclic group $C_{\varphi(3^m)}$.*

Proof. **(a)** By Euler's theorem (e.g., [2, Theorem 2.4.4.3, p. 32]) $a^{\varphi(n)} \equiv 1 \pmod{n}$, provided $\gcd(a,n) = 1$. Now $\gcd\left(\frac{3^m+1}{2}, 3^m\right) = \gcd\left(\frac{3^m+1}{2}, 3\right) = 1$ because $\frac{3^m+1}{2} \equiv \frac{1}{2} \pmod{3^m}$ (see above), and hence $\frac{3^m+1}{2} \not\equiv 0 \pmod{3^m}$. This shows that $L_0(m)$ is a period length, but we have to show that it is in fact the length of the primitive period, i.e., we have to prove that the order of $\frac{3^m+1}{2}$ modulo 3^m is $L_0(m)$. (See e.g., [2, Definition 2.4.4.1, p. 31], for the order definition.) In other words we want to show that $\frac{3^m+1}{2}$ is a primitive root (of 1) modulo 3^m. Assume that $k(m)$ is this order (the existence is certain due to Euler's theorem). Hence $(\frac{1}{2})^{k(m)} \equiv 1 \pmod{3^m}$ and $k(m)|L_0(m)$. It is known that the modulus 3^m possesses primitive roots, and the theorem on the primitive roots says that there are precisely $\varphi(\varphi(3^m))$ incongruent ones (e.g., Niven et al. [6, pp. 205, 207], or Nagell [5, Theorem 62.3, p. 104 and Theorem 65, p. 107]). In our case this number is $\varphi(2 \cdot 3^{m-1}) = 2 \cdot 3^{m-2}$ if $m \geq 1$. The important point, proven in Nagel [5, Theorem 65.3, p. 107], is that if we have a primitive root r modulo an odd prime, here 3, then, if $r^{3-1} - 1$ is not divisible by 3^2, it follows that r is in fact a primitive root for any modulus 3^q, with $q \in \mathbb{N}_0$. One of the primitive roots modulo 3 is 2, because $2^2 = 4 \equiv 1 \pmod{3}$ and $2^1 \not\equiv 1 \pmod{3}$. Also $2^{3-1} - 1 = 3$ is not divisible by 3^2, hence 2 is a primitive root of any modulus 3^q for $q \in \mathbb{N}_0$. From this we proof that $\frac{3^m+1}{2} \equiv \frac{1}{2} \pmod{3^m}$ is a primitive root modulo 3^m. Consider $\left(\frac{3^m+1}{2}\right)^k \equiv \frac{1}{2^k} \pmod{3^m}$ for $k = 1, 2, \ldots, \varphi(3^m)$. In order to have $\left(\frac{1}{2}\right)^k \equiv 1 \pmod{3^m}$ one needs $2^k \equiv 1 \pmod{3^m}$. But due to Nagel [5, Theorem 65.3, p. 107], for $p = 3$, a primitive root modulo 3^m is 2, and the smallest positive k is therefore $\varphi(3^m)$, and hence $\frac{3^m+1}{2}$ is a primitive root (of 1) of modulus 3^m.

(b) $x_0(m, n + \varphi(3^m)) = \frac{1+2^n 2^{\varphi(3^m)} y_0(m,n)}{3^m}$ from the periodicity of y_0. Rewritten, this is

$$\begin{aligned}\frac{2^{\varphi(3^m)}\left((2^{-\varphi(3^m)} - 1) + (1 + 2^n\, y_0(m,n)\right)}{3^m} &= -\frac{1}{3^m}\left(2^{\varphi(3^m)} - 1\right) + 2^{\varphi(3^m)} x_0(m,n) \\ &= q(m) x_0(m,n) - r(m)\end{aligned}$$

with the values given in the proposition.

