Collatz High Cycles Do Not Exist

Kevin Knight

[a] *Threeven Labs, 134 Roma Ct., Marina del Rey, 90292, CA, USA*

Abstract

The Collatz function takes odd n to $(3n+1)/2$ and even n to $n/2$. Under the iterated Collatz function, every positive integer is conjectured to end up in the trivial cycle 1-2-1. Two types of cycles are of special interest. Consider the set S consisting of the smallest members of all cycles containing the same number of odd terms. The *circuit* contains the smallest member of S, while the *high cycle* contains the largest. It is known that no circuits of positive integers exist (except 1-2-1); this paper shows that there are likewise no high cycles of positive integers.

Keywords: number theory
2000 MSC: 11A99

1. Introduction

The Collatz function is

$$T(n) = \begin{cases} \dfrac{3n+1}{2}, & \text{if } n \text{ is odd;} \\ \dfrac{n}{2}, & \text{otherwise.} \end{cases}$$

Iterating this function famously yields interesting sequences. For example: $61 \to 92 \to 46 \to 23 \to 35 \to 53 \to 80 \to 40 \to 20 \to 10 \to 5 \to 8 \to 4 \to 2 \to 1 \to 2 \to 1 \to \ldots$

The Collatz conjecture posits that every positive integer eventually reaches 1. The conjecture was verified in 2021 for all $1 \leq n \leq 10^{21}$ [2], but it has not yet been proven or refuted.

Email address: `kevin.crawford.knight@gmail.com` (Kevin Knight)

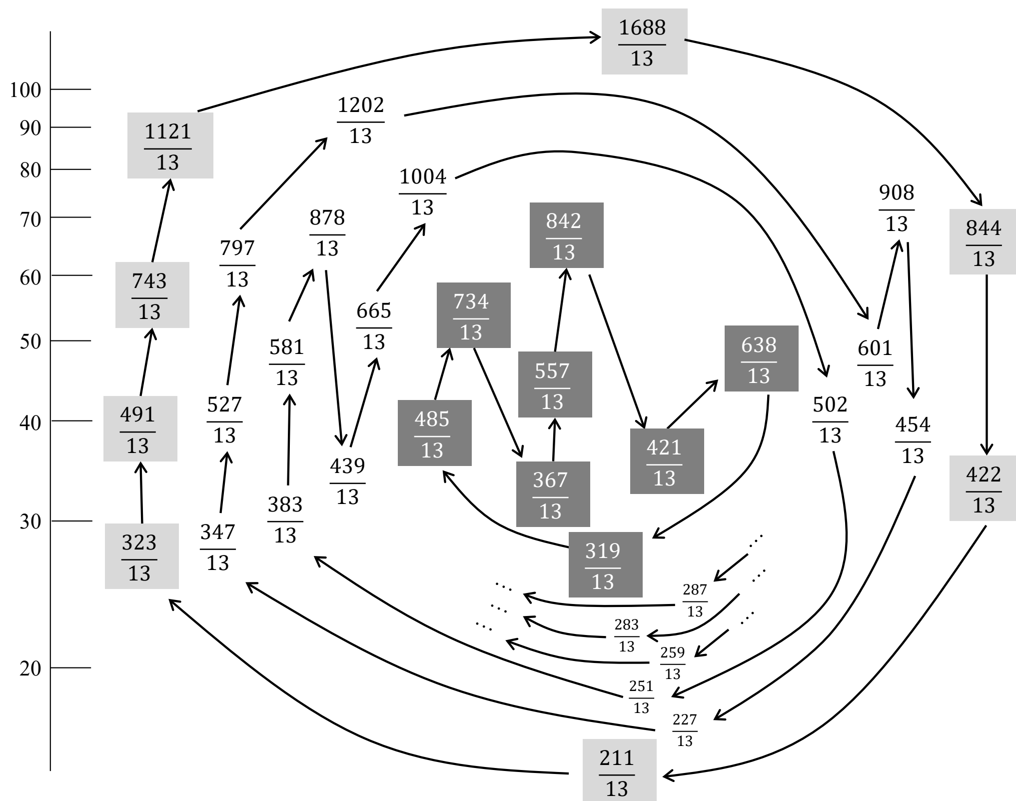

Figure 1: All of the rational Collatz cycles of length $k = 8$ with $x = 5$ odd terms. Of special note are the outermost cycle (the *circuit*), in light gray, and the innermost cycle (the *high cycle*), in dark gray.

