American Journal of Computational Mathematics, 2021, 11, 226-239
https://www.scirp.org/journal/ajcm
ISSN Online: 2161-1211
ISSN Print: 2161-1203

Collatz Sequences and Characteristic Zero-One Strings: Progress on the 3*x* + 1 Problem

David C. Kay*

Asheville, USA
Email: dck.gm09@yahoo.com

How to cite this paper: Kay, D.C. (2021) Collatz Sequences and Characteristic Zero-One Strings: Progress on the 3*x* + 1 Problem. *American Journal of Computational Mathematics*, **11**, 226-239.
https://doi.org/10.4236/ajcm.2021.113015

Received: May 25, 2021
Accepted: September 27, 2021
Published: September 30, 2021

Abstract

The unsolved number theory problem known as the 3*x* + 1 problem involves sequences of positive integers generated more or less at random that seem to always converge to 1. Here the connection between the first integer (*n*) and the last (*m*) of a 3*x* + 1 sequence is analyzed by means of characteristic zero-one strings. This method is used to achieve some progress on the 3*x* + 1 problem. In particular, the long-standing conjecture that nontrivial cycles do not exist is virtually proved using probability theory.

Keywords

Generator, Resultant, 3*x* + 1 Cycle

1. Introduction

Everett [1] (Iteration of the number-theoretic function $f(2n) = n$, $f(2n + 1) = 3n + 2$) introduced the concept of parity vectors to obtain early results concerning the 3*x* + 1 problem. A different aspect of that approach is to focus mainly on the elements of such vectors (zeros and ones) and to index them differently. The 3*x* + 1 problem involves using the following number theory algorithm: starting with any positive integer *n*, if *n* is even divided by 2, or if *n* is odd multiply by 3, add 1, then divided by 2. This generates the Collatz sequence $\{n = n_1, n_2, \cdots\}$ (named after Lothar Collatz who introduced the problem). Experimentation suggests that such sequences always end in the trivial cycle {1, 2, 1, 2, …}. The 3*x* + 1 problem is to prove this for all Collatz sequences $C(n)$. Alternatively, one must prove that every Collatz sequence $C_k(n)$ of finite length *k* converges to 1 for all positive integers *n*, *k* large enough. This has been numerically verified for all $n < n^* = 20 \times 2^{58} = 5.7646 \times 10^{18}$ (Lagarias [2]), and more recently to $n^* = 8.7 \times 10^{18}$.

*Formerly worked in the University of North Carolina at Asheville, Asheville, USA.

A parity vector corresponding to a finite Collatz sequence is defined as $\boldsymbol{v}=(x_1,x_2,\cdots x_k)$ where $x_i=n_i\pmod 2$, $1\le i\le k$. The characteristic zero-one string of $C_k(n)$ is the set of zeros and ones in $\boldsymbol{v}$, written as:

$$L=0^{s_0}1^{r_1}0^{s_1}1^{r_2}0^{s_2}1^{r_q}0^{s_q}$$

where r_i is the number of ones occurring in a group of consecutive ones in $\boldsymbol{v}$, and s_i is the number of zeros in a following consecutive group of zeros, $0 \le i \le q$. Here, $r_i \ge 1$ and $s_i \ge 1$ for all i, except for s_0 and s_q which can individually be zero.

For example, consider:

$$C_8(67)=\{67,101,152,76,38,19,29,44\}$$

Its parity vector is $\boldsymbol{v}=(1,1,0,0,0,1,1,0)$ and its characteristic zero-one string is $l=1^2 0^3 1^2 0^1$.

The total number of ones in a finite string will always be denoted r and the total number of zeros by s. Thus, in general:

$$r=r_1+r_2+\cdots+r_q \quad \text{and} \quad s=s_0+s_1+\cdots+s_q$$

The length of a string is denoted $l(L) = r + s = k$, and its norm by $\|L\|=\dfrac{3^r}{2^k}$. If L is the characteristic string of $C_k(n)$, $n = n_1$ is called a generator of L, n_k a pre-resultant, and n_{k+1} a *resultant*. If n is a generator of L and m a resultant, this relationship will be denoted nLm. A cycle, if it exists, will be denoted nLn (where L is an appropriate finite string). At first we consider only finite strings (corresponding to finite Collatz sequences).

At this point, it is clear that each positive integer is a generator of a string, but not so clear that a given (finite) string will have a generator. Note that a string can have more than one generator; for example, the string $1^1 0^1 1^3$ is generated by $C_5(9)=\{9,14,7,11,17\}$ and $C_5(41)=\{41,62,31,47,71\}$.

Much work has been done on what is called the stopping time function, $\sigma(n)$ (see Terras [3]). This function is defined to be the least positive integer k such that $n_k < n$, or $= \infty$ otherwise. Terras proved that almost all integers have a finite stopping time; that is, $\lim_{n\to\infty}\#\{\sigma(n)<\infty\}/n=1$.

It is useful to represent the $3x + 1$ problem graphically. **Figure 1** shows the graph of the Collatz sequence $C(11)$, while **Figure 2** illustrates such a convergent sequence with a large generator: $C_{20}(11, 111)$.