(c) This follows from the reduced residue system modulo 3^m for $m \in \mathbb{N}_0$,

$$\left\{\left(\frac{1}{2}\right)^0, \left(\frac{1}{2}\right)^1, \ldots, \left(\frac{1}{2}\right)^{\varphi(3^m)-1}\right\},$$

because $\frac{1}{2}$ is a primitive root modulo 3^m (from part **b)**). With $a(m) := \frac{3^m+1}{2}$ one has $1 = \gcd(a(m), 3) = \gcd(a(m), 3^m) = \gcd(a(m)^{b(m)} 3^m)$ with $b(m) := 3^{m-1}$, and

$$\left\{a(m)^{b(m)}\left(\frac{1}{2}\right)^0, a(m)^{b(m)}\left(\frac{1}{2}\right)^1, \ldots, a(m)^{b(m)}\left(\frac{1}{2}\right)^{\varphi(3^m)-1}\right\}$$

is a reduced residue system modulo 3^m (see Apostol [1, Theorem 5.16, p. 113]). Thus

$$Y_0(m) \equiv \{a(m)^{b(m)} 1, a(m)^{b(m)+1}, \ldots, a(m)^{b(m)+\varphi(3^m)-1}\}$$

is a reduced residue system modulo 3^m. Therefore this gives a permutation of the reduced residue system modulo 3^m with the smallest positive integers sorted increasingly. □

Example 4. For $m = 3$, $\varphi(3^3) = 2 \cdot 3^2 = 18 = L_0(3)$, and

$$\{y_0(3,n)\}_{n=0}^{17} = \{26, 13, 20, 10, 5, 16, 8, 4, 2, 1, 14, 7, 17, 22, 11, 19, 23, 25\}$$

is a permutation of the standard reduced residue system modulo 27, obtained by re-sorting the found system in increasing order. See A239125. For $m = 1, 2$, and 4 see A007583, A234038 and A239130 for the solutions $(x_0(m,n), y_0(m,n))$.

4 Recurrences and their solution

After these preparations it is straightforward to derive the recurrence for the start and end numbers M and N for any given $CW(\vec{n}_S)$, for $S \in \mathbb{N}$.

(A) We first consider the case of words with $n_S = 1$, i.e., $\vec{n}_S = [n_0, n_1, \ldots, n_{S-1}, 1]$. This is the word $CW(\vec{n}_S) = \overrightarrow{\prod}_{j=0}^{S-1} d^{n_j} s$ (with an ordered product, beginning with $j = 0$ at the left-hand side). In order to simplify the notation we use $M(S)$, $N(S)$, $y_0(S)$, $x_0(S)$, and $c(S)$ for $M(\vec{n}_S)$, $N(\vec{n}_S)$, $y_0(S, n_S)$, $x_0(S, n_S)$ and $c(S, n_S)$, respectively. For $S = 1$, the input for the recurrence, one has

$$M(1;k) = 2^{n_0}(2\,k+1) \text{ and } N(1;k) = 3k+2, \text{ for } k \in \mathbb{N} \ , \tag{13}$$

because there are n_0 factors of 2 from d^{n_0}, and then an odd number $2k+1$ leads after application of s to $3k+2$. Thus $M(1) = 2^{n_0}$ and $N(1) = 2$.

Proposition 5. *Recurrences for $M(S)$ and $N(S)$ with $n_S = 1$.*

(a) *The coupled recurrences for $M(S,t)$ and $N(S,t)$, the first and last entry of the Collatz sequences $CS(\vec{n}_S; t)$ for the word $CW(\vec{n}_S)$ with $\vec{n}_S = [n_0, n_1, \ldots, n_{S-1}, 1]$ ($n_S = 1$) are*

$$\begin{aligned} M(S,t) &= M(S) \ + \ 2^{\hat{D}(S)} t, \\ N(S,t) &= N(S) \ + \ 3^S t, \text{ with } t \in \mathbb{Z} \ , \end{aligned} \tag{14}$$

where $\hat{D}(S) := \sum_{j=0}^{S-1} n_j$ (we prefer to use a new symbol for the $n_S = 1$ case), and the recurrences for $M(S)$ and $\widetilde{N}(S) = N(S) - 2$ are

$$\begin{aligned} M(S) = M(S-1) \ &+ \ 2^{\hat{D}(S-1)} c(S-1) x_0(S-1) \ , \\ \widetilde{N}(S) \ &= \ 3 y_0(S-1) c(S-1), \end{aligned} \tag{15}$$

with

$$c(S-1) = 2(2^{n_{S-1}-2} - 1) - \widetilde{N}(S-1) =: A(S-1) - \widetilde{N}(S-1) \ . \tag{16}$$