If $T^i(n) = n$ for some positive integer n, there is a Collatz *cycle* of length i whose first term is n. The only known T-cycle is the trivial one: $1 \to 2 \to 1$.

The unresolved Weak Collatz conjecture claims that there are no positive integer cycles. (It is called "weak" because it allows for the possibility of a start number that diverges to infinity, never reaching 1.)

We can also consider cycles with rational values. Figure 1 shows all cycles of length $k = 8$ with $x = 5$ odd terms. Considering the term $\frac{211}{13}$ to be odd (by its numerator), the next term produced by the Collatz process is $(3(\frac{211}{13}) + 1)/2 = \frac{323}{13}$ (also odd).

For any k and x, the outermost and innermost cycles are of special interest. Let S contain the smallest members of each cycle (here, $\frac{211}{13}$, $\frac{227}{13}$, $\ldots$ $\frac{319}{13}$).

The outermost cycle (called the *circuit*) contains the lowest member in S, while the innermost cycle (called the *high cycle*) contains the highest member in S.

Steiner [11] showed that no positive integer circuits exist for any k, x (except $k = 2, x = 1$). The current paper shows likewise that no high cycles of positive integers exist.

2. Parity vectors

Associated with every cycle member m is a unique *parity vector* that records the even (0) and odd (1) terms obtained from k iterations of the Collatz function.

Example 2.1 (Cycle member to parity vector)**.** *Starting with* $m = \frac{211}{13}$*, we encounter* 5 *odd terms followed by* 3 *even terms, so the corresponding parity vector is* 11111000*.*

Likewise, given a parity vector $\mathbf{v}$, we can compute a cycle member m.

Example 2.2 (Parity vector to cycle member)**.** *Applying* $\mathbf{v} = 110$ *to arbitrary start number* n *yields* $[3\frac{3n+1}{2} + 1]/2/2 = \frac{9}{8}n + \frac{5}{8}$*. To make a cycle, we set* $n = \frac{9}{8}n + \frac{5}{8}$*. Solving this equation yields* $n = -5$*, which is part of the* $-5 \to -7 \to -10 \to -5$ *cycle.*

Generalizing this idea, Böhm and Sontacchi [4] give the cycle member associated with arbitrary parity vector $\mathbf{v}$ as

$$f(\mathbf{v}) = \frac{\sum_{i=0}^{x-1} 2^{d_i(\mathbf{v})}\, 3^{x-i-1}}{2^k - 3^x}, \tag{2.1}$$

where $d_0(\mathbf{v}) < \ldots < d_{x-1}(\mathbf{v})$ are the (zero-based) indices of 1s in $\mathbf{v}$. Since we are only interested in Collatz cycles with positive terms, the denominator must be positive, so $k > x \log_2 3$.

The Weak Collatz Conjecture can be restated: "Is there a parity vector $\mathbf{v}$ such that $f(\mathbf{v})$ is an integer greater than 2?"

It will sometimes be useful to refer to the numerator only:

$$g(\mathbf{v}) = \sum_{i=0}^{x-1} 2^{d_i(\mathbf{v})}\, 3^{x-i-1}. \tag{2.2}$$

Note that rotating a parity vector produces another term in the same rational cycle.

Example 2.3 (Rotation). *$f(100) = \frac{1}{5}$, $f(001) = \frac{4}{5}$, and $f(010) = \frac{2}{5}$. These are all entry points into the $\frac{1}{5} \to \frac{4}{5} \to \frac{2}{5} \to \frac{1}{5}$ cycle.*

Finally, we can restrict ourselves to aperiodic parity vectors. Consider $f(1101011010) = \frac{23}{5}$, which is part of the double cycle $\frac{23}{5} \to \frac{37}{5} \to \frac{58}{5} \to \frac{29}{5} \to \frac{46}{5} \to \frac{23}{5} \to \frac{37}{5} \to \frac{58}{5} \to \frac{29}{5} \to \frac{46}{5} \to \frac{23}{5}$. Had these been integers, the aperiodic $f(11010)$ would also be an integer.