2. Existence of Generators

The existence of a generator for a given zero-one string depends on the following result.

Theorem A: suppose that n is the smallest generator of a string L of length k with r odd members, and it is required to find the generator n' for the string Lx of length $k + 1$, where x = either 0 or 1. If the resultant of L matches x (that is, $T^{(k)}(n)\ (mod\ 2) = x$), then $n' = n$ is the solution. But if $T^{(k)}(n)$ and x are mismatched (that is, $T^{(k)}(n)(mod\ 2) \ne x$), then $n' = n + 2^k$ is the solution. Furthermore, if m = $T^{(k)}(n)$ is the resultant of L, then $m' = m + 3^r$ is the pre-resultant of Lx.

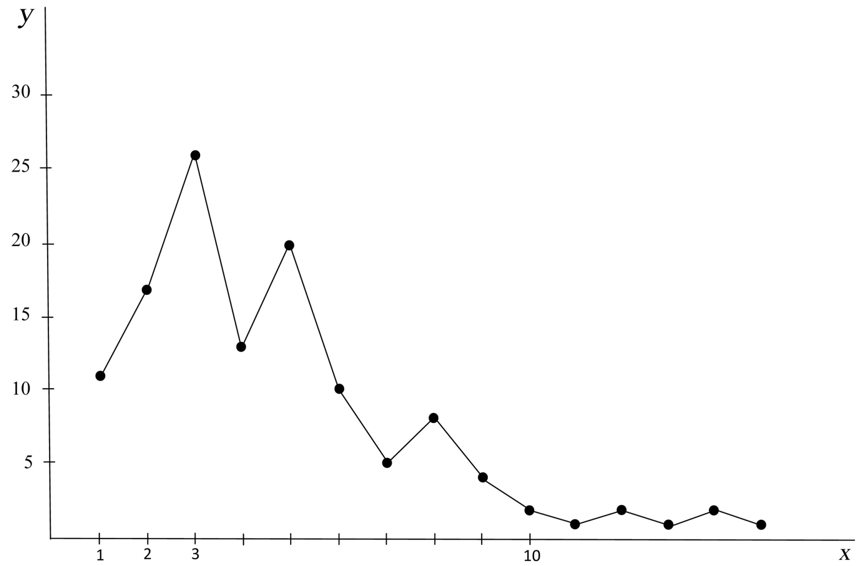

Figure 1. Collatz graph for $C(11)$.

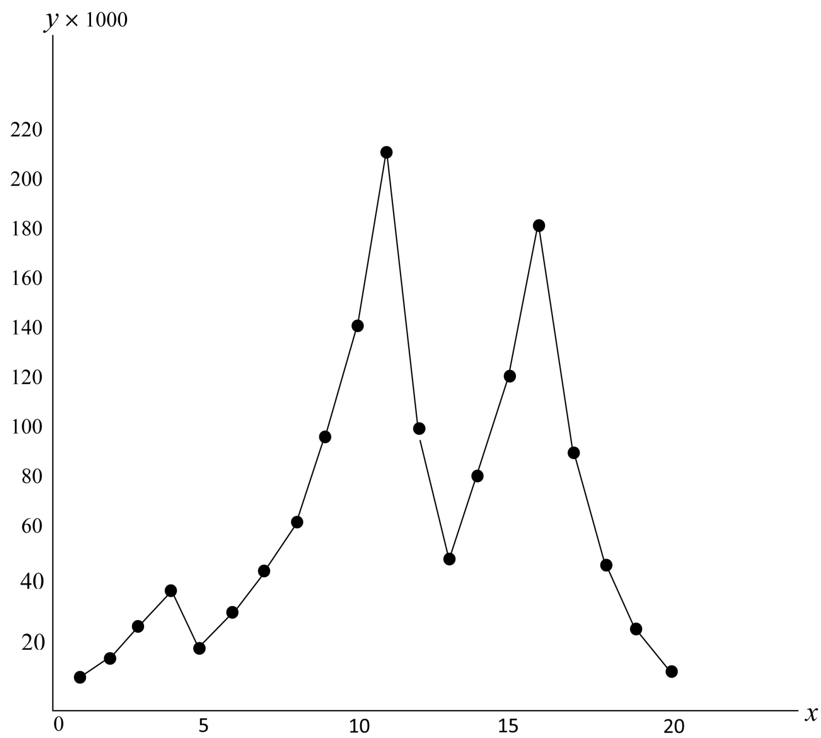

Figure 2. Graph of $C_{20}(11,111)$.

Proof: We need only prove this when the first mismatch occurs at n_k. Consider $T^{(i)}\left(n+2^k\right)$, $i \geq 1$. By induction:

$$T^{(1)}\left(n+2^k\right)=T^{(1)}(n)+u_1 2^{k-1}$$

where u_1 is either 1 or 3 according as n is even or odd:

$$T^{(2)}\left(n+2^k\right)=T^{(2)}(n)+u_1 u_2 2^{k-2}$$

where u_2 is either 1 or 3 according as $T^{(1)}(n)$ is even or odd,

$\cdots$

$$T^{(k)}\left(n+2^k\right)=T^{(k)}(n)+u_1u_2\cdots u_k$$

where u_k is either 1 or 3 according as $T^{(k-1)}(n)$ is even odd.