The recurrence for $c(S)$ is

$$c(S) = -3y_0(S-1)c(S-1) + A(S),\ S \geq 2, \tag{17}$$

and the input is $M(1) = 2^{n_0}$, $\widetilde{N}(1) = 0$ and $c(1) = A(1)$.

(b) *The general positive integer solution is*

$$\begin{aligned} M(S;k) &= M(S) + 2^{\hat{D}(S)} t_{\min}(S-1) + 2^{\hat{D}(S)} k, \\ N(S;k) &= 2 + \widetilde{N}(S) + 3^S t_{\min}(S-1) + 3^S k,\ k \in \mathbb{N}_0, \end{aligned} \tag{18}$$

where

$$t_{\min}(S) = t_{\min}(S, n_S, \mathrm{sgn}(c(S))) = \begin{cases} \left\lceil |\, c(S)\, | \frac{x_0(S)}{2^{n_S}} \right\rceil, & \text{if } c(S) < 0; \\ \left\lceil -c(S) \frac{y_0(S)}{3^S} \right\rceil, & \text{if } c(S) \geq 0. \end{cases} \tag{19}$$

Corollary 6.

$$\begin{aligned} M(S;k) &\equiv M(S) + 2^{\hat{D}(S)} t_{\min}(S-1) \pmod{2^{\hat{D}(S)}}, \\ N(S;k) &\equiv \widetilde{N}(S) + 3^S t_{\min}(S-1) \pmod{3^S}. \end{aligned} \tag{20}$$

In Terras' article [8] the first congruence corresponds to Theorem 1.2, where the encoding vector $E_k(n)$ refers to the modified Collatz tree using only d and s operations.

Proof. **(a)** By induction over S. For $S = 1$ the input $M(1) = 2^{n_0}$, $N(1) = 2$ or $\widetilde{N}(1) = 0$ provides the start of the induction. Assume that part **(a)** of the proposition is true for S values $1, 2, \ldots, S-1$. To find $M(S)$ one has to make sure that $d^{n_{S-1}} s$ can be applied to $N(S-1;k)$, the end number of step $S-1$ sequence $CS(\vec{n}_{S-1};t)$ which is $N_{int}(S-1,t) = N(S-1) + 3^{S-1}t$, with integer t, by the induction hypothesis. This number has to be of the form $2^{n_{S-1}-1}(2m+1)$ (one has to have an odd number after n_{S-1} $d-$steps such that s can be applied). Thus $3^{S-1}t - 2^{n_{S-1}}m = 2^{n_{S-1}-1} - N(S-1) = A(S-1) - \widetilde{N}(S-1) =: c(S-1)$, where $\widetilde{N}(S-1) = N(S-1) - 2$ and $A(S-1) = 2(2^{n_{S-1}-2} - 1)$. Due to Lemma 1 the general solution, with $t \to x(S-1, n_{S-1}; t) \hat{=} x(S-1;t)$, $m \to y(S-1, n_{S-1}; t) \hat{=} y(S-1;t)$, to shorten the notation, is

$$\begin{aligned} t \to x(S;t) &= c(S-1)x_0(S-1) + 2^{n_{S-1}} t, \\ m \to y(S;t) &= c(S-1)y_0(S-1) + 3^{S-1} t,\ t \in \mathbb{Z}\ . \end{aligned} \tag{21}$$

Therefore the first entry of the sequence $CS(\vec{n}_S;t)$ is $M(S;t) = M(s-1, x(S-1,t))$ which is

$$M_{int}(S;t) = M(S-1) + 2^{\hat{D}(S-1)} c(S-1)x_0(S-1) + 2^{\hat{D}(S)} t, \tag{22}$$

hence $M(S) = M(S-1) + 2^{\hat{D}(S-1)} c(S-1)x_0(S-1)$, the claimed recurrence for $M(S)$.