3. No integer circuits

The parity vector $\mathbf{v_c} = 1^x 0^{k-x}$ is associated with the smallest member of the (k, x)-circuit. Following Equation 2.1,

$$f(\mathbf{v_c}) = \frac{\sum_{i=0}^{x-1} 2^i 3^{x-i-1}}{2^k - 3^x}.$$

We provide a novel proof that (non-trivial) Collatz circuits do not exist, which is simpler than those of Steiner [11] and Rozier [10].

Lemma 3.1. *$f(\mathbf{v_c})$ simplifies to $\dfrac{3^x - 2^x}{2^k - 3^x}$.*

Proof. By induction, omitted. □

Lemma 3.2. *If there is a non-trivial positive integer circuit of length k with x odd terms, then $2 \cdot (\frac{3}{2})^x \geq 2^{k-x} - 1$.*

Proof. This follows from the fact that the putative circuit's smallest member is at least 1.

$$\begin{aligned}
\frac{3^x - 2^x}{2^k - 3^x} &\geq 1 \\
3^x - 2^x &\geq 2^k - 3^x \\
2 \cdot 3^x &\geq 2^k + 2^x \\
2 \cdot (\frac{3}{2})^x &\geq 2^{k-x} - 1.
\end{aligned}$$

□

Lemma 3.3 (Ellison [7])**.** *For integers k and x, with $x > 17$, we have $2^k - 3^x > 2.56^x$.*

Proof. We start with Ellison's original bound (for $k > 27$) and use the fact that $k > x \log_2 3$ in positive cycles.

$$2^k - 3^x > \frac{2^k}{e^{k/10}} > 1.8^k > 1.8^{x \log_2 3} > 2.56^x, \text{ for } x > 17. \qquad \square$$

Theorem 3.4 (Steiner, 1977)**.** *Non-trivial Collatz circuits do not exist.*

Proof. Assume $f(\mathbf{v_c}) = \dfrac{3^x - 2^x}{2^k - 3^x}$ is a positive integer. Then the following are also positive integers:

$$\frac{3^x - 2^x}{2^k - 3^x} + 1,$$

$$\frac{3^x - 2^x}{2^k - 3^x} + \frac{2^k - 3^x}{2^k - 3^x},$$

$$\frac{2^k - 2^x}{2^k - 3^x},$$

$$\frac{2^x(2^{k-x} - 1)}{2^k - 3^x}, \text{ and } \frac{2^{k-x} - 1}{2^k - 3^x}.$$

We can remove the 2^x in the last step without affecting the expression's purported integrality, since the denominator is odd.

However, the last term is actually less than one; we invoke Lemmas 3.3 and 3.2 to show that the denominator outstrips the numerator. For $x > 17$,

$$2^k - 3^x > 2.56^x > 2 \cdot (1.5)^x \geq 2^{k-x} - 1.$$

$2^{k-x} - 1 = 0$ only when $k = 2, x = 1$; setting aside this case, we obtain the contradiction $0 < \frac{2^{k-x}-1}{2^k - 3^x} < 1$. Cases of $x \leq 17$ are handled because any non-trivial Collatz cycle must contain many millions of odd terms [6]. $\square$

To date, all no-circuit proofs depend on deep lower bounds for $2^k - 3^x$, such as Ellison's, which are all derived from Alan Baker's pioneering transcendental number theory work [1]. As Jeffrey Lagarias [9] remarks, "The most remarkable thing about Theorem [...] is the weakness of its conclusion compared to the strength of the methods used in its proof."[1]

4. The high cycle

In this section, we characterize the Collatz high cycle.

Definition 4.1. *Let* $\mathbf{v_h}$ *be a* (k,x)*-parity vector with 1s indexed by* $d_i(\mathbf{v}) = \lfloor \frac{k}{x} i \rfloor$*, for* $0 \le i \le x-1$.

Example 4.2. *For* $k = 8, x = 5$, $\mathbf{v_h} = 11011010$.

Example 4.3. *For* $k = 21, x = 13$, $\mathbf{v_h} = 110110101101101011010$.