But $u_1u_2\cdots u_k=3^r$, and $T^{(k)}(n)=m$ is the resultant of L, while: $T^{(k)}\left(n+2^k\right)=m+3^r$ is the pre-resultant of Lx.

To apply Theorem A in a specific situation, suppose n is the smallest generator of L (length k) and the resultant of L is $m = T^{(k)}(n)$. Consider the string Lx. If the next member of the Collatz sequence $T(m)$ matches x, then n is the generator of Lx. If not, then add 2^k to n which (by Theorem A) will generate Lx with pre-resultant $T(m) = m + 3^r$ where r is the number of ones in L.

An example will show how this procedure works. Suppose we want to find the smallest generator of $L = 101^30^41 = 1011100001$. Start with $n = 1$ (2 if the leading element of L is zero). Then:

[9] $L \rightarrow$	1	0	1	1	1	0	0	0	0	1
[6] $C(1) \rightarrow$	1	2	1	2						
	$+\underline{2^3}$			$+\underline{3^2}$						
[5] $C(9) \rightarrow$	9			11	17	26	13			
	$+\underline{2^6}$						$+\underline{3^4}$			
[4] $C(73) \rightarrow$	73						94	47		
	$+\underline{2^7}$							$+\underline{3^4}$		
[3] $C(201) \rightarrow$	201							128	64	32
	$+\underline{2^9}$									$+\underline{3^4}$
[3] $C(713) \rightarrow$	713									113

Thus, 713 is the smallest generator of L, 113 is its pre-resultant, and $(3 \times 113 + 1)/2 = 170$ is its resultant.

Corollary: every finite string has a generator.

3. A Formula for the Resultant

One starts with the special case $L = 1^r0^s$, with n a generator of L Induction on r, starting with $s = 0$, shows that:

$$m=\frac{3^r n+3^r-2^r}{2^{r+s}} \tag{3.1}$$

which can be put in the form $m = \lambda n + d$ where $\lambda = 3^r/2^{r+s}$ and $d=\left[\left(\frac{3}{2}\right)^r-1\right]\Big/2^s$.

In general, if:

$$L=0^{s_0}1^{r_1}0^{s_1}1^{r_2}0^{s_2}1^{r_q}0^{s_q}$$

with $C_k(n)$ the Collatz sequence corresponding to L, let m_i be the member of the sequence that generates the substring of L beginning with 1^{r_i}. Then for each i, $1 \le i \le q$:

$$m_i=\lambda_i m_{i-1}+d_i \quad (1\le i\le q)$$

where $m_0 = n$, $\lambda_i = 3^{r_i}/2^{r_i+s_i}$, and $d_i = \left[\left(\frac{3}{2}\right)^{r_i} - 1\right] \Big/ 2^{s_i}$. Thus, performing the indicated substitutions, each m_i can be determined, and it follows that:

$$m_q = \lambda_q \lambda_{q-1} \lambda_{q-2} \cdots \lambda_1 n + \lambda_q \lambda_{q-1} \lambda_{q-2} \cdots \lambda_2 d_1 + \lambda_q \lambda_{q-1} \lambda_{q-2} \cdots \lambda_3 d_2 + \cdots + \lambda_q d_{q-1} + d_q$$

Since m_q is the resultant of the sequence corresponding to L and $\lambda_q \lambda_{q-1} \lambda_{q-2} \cdots \lambda_1 = \|L\|$, the final result is:

$$m = \|L\| n + \sum \left(\frac{3}{2}\right)^{u_i} \left(\frac{1}{2}\right)^{v_i} d_i \quad (1 \le i \le q) \tag{3.2}$$

where $u_i = r_{i+1} + r_{i+2} + \cdots + r_q$ and $v_i = s_{i+1} + s_{i+2} + \cdots + s_q$ ($u_q = v_q = 0$). This formula is a special case of a similar one obtained by Bohm and Sontacchi in 1978.

The validity of (3.2) can be checked, using the example in Section 1, where $L = 101^3 0^4 1$ and it was determined that $n = 713$ and $m = 170$. In this case, $q = 3$, $s_0 = s_q = 0$. First, compute u_i, v_i, and d_i: for each $i \le 3$):

$$u_1 = 4,\ v_1 = 4,\ u_2 = 1,\ v_2 = 0, \text{ and } u_3 = v_3 = 0$$

$$d_1 = 1/4,\ d_2 = 19/128, \text{ and } d_3 = 1/2.$$

Then:

$$\begin{aligned} m &= \frac{3^5}{2^{10}} n + \left(\frac{3}{2}\right)^4 \left(\frac{1}{2}\right)^4 d_1 + \left(\frac{3}{2}\right)^1 \left(\frac{1}{2}\right)^0 d_2 + d_3 \\ &= \frac{3^5}{2^{10}} n + \left(\frac{3}{2}\right)^4 \left(\frac{1}{2}\right)^4 \frac{1}{2^2} + \left(\frac{3}{2}\right)^1 \left(\frac{1}{2}\right)^0 \frac{19}{2^7} + \frac{1}{2} \\ &= \frac{3^5(713) + 3^4 + 3 \times 19 \times 2^2 + 2^9}{2^{10}} \\ &= \frac{173259 + 81 + 228 + 512}{1024} = \frac{174080}{1024} = 170 \end{aligned}$$