The last member of $CS(\vec{n}_{S-1};t)$ is $3m+2$ (after applying s on $2m+1$ from above). Thus $N_{int}(S;t) = 3y(S;t) + 2$, or $N_{int}(S;t) - 2 = 3c(S-1)y_0(S-1) + 3^S t$. Therefore,

$\widetilde{N}(S) = N(S) - 2 = 3c(S-1)y_0(S-1)$ which is the claim for the $\widetilde{N}$ recurrence. Note that the remainder structure of eqs. 18, expressed also in the Corollary 6, has also been verified by this inductive proof. The recurrence for $c(S) = A(S) - \widetilde{N}(S)$ follows from the one for $\widetilde{N}(S)$.

(b) Positive integer solutions from $M_{int}(S;t)$ and $N_{int}(S;t)$ of part **(a)** are found from the second part of Lemma 1 applied to the equation $3^{S-1}x - 2^{n_S-1}y = c(S-1)$, determining $t_{\min}(S-1)$ as claimed. This leads finally to the formulae for $M(S;k)$ and $N(S;k)$ with $k \in \mathbb{N}_0$. □

Example 7. $(sd)^{S-1}s$ Collatz sequences

Here $n_0 = 0 = n_S$ and $n_j = 2$ for $j = 1, 2, \ldots, S-1$. The first entries $M(S;k)$ and the last entries $N(S;k)$ of the Collatz sequence $CS([0,2,\ldots,2];k)$ (with $S-1$ times a 2), whose length is $3S$, are $M(S;k) = 1 + 2^{2S-1}k$ and $N(S;k) = 2 + 3^S k$. For $S = 3$ a complete Collatz sequence $CS([0,2,2];3)$ of length 9 is $[97, 292, 146, 73, 220, 110, 55, 166, 83]$ which is a special realization of the word $sdsds$ with starting number $M(3;3) = 97$ ending in $N(3;3) = 83$. Note that for this $u-d$ pattern the start and end numbers have remainders $M(S;0) = M(1;0) = 1$ and $N(S;0) = N(1;0) = 2$. See the tables A240222 and A240223.

The recurrences for $M(S) \hat{=} M(\vec{n}_{S-1})$, $\tilde{N}(S) \hat{=} \tilde{N}(\vec{n}_{S-1})$ or $N(S) \hat{=} N(\vec{n}_{S-1})$ and $c(S) \hat{=} c(\vec{n}_{S-1})$ are solved by iteration with the given inputs $M(1) = 2^{n_0}$, $\tilde{N}(1) = 0$ and $c(1) = A(1) = 2(2^{n_0-2} - 1)$.

Proposition 8. *Solution of the recurrences for $n_s = 1$*

The solution of the recurrences of Proposition 5 with the given inputs are, for $S \in \mathbb{N}$:

$$\begin{aligned} c(S) &= A(S) + \sum_{j=1}^{S-1} (-3)^j A(S-j) \prod_{l=1}^{j} y_0(S-l), \\ \tilde{n}(S) &= A(S) - c(S) = -\sum_{j=1}^{S-1} (-1)^j A(S-j) \prod_{l=1}^{j} y_0(S-l), \\ N(s) &= \tilde{N}(S) + 2, \\ M(S) &= 2^{n_0} + \sum_{j=1}^{S-1} R(S-j), \end{aligned} \tag{23}$$

with $\hat{D}(S) := 1 + \sum_{j=0}^{S-1} n_j$, $A(S) := 2(2^{n_S-2} - 1)$, $R(S) := 2^{\hat{D}(S)} x_0(S) c(S)$, and $y_0(S) \hat{=} y_0(S, n_S)$, $x_0(S) \hat{=} x_0(S, n_S)$, given in Lemma 1.