The vector $\mathbf{v_h}$ is the well-known *(upper) Christoffel word* [3], constructed so that its 1s are roughly evenly spread among its 0s. The cycle member corresponding to $\mathbf{v_h}$ is

$$f(\mathbf{v_h}) = \frac{\sum_{i=0}^{x-1} 3^{x-i-1} 2^{\lfloor \frac{k}{x} i \rfloor}}{2^k - 3^x}.$$

Now we list a number of properties of $\mathbf{v_h}$.

First, $f(\mathbf{v_h})$ is the smallest member of its cycle.

Theorem 4.4 (Halbeisen and Hungerbühler [8])**.** *For every* (k,x) *parity vector* $\mathbf{v}$ *that is a rotation of* $\mathbf{v_h}$*, we have* $f(\mathbf{v_h}) \le f(\mathbf{v})$.

[1]Note that certain circuits can be ruled out without these deep bounds. For example, if k and x are not co-prime, then we can factor $2^k - 3^x$. E.g., $(2^{k/2} + 3^{x/2})(2^{k/2} - 3^{x/2}) > 2^{k/2} > 2^{x log 3/2} > 1.72^x > 2^{k-x} - 1$. Or for the case of $\mathbf{v_c} = 1111111100000$, it follows that if $f(\mathbf{v_c}) = \dfrac{3^8 - 2^8}{2^{13} - 3^8}$ is an integer, then so is $31 f(\mathbf{v_c}) + 206$, but it happens that $31 \dfrac{3^8 - 2^8}{2^{13} - 3^8} + 206 \dfrac{2^{13} - 3^8}{2^{13} - 3^8} = \dfrac{3^{12}}{2^{13} - 3^8}$, which is not integral.

Remark 4.5. *It is known that a Christoffel word* $\mathbf{v_h}$ *is at a lexicographic extreme among all its rotations [3]. However, lexicographic ordering does not always coincide with f-ordering. For example,* $f(11110010) = \frac{259}{13}$, $f(11100101) = \frac{395}{13}$, *and* $f(10111100) = \frac{341}{13}$. *Of course, lexicographic ordering correlates roughly with f-ordering, as left-heavy vectors tend to be associated with smaller cycle members.*

Theorem 4.6. *Among all the (k, x) upper Christoffel words, $f(\mathbf{v_h})$ is maximized when $k = \lceil x \log_2 3 \rceil$.*

Proof. Let $\mathbf{v_h}$ be an upper Christoffel word with $k = \lceil x \log_2 3 \rceil$. Let $\mathbf{u}$ be a longer $(k+1, x)$-high-cycle parity vector. Replacing k by $k+1$ increases the numerator of $f(\mathbf{v})$ by no more than a factor of 2, while increasing the denominator by more than a factor of 2.

$$
\begin{aligned}
f(\mathbf{u}) &= \frac{\sum_{i=0}^{x-1} 3^{x-i-1} 2^{\lfloor \frac{k+1}{x} i \rfloor}}{2^{k+1} - 3^x} = \frac{\sum_{i=0}^{x-1} 3^{x-i-1} 2^{\lfloor \frac{k}{x} i + \frac{i}{x} \rfloor}}{2^{k+1} - 3^x} \\
&< \frac{\sum_{i=0}^{x-1} 3^{x-i-1} 2^{\lfloor \frac{k}{x} i \rfloor + 1}}{2^{k+1} - 2 \cdot 3^x} = \frac{2 \cdot \sum_{i=0}^{x-1} 3^{x-i-1} 2^{\lfloor \frac{k}{x} i \rfloor}}{2(2^k - 3^x)} = f(\mathbf{v}). \qquad \square
\end{aligned}
$$

Similar reasoning holds if $\mathbf{u}$ is a $(k+n, x)$ upper Christoffel word, for any $n \geq 1$.

Next, we confirm the reader's suspicion that $f(\mathbf{v_h})$ is indeed a member of the high cycle. Let S contain the smallest members of each (k, x)-cycle.