4. The Diophantine Equation

The above Equation (3.2) is equivalent to:

$$2^k m = 3^r n + 2^k Q$$

where Q is the summation term in (3.2). Since both $2^k m$ and $3^r n$ are positive integers, the term $2^k Q \equiv N$ is an integer and we obtain the well-known Diophantine equation for the $3x + 1$ problem:

$$2^k m - 3^r n = N \tag{4.1}$$

where N is the product of 2^k and the summation term in (3.2). Since 2^k and 3^r are relatively prime, (4.1) has infinitely many positive solutions of the form:

$$n = m_0 + 2^k t,$$

$$m = n_0 + 3^r t \quad (t = 0, 1, 2, \cdots)$$

where n_0 and m_0 are the smallest positive solutions of (4.1). Thus every zero-one string has infinitely many generators of the form listed above.

Moreover, by setting $m = n$, one obtains an explicit formula for the generator of a cycle:

$$n = \frac{2^k \sum \left(\frac{3}{2}\right)^{u_i} \left(\frac{1}{2}\right)^{v_i} d_i}{3^r - 2^k} \tag{4.2}$$

In general, the numbers in this equation are incredibly large, but if it could be shown that for $n \geq 3$ a prime divisor of $3^r - 2^k$ does not divide $\sum a^{u_i} b^{v_i} d_i$ evenly, a proof that nontrivial cycles do not exist is obtained. Note that if n is a generator of 1010… 10 of length $2q$ then $n = 1$ and $3^r - 2^k$ divides $\sum \left(\frac{3}{2}\right)^{u_i} \left(\frac{1}{2}\right)^{v_i} d_i$ evenly.

5. Result of Terras and Everett

Theorem A leads to a surprising relationship between all strings of length k and their least generators. It was discovered independently by Terras [3] and Everett [1] in their work on the $3x + 1$ problem in 1977-78. There are clearly 2^k possible zero-one strings of length k, so starting with the 4 strings of length 2, we can easily deduce their least generators (and resultants):

4 (00) 1; 2 (01) 2; 1 (10) 1; 3 (11) 8

Suppose the strings of length k are $L_1, L_2, L_3, \ldots, L_p$, $p = 2^k$, and we seek the generators of the strings of length $k + 1$, which are of the form $0L_i$ and $1L_i$, $1 \leq i \leq 2^k$. Those with leading term zero in L_i are obviously generated by the even integers 2, 4, 6, …, 2^{k+1}. The remaining strings with leading term one must have odd generators. Half of them correspond to strings ending in zero, the other half to strings ending in one.

Theorem B: The strings of length $k + 1$ are uniquely generated by the positive integers less than or equal to 2^{k+1}. Accordingly, there is a one-to-one correspondence between the strings of length $k + 1$ and their least generators.

Proof: A careful analysis of the previous discussion reveals that all the integers from 1 to 2^{k+1} have been accounted for.

Table 1 illustrates Theorem B, showing all the possible characteristic strings of lengths 3, 4 and 5 and the corresponding generators and resultants. At this point we use the notation nLm only when n is the unique least generator of the string L and m is its resultant.

6. Inequalities

Suppose nLm with L of length k, and $\|L\| = 3^r/2^k$. One obtains from (3.2):

$$\begin{aligned} m &= \|L\| n + \sum \left(\frac{3}{2}\right)^{u_i} \left(\frac{1}{2}\right)^{v_i} d_i \quad (1 \leq i \leq q) \\ &< \|L\| n + \sum \left(\frac{3}{2}\right)^{u_i} \left(\frac{1}{2}\right)^{v_i} \left(\frac{3}{2}\right)^{r_i} \left(\frac{1}{2}\right)^{s_i} < \|L\| n + \left(\frac{3}{2}\right)^r \sum \left(\frac{1}{2}\right)^i \\ &< \|L\| n + \left(\frac{3}{2}\right)^r \left(\frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^q}\right) < \|L\| n + \left(\frac{3}{2}\right)^r \end{aligned}$$

Table 1. List of Generators for Strings of Lengths 3, 4, and 5

L	n	m	L	n	m	L	n	m	L	n	m
000	8	1	0000	16	1	00000	32	1	10000	21	2
001	4	2	0001	8	2	00001	16	2	10001	5	2
010	2	1	0010	4	1	00010	8	1	10010	13	4
011	6	8	0011	12	8	00011	24	8	10011	29	26
100	5	2	0100	10	2	00100	20	2	10100	17	5
101	1	2	0101	2	2	00101	4	2	10101	1	2
110	3	4	0110	6	4	00110	12	4	10110	25	22
111	7	26	0111	14	26	00111	28	26	10111	9	26
			1000	5	1	01000	10	1	11000	3	1
			1001	13	8	01001	26	8	11001	19	17
			1010	1	1	01010	2	1	11010	11	10
			1011	9	17	01011	18	17	11011	27	71
			1100	3	2	01100	6	2	11100	23	20
			1101	11	20	01101	22	20	11101	7	20
			1110	7	13	01110	14	13	11110	15	40
			1111	15	80	01111	30	80	11111	31	242