Proof. This is obvious. □

(B) The general case $n_S = 1$ can now be found by appending the operation d^{n_S-1} to the above result. This leads to the following theorem.

Theorem 9. *The general case $\vec{n}_S$.*

For the Collatz word $CW(\vec{n}_S) = d^{n_0} \overrightarrow{\prod}_{j=1}^{S}(sd^{n_j-1}) = d^{n_0} s \overrightarrow{\prod}_{j=1}^{S}(d^{n_j-1}s)d^{n_S-1}$ (the ordered product begins with $j=1$ on the left-hand side) with $n_0 \in \mathbb{N}_0$, $n \in \mathbb{N}$, the first and last entries of the corresponding Collatz sequences $\{CS(\vec{n}_S;k)\}$, of length $L(S) = n_0 + 2\sum_{j=1}^{S} n_j$, for $k \in \mathbb{N}_0$, are

$$\begin{aligned} M(\vec{n}_S;k) &= M(S) - 2^{\hat{D}(S)} N(S) x_0(S, n_S - 1) + 2^{D(S)} t_{\min}(S, n_S - 1, \mathrm{sgn}(c_{new}(S)) \\ &\quad + 2^{D(S)} k, \\ N(\vec{n}_S;k) &= c_{new}(S) y_0(S, n_S - 1) + 3^S t_{\min}(S, n_S - 1, \mathrm{sgn}(c_{new}(S)) + 3^S k, \end{aligned} \tag{24}$$

with $c_{new}(S) := -N(S)$, $\hat{D}(S) = 1 + \sum_{j=0}^{S-1} n_j$, $D(S) = \sum_{j=0}^{S} n_j$.

Proof. In order to be able to apply to the Collatz sequences $CS([n_0, n_1, \ldots, n_S - 1, 1])$ (with the results from part (A) above) the final d^{n_S-1} operation one needs for the last entries $N_{int}(S;t) = N(S) + 3^S t = 2^{n_S-1} m$ with some (even or odd) integer m. The new last entries of $CS([n_0, n_1, \ldots, n_S]; t)$ are then m. The general solution of $3^S - 2^{n_S-1} m = -N(S) =: c_{new}(S)$ is, according to Lemma 1, given by

$$\begin{aligned} t \to x(S;t) &= c_{new}(S) x_0(S, n_S - 1) + 2^{n_S-1} t, \\ m \to y(S;t) &= c_{new}(S) y_0(S, n_S - 1) + 3^S t, \ t \in \mathbb{Z}. \end{aligned} \tag{25}$$

This leads to positive integer solutions after the shift $t \to t_{\min} + k$, with $t_{\min} = t_{\min}(S, n_S - 1, \mathrm{sgn}(c_{new}(S)))$ to the claimed result $N(S;k)$ for the new last number of $CS(\vec{n};k)$, with $k \in \mathbb{N}_0$. The new start value $M(S;k)$ is obtained by replacing $t \to x(S;t)$ in the old $M_{int}(S;t)$ (with $n_S = 1$). $M(S;k) = M_{int}(S, x(S;t))$ with $t \to t_{\min} + k$, also leading to the claimed formula. □

The remainder structure modulo $2^{D(S)}$ for $M(\vec{n}_S, k)$ and modulo 3^S for $N(\vec{n}_S, k)$ is manifest.

The explicit sum versions of the results for case $n_S = 1$, given in Proposition 9, can be inserted here.