Thus it follows that:

$$\|L\|n < m < \|L\|n + \left(\frac{3}{2}\right)^r \tag{6.1}$$

Due to the term $\left(\frac{3}{2}\right)^r$ one cannot conclude that $m < n$ if $\|L\| < 1$. However this is true if $\|L\|$ is small enough, as one might expect. In fact, if L has as many zeros as ones, then $m < n$, as will be shown. A cyclic string with $m = n$ must have $\|L\| < 1$, due to (6.1). Thus $3^r < 2^k$. Taking the logarithm to the base 2, we have:

$$k > r\log_2 3 \tag{6.2}$$

Define the constants $\omega_1 = \log_2 3 \approx 1.584962501$ and: $\omega_0 = \log_2 3 - 1 \approx 0.584962501$. Since $k = r + s$, we have for any string L such that $\|L\| < 1$:

$$s > r\omega_0 \quad \text{and} \quad k > r\omega_1 \tag{6.3}$$

7. The Ratio s/r and Finite Stopping Time for Certain Collatz Sequences

The numerical order relationship between r and s sometimes determines whether the generator of a string has a finite stopping time.

Theorem C: Suppose that $s \geq r$ in the Collatz sequence mLn where $m \geq 3$. Then $m < n$ and n has a finite stopping time.

Proof: Define the sequence $\{v_i\}$, $1 \le i \le k$, as follows: if n_i is even, set $v_i = 1/2$; if n_i is odd, set $v_i = 2$. Then for each i, if n_i is even $n_{i+1} = v_i n_i$, and if n_i is odd $n_{i+1} = (3n_i + 1)/2 < 2n_i = v_i n_i$. It follows that:

$$n_2 \le v_1 n_1, \quad n_3 \le v_2 n_2 \le v_2 v_1 n_1, \quad n_4 \le v_3 n_3 \le v_3 v_2 v_1 n_1, \quad \cdots$$

and:

$$m = n_{k+1} \le v_1 v_2 v_3 \cdots v_k n_1 < \left(\frac{1}{2}\right)^s 2^r n \le \left(\frac{1}{2}\right)^s 2^s n = n$$

A more general result can be shown by substantially the same argument.

Theorem D: suppose that nLm, $m > 1$, and for some positive integer p, $s = r - p$. Then if $r \ge 3p$, n has a finite stopping time. The number 3 is the least integer possible.

Proof: Set $\varepsilon = 2^{-63}$. Let $C_k(n)$ be the Collatz sequence corresponding to the string L. If $n_i < \frac{1}{\varepsilon}$ for at least one i, $1 \le i \le k$, then $n_i < 2^{63} < 20 \times 2^{58} = n^*$ and n has a finite stopping time. Thus it may be assumed that $n_i > \frac{1}{\varepsilon}$ for all i, or $1 < \varepsilon n_i$. Thus:

$$\frac{3n_i + 1}{2} < \frac{3n_i + \varepsilon n_i}{2} = \frac{3 + \varepsilon}{2} n_i$$

Define v_i as follows: If n_i is even take $v_i = 1/2$; if n_i is odd, $v_i = \frac{3+\varepsilon}{2}$. Then it follows that for $1 \le i \le k$, $n_{i+1} \le v_i n_i$. Hence, as in Theorem C, $m = n_{k+1} < v_1 v_2 \cdots v_k n_1$ and we obtain:

$$m < \left(\frac{1}{2}\right)^s \left(\frac{3+\varepsilon}{2}\right)^r n = \left(\frac{1}{2}\right)^{r-p} \left(\frac{3+\varepsilon}{2}\right)^r n = 2^p \left(\frac{3+\varepsilon}{4}\right)^r n$$

$$\le 2^{r/3} \left(\frac{3+0.1}{4}\right)^r n = \left[\sqrt[3]{2}\left(\frac{31}{40}\right)\right]^r n < n$$

Corollary: suppose that nLm, $m > 1$, and $s/r \ge 2/3$. Then n has finite stopping time.

Proof: If $s \ge r$, then Theorem C applies and $m < n$. Otherwise, $s < r$ and there exists a positive integer p such that $s = r - p$. By hypothesis:

$$\frac{r-p}{r} \ge \frac{2}{3}$$

$$1 - \frac{p}{r} \ge \frac{2}{3}$$

from which it follows that $r \ge 3p$. Therefore by Theorem D n has finite stopping time.

It might be imagined that if a cycle exists the effect of the odd members in the cycle essentially cancels that of the even members, and s is roughly equal to r. As a matter of fact, the number of even members is about 63% the number of odd members: if $C_k(n)$ is a nontrivial cycle, then $\|L\| < 1$ and $3^r < 2^k$, which leads to $k < r\log_2 3 = r\omega_1$ where $\omega_1 = \log_2 3 = 1.584962\cdots$. Also, the above corollary implies that $s/r < 2/3$, and $k < 5/3$. Thus:

$$\omega_1 < \frac{k}{r} < \frac{5}{3} \quad \text{and} \quad 0.585 \approx \omega_0 < \frac{s}{r} < \frac{2}{3} \approx 0.667 \tag{7.1}$$

Using a theorem of Crandall [4], Lagarias showed in [2] that a non1rivial cycle, if it exists, must have length at least $k = 275{,}000$. More recently a paper by Halbeisen and Hungerbühler [5] in 1997 showed that $k > 102{,}225{,}496$. The article by Eliahou [6] established k in terms of the minimum element of a cycle. Brox [7] showed that certain cycles occur only finitely many times.

8. Another Formula for the Resultant

Define:

$$v_i = u_i = \frac{1}{2}$$

if n_i is even:

$$v_i = \frac{3}{2}\left(1 + \frac{1}{3n_i}\right) = u_i\left(1 + \frac{1}{3n_i}\right)$$

if n_i is odd:

Then $n_{i+1} = u_i n_i$ if n_i is even, and $n_{i+1} = \frac{3n_i + 1}{2} = u_i\left(1 + \frac{1}{3n_i}\right)n_i = v_i n_i$ if n_i is odd. Thus:

$$m = v_1 v_2 v_3 \cdots v_k n$$

The product of the factors u_i is $\|L\|$ and the factor $1 + \frac{1}{3n_i}$ appears only when n_i is odd. Hence:

$$m = \|L\| n \prod \left(1 + \frac{1}{3n_j}\right) \tag{8.1}$$

where the product is taken over all the odd members n_j of the sequence, $1 \le j \le r$. (If there are no odd members, then we interpret (8.1) as simply $m = \|L\| n$, which is trivially true).

Solving for $||L||$ in (6.1) produces an interesting formula for the norm:

$$\|L\| = m\left[n \prod \left(1 + \frac{1}{3n_j}\right)\right]^{-1} \quad (1 \le j \le r) \tag{8.2}$$

where the product is taken over all the odd members of the sequence.

For a numerical test of (8.2), consider $7(1^3 0^1 1^1 0^2 1^1)8$. The odd members of the designated Collatz sequence are 7, 11, 17, 13, and 5. The expression inside the brackets in (8.2) equals:

$$7\left(1 + \frac{1}{21}\right)\left(1 + \frac{1}{33}\right)\left(1 + \frac{1}{51}\right)\left(1 + \frac{1}{39}\right)\left(1 + \frac{1}{15}\right)$$

$$= 7 \times \frac{22}{21} \times \frac{34}{33} \times \frac{52}{51} \times \frac{40}{39} \times \frac{16}{15} = \frac{2 \times 2 \times 4 \times 8 \times 16}{3 \times 3 \times 3 \times 3 \times 3} = \frac{2^{11}}{3^5}$$

from which it follows that $m \cdot \frac{3^5}{2^{11}} = 8 \cdot \frac{3^5}{2^{11}} = \frac{3^5}{2^8}$, the correct value for $\|L\|$.

An inequality can be established using (6.1). Suppose that p is an upper bound for the odd members of the Collatz sequence nLm. That is, $n_j < p$ for all j. Then:

$$\prod\left(1+\frac{1}{3n_j}\right) > \prod\left(1+\frac{1}{3n_j}\right) = \left(1+\frac{1}{3p}\right)^r \quad (1 \le j \le r)$$

and it follows from (6.1) that:

$$m > \|L\| n (1+q)^r \tag{8.3}$$

where $q = 1/3p$, slightly stronger than (6.1).

9. Relation between k and n^*

A formula for the product term in (8.1) will be obtained for all Collatz sequences that do not have a finite stopping time. Starting with (8.1):

$$m = \|L\| n \prod\left(1+\frac{1}{3n_j}\right).$$

which is equivalent to:

$$\frac{2^k m}{n} = \prod\left(3+\frac{1}{n_j}\right)$$

Let c be the product term in the previous equation, and consider the continuous function:

$$f(x) = (3+x)^r$$

for fixed r. Then:

$$c > 3^r = f(0).$$

Also, for Collatz sequences with no finite stopping time and having r odd members, n_j ($1 \le j \le r$), we must have $n_j \ge n^*$ for all j, and thus:

$$c = \prod\left(3+\frac{1}{n_j}\right) < \prod\left(3+\frac{1}{n^*}\right) = 3+\frac{1}{n^*} = f\left(\frac{1}{n^*}\right)$$

Accordingly, there exists $0 < \varepsilon < \frac{1}{n^*}$ such that $f(\varepsilon) = c$. Or:

$$3+\frac{1}{n^*} > (3+\varepsilon)^r = \prod\left(3+\frac{1}{n_j}\right) = \frac{2^k m}{n} > 2^k$$

and it follows that:

$$2^k < \left(3+\frac{1}{n^*}\right)^r \tag{9.1}$$

10. The $3x + 1$ Cycles Conjecture

The result (8.1) with $m = n$ yields the relation:

$$\prod\left(3+\frac{1}{n_j}\right) = 2^k \tag{10.1}$$

Since all the factors of the above product are fractional, it seems unlikely that the product is a power of 2.