Example 10. $ud^m = sd^{m-1}$

For $m = 1, 2, 3$ and $k = 0, 1, \ldots, 10$ one finds for $N([0, m], k)$:

$$[2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32],$$
$$[1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31],$$
$$[2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32],$$

and for $M([0, m], k)$:

$$[1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21],$$
$$[1, 5, 9, 13, 17, 21, 25, 29, 33, 37, 41],$$
$$[5, 13, 21, 29, 37, 45, 53, 61, 69, 77, 85].$$

Only the odd members of $N([0,m],k)$, that is, the odd-indexed entries, and the corresponding $M([0,m],k)$ appear in the article of Trümper [9] as Example 2.1. See A238475 for $M([0,2n],k)$ and A238476 for $M([0,2n-1],k)$. The odd $N([0,2n],k)$ values are the same for all n, namely $5+6k$, and $N([0,2n-1],k) = 1+6k$ for all $n \in \mathbb{N}$.

Example 11. $(ud)^n = s^S, S \in \mathbb{N}$

$\vec{n}_S = [0,1,\ldots,1]$ with S times a 1. For $S = 1,2,3$ and $k = 0,1,\ldots,10$ one finds for $N(\vec{n}_S,k)$:

$$[5,8,11,14,17,20,23,26,29,32],$$
$$[17,26,35,44,53,62,71,80,89,98],$$
$$[53,80,107,134,161,188,215,242,269,296],$$

and for $M(\vec{n}_S,k)$:

$$[3,5,7,9,11,13,15,17,19,21],$$
$$[7,11,15,19,23,27,31,35,39,43],$$
$$[15,23,31,39,47,55,63,71,79,87].$$

For odd N entries, and corresponding M entries this is in the article of Trümper [9, Example 2.1].

See A239126 for these M values, and A239127 for these N values, which are here S dependent.

In conclusion the author does not think that the knowledge of all Collatz sequences with a given up-down pattern (a given Collatz word) helps to prove the Collatz conjecture. Nevertheless the problem considered in this paper is a nice application of the solution of a simple Diophantine equation.

5 Acknowledgment

Thanks go to Peter Bala who answered the author's question for a proof that all triangle A107711 entries are non-negative integers. See the history there, dated February 28 2014.

6 Note added

The referee commented that "... the line of reasoning of the author shows resemblance with the derivation of cycles for the general Collatz problem ($T(n) = \frac{pn+q}{2}$ if n is odd) based on Lyndon words [3] and the extra equality $x = y$" (in the present work $x = M$ and $y = N$).

References

[1] T. M. Apostol, *Introduction to Analytic Number Theory*, Springer-Verlag, 1976.

[2] B. Fine and G. Rosenberger, *Number Theory*, Birkhäuser, 2007.

[3] T. Laarhoven and B. de Weger, The Collatz conjecture and De Bruijn graphs, *Indag. Math. (N.S.)* **24** (2013) 971–983. Preprint available at http://arxiv.org/abs/1209.3495.

[4] J. Lagarias, The $3x+1$ problem and its generalization, *Amer. Math. Monthly* **92** (1985), 3–23.

[5] T. Nagell, *Introduction to Number Theory*, Chelsea Publishing Company, 1964.

[6] I. Niven, H. S. Zuckerman, and H. L. Montgomery, *An Introduction to the Theory Of Numbers*, Fifth Edition, John Wiley and Sons, Inc., 1991.

[7] The On-Line Encyclopedia of Integer Sequences (2010), published electronically at http://oeis.org.

[8] R. Terras, A stopping time problem on the positive integers, *Acta Arith.* **30** (1976), 241–252.

[9] M. Trümper, The Collatz problem in the light of an infinite free semigroup, *Chinese J. Mathematics*, **2014** (2014), Article ID 756917, http://www.hindawi.com/journals/cjm/aip/756917/.

2010 *Mathematics Subject Classification*: Primary 11D04; Secondary 32H50.
Keywords: Collatz problem, Collatz sequences, Collatz tree, recurrence, iteration, linear Diophantine equation.

(Concerned with sequences A000010, A005186, A007583, A016957, A107711, A127824, A152007, A176866, A234038, A238475, A238476, A239125, A239126, A239127, A239130, A240222, and A240223.)

Received April 10 2014; revised versions received October 2 2014; November 5 2014; November 11 2014. Published in *Journal of Integer Sequences*, December 12 2014.

Return to Journal of Integer Sequences home page.