It was proved in the preceding section that the product term equals $(3+\varepsilon)^r$. Then for nontrivial cycles one obtains:

$$(3+\varepsilon)^r = 2^k$$

It might be imagined that $(3+\varepsilon)^r$ is not an integer since ε is a very small positive number, thus providing an immediate proof that cycles do not exist. But this is false in general, for the equation:

$$(3+x)^r = 2^u$$

has a positive solution in x for certain values of u. Suppose u lies between $r\omega_1$ and $2r$. Then:

$$(3+x)^r > 2^{r\omega_1} = 3^r$$

$$3+x > 3$$

$$x > 0$$

Also:

$$(3+x)^r < 2^{2r}$$

$$3+x < 4$$

$$x < 1$$

If $u = k$ then one observes that the above condition for u is satisfied for a nontrivial cycle and there is a solution for:

$$(3+x)^r = 2^k \quad (0 < x < 1)$$

For example:

$$(3.17480210389\cdots)^9 = 2^{15}$$

The question is whether $x = \varepsilon$. In fact, the above analysis proves that ε must have the exact value:

$$\varepsilon = 2^{k/r} - 3$$

and it remains to obtain a contradiction.

It is important in our analysis to show that ε is irrational. Note that for a nontrivial cycle:

$$3+\varepsilon = 2^{k/r}$$

where $\omega_1 < k/r < 5/3$. Thus:

$$\omega < 2^{k/r} < 5/3$$

so that $(2^k)^{1/r}$ is not an integer. It is well known that, accordingly, $(2^k)^{1/r}$ is irrational, proving that $3+\varepsilon$ and ε are irrational.

11. Probabilistic Proof That Cycles Do Not Exist

Consider the product:

$$(3+\varepsilon)^{r-1}(3+\varepsilon)=(3+\varepsilon)^{r}$$

(which equals 2^k if a nontrivial cycle exists). The decimal form of the left side is:

$$(a+0.a_1a_2\cdots)(3+0.0.b_1b_2\cdots)$$

where a is a large positive integer and the first 18 decimals of $0.b_1b_2\cdots$ are zero. (It may be assumed that $(3+\varepsilon)^{r-1}$ is irrational, for if not, then $(3+\varepsilon)^{r}$ is irrational and $\neq 2^k$, ending the proof). Multiplying out, one obtains:

$$3a+3(0.a_1a_2\cdots)+a(0.0.b_1b_2\cdots)+(0.c_1c_2\cdots)$$

Let b be the integral part of $3(0.a_1a_2\cdots)$ and c the integral part of $a(0.0.b_1b_2\cdots)$. Then we obtain for certain decimals:

$$3a+b+(0.d_1d_2\cdots)+c+(0.e_1e_2\cdots)+(0.c_1c_2\cdots)$$

The fractional part of the final result equals:

$$(0.d_1d_2\cdots)+(0.e_1e_2\cdots)+(0.c_1c_2\cdots)=u_0.u_1u_2\cdots$$

The final decimal $u_0.u_1u_2\cdots$ equals an integer only if $u_i = 9$ for all $i \geq 1$ and, accordingly, equals:

$$u_0.9999\cdots9\cdots=u_0+1$$

The probability that the first p decimals equal 9 is $1/10^p \approx 0$, and the probability that all the decimals equal 9 is virtually zero. The probability of the existence of a nontrivial cycle is thus virtually impossible.

12. Infinite Strings

The essence of the $3x + 1$ problem involves infinite strings. Several elementary results concerning infinite strings can be established. First observe that not all infinite strings have generators; simple examples are:

$$11111\cdots,\ \ 00000\cdots,\ \ 10101010\cdots$$

In order for an infinite string to have a generator, it must at least have a mixture of zeros and ones that continues indefinitely. One somewhat surprising fact is:

Theorem E: if a generator of an infinite string exists, it is unique.

Proof: suppose L is an infinite string having n as a generator. There must at some point be an odd member of the sequence and there is no loss if it is assumed that n is odd. If a second generator $n' = n + p$ exists, let 2^q be the highest power of 2 that divides p. and perform the Collatz algorithm on n' repeatedly: $n_1' = n_1 + 2^q p'$, $n_2' = \left[3\left(n_1 + 2^q p'\right)+1\right]/2 = n_2 + 3\cdot 2^{q-1} p'$, $\cdots$, $n_q' = n_q + 3^q p'$. The next term will be a non-integer and n' cannot be a generator.

Suppose that C is a Collats sequence with infinitely many terms different from 1. Its corresponding zero-one string is L, consisting of infinitely many zeros and ones. The generator of C is n, a finite positive integer. As we progress through the algorithm to obtain successive values of C, no discrepancy ever exists, an unlikely circumstance (see Section 4). But this has, as yet, resisted proof.

If we assume that no nontrivial cycles exist, there can be no upper bound for any Collatz sequence, hence $\lim_{i\to\infty} n_i = \infty$. As a matter of fact, it follows that every Collatz sequence that does not converge to 1 contains an increasing subsequence of the form $n_0 + 2^k t_i$ where n_0 is a positive integer, k is an arbitrarily large integer, and $\{t_i\}$ is an increasing sequence. Suppose C is a collatz sequence that does not converge to 1, and let L^* be its characteristic string, with odd generator n.

Theorem F: The string L^* is of the form $K_1LK_2 \cdots K_{i-1}LK_i \cdots$ where L is a finite substring repeated infinitely often of arbitrarily large length and with beginning element 1, and K_i is a finite substring of L^* or the empty set if two substrings L juxtoposed.

Proof (by induction on k, the length of L): to obtain a string L of length $k = 1$, take the first element 1 of L^* and observe that neither 000… 0… nor 111… 1… have generators. Thus a repeated element 1 must occur infinitely often, and we take $L = 1$. Assume that a finite string L of length k beginning with 1 occurs infinitely often. Let the element immediately following L be denoted x, which must be repeated infinitely often, producing the string Lx of length $k + 1$, where Lx must occur infinitely often.

Consider the string $L^* = K_1LK_2 \cdots K_{i-1}LK_i \cdots$ with generator n. Let n_0 be the least generator of L From a previous result, the generator of the string from this point on must be of the form:

$$n_0 + 2^k t_1$$

where k is the length of L and t_1 is a positive integer. For the subsequent repeated strings which are identical to L, their generators must also be of the form $n_0 + 2^k t_i$, where t_i is a positive integer. We can assume that the sequence $\{t_i\}$ is increasing since it must contain one that is. Thus, we have proved the previous claim made above.

The resultants of these repeating strings must have the form $m_0 + 3^r t_i$ Thus If n_i and m_i are the elements of C immediately preceding and following a string L, the limit of their quotients converges to the norm of L:

$$\lim_{i\to\infty} \frac{m_i}{n_i} = \frac{m_0 + 3^r t_i}{n_0 + 2^k t_i} = \frac{3^r}{2^k} = \|L\|$$

The product Formula (8.1) leads to another relation involving the string L^*. Again, assuming nontrivial cycle exists, it was determined that $3^r c$ is not a power of 2. Hence, from (7.1):

$$\frac{m}{n} \neq 2^p$$

where, since n_0 is odd, then $n_0 + 2^k t_i$ is odd and both m and n may be assumed to be odd integers with m much larger than n. It may be possible to derive a contradiction here.

Another possibility derives from an inequality, also with m and n odd and m much larger than n.

Let the string corresponding to the Collatz sequence beginning with $n = n_0 + 2^{k'} t_0$ and ending with $m = n_0 + 3^{r'} t_i$ be *K* of, length *k* with *r* ones and *s* zeros. If $\|K\| < 1$ (or $k > r\omega_1$), one obtains from the inequality (4.2):

$$3^{r'} t_i > 3\left(\frac{3}{2}\right)^r + \|K\| 2^{k'} t_0 \tag{12.1}$$

where *r'* and *k'* are the indices corresponding to *L*. This could lead to a contradiction.

13. Conclusion

The main goal of this article was to find an elementary proof involving zero-one strings to show that nontrivial cycles do not exist. A proof no doubt depends ultimately on advanced results in number theory, as yet not identified. As shown by many authors, such as the group here in the references, a proof must involve methods defined by multiple new or old results involving intricate details. Of course, the problem as a special case of the original $3x + 1$ problem could be undecidable, not yet determined. It can only be hoped that untried methods of proof showing up in the literature can trigger the discovery of a connection to known results that finally provides a contradiction.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Everett, C.J. (1977) Iteration of the Number Theoretic Function $f(2n) = n$, $f(2n + 1) = 3n + 2$. *Advances in Mathematics*, **25**, 42-45. https://doi.org/10.1016/0001-8708(77)90087-1

[2] Lagarias, J.C. (2010) The Ultimate Challenge: The 3x + 1 Problem. American Mathematical Society, Providence. https://doi.org/10.1090/mbk/078

[3] Terras, R. (1976) A Stopping Time Problem on the Positive Integers. *Acta. Arithmetica*, **30**, 241-252. https://doi.org/10.4064/aa-30-3-241-252

[4] Crandall, R.E. (1978) On the $3x + 1$ Problem. *Mathematics of Computation*, **32**, 1281-1292. https://doi.org/10.2307/2006353

[5] Halbeisen, L. and Hungerbühler, N. (1997) Optimal Bounds for the Length of Rational Collatz Cycles. *Acta Arithmetica*, **78**, 227-239. https://doi.org/10.4064/aa-78-3-227-239

[6] Eliahon, S. (1993) The $3x + 1$ Problem: New Lower Bounds on Nontrivial Cycle Lengths. *Discrete Mathematics*, **118**, 45-56. https://doi.org/10.1016/0012-365X(93)90052-U

[7] Brox, T. (2000) Collatz Cycles with Few Descents. *Acta Arithmetica*, **92**, 181-188. https://doi.org/10.4064/aa-92-2-181-188