Theorem 4.7 (Halbeisen and Hungerbühler [8])**.** *Let* $\mathbf{v}$ *be a parity vector associated with the smallest member of any (k, x)-cycle, and let* $\mathbf{v_h}$ *be an upper Christoffel word with $k = \lceil x \log_2 3 \rceil$. Then,* $f(\mathbf{v_h}) \geq f(\mathbf{v})$.

Remark 4.8 (Tangential)**.** *This theorem is useful for proving that any Collatz cycle must be long. Just for example, it is easy to verify empirically that* $f(\mathbf{v_h}) < 10^{13}$ *for all* $x < 10{,}000{,}000$, *implying that* every *cycle (not just the high cycle) with fewer than ten million odd terms has some member less than* 10^{13}. *Since no Collatz counter-examples exist among the first* 10^{20} *positive integers [2], a purported Collatz cycle must have more than ten million odd terms. Stronger results have been obtained through analytical means [8]. Of course,* $f(\mathbf{v_h})$ *will outstrip any finite confirmation, after some x, because* $O(3^x x)$ *exceeds* $O(2^k - 3^x)$.

Recall the cycle member corresponding to $\mathbf{v_h}$.

$$f(\mathbf{v_h}) = \frac{\sum_{i=0}^{x-1} 3^{x-i-1} 2^{\lfloor \frac{k}{x} i \rfloor}}{2^k - 3^x},$$

Due to the floor function, we cannot simplify this expression as we did for $f(\mathbf{v_c})$. Its value can only be partially wrangled through upper and lower bounds.

Theorem 4.9 (Halbeisen and Hungerbühler [8])**.**

$$\frac{3^x(x/20)}{2^{\lceil x \log_2 3 \rceil} - 3^x} < f(\mathbf{v_h}) < \frac{3^x(7x/10)}{2^{\lceil x \log_2 3 \rceil} - 3^x}.$$

These bounds can be improved from $(x/20, 7x/10)$ to $(x/6, x/2)$; we omit the proof.

5. No integer high cycles

In this section, we prove our main result that members of the (non-trivial) high cycle are not integers. To do this, we need two more facts about Christoffel words. First, $f(\mathbf{v_h^R})$ is a member of the same cycle as $f(\mathbf{v_h})$.

Theorem 5.1 (Cohn [5])**.** *The reverse* $\mathbf{v_h^R}$ *of an upper Christoffel word* $\mathbf{v_h}$ *is also a rotation of it.*

Therefore, $f(\mathbf{v_h})$ and $f(\mathbf{v_h^R})$ are in the same cycle. Indeed, $f(\mathbf{v_h^R})$ is the largest member of the cycle, though we do not need this fact; also for interest, the left-rotation distance is the multiplicative inverse of x modulo k.

Theorem 5.2 (Berstel et al [3])**.** *For aperiodic* $\mathbf{v_h}$*, we have* $\mathbf{v_h} = 1\mathbf{u}0$*, and* $\mathbf{v_h^R} = 0\mathbf{u}1$*.*

Example 5.3. $\mathbf{v_h} = 11011010$*,* $\mathbf{v_h^R} = 01011011$*, and* $\mathbf{u} = 101101$*. Incidentally,* $\mathbf{u}$ *is always a palindrome, though we do not need this fact.*

Theorem 5.4 (Main result)**.** *No high cycle consists of integers, other than the trivial cycle.*

Proof. Rotating $\mathbf{v_h}$ and $\mathbf{v_h^R}$ to the left (by one position each) yields two more parity vectors, $\mathbf{u}01$ and $\mathbf{u}10$, so that $f(\mathbf{u}01)$ and $f(\mathbf{u}10)$ are additional members of the high cycle.

Because $\mathbf{u}01$ and $\mathbf{u}10$ differ only in their last two components, we expect $f(\mathbf{u}01)$ and $f(\mathbf{u}10)$ to have similar summands. We express both in terms of $\mathbf{u}$, via Equations 2.1 and 2.2.

$$f(\mathbf{u}01) = \frac{3g(\mathbf{u}) + 2^{k-1}}{2^k - 3^x}$$
$$f(\mathbf{u}10) = \frac{3g(\mathbf{u}) + 2^{k-2}}{2^k - 3^x}$$

Assume, by way of contradiction, that these two high-cycle members are both integers. Then so are

$$f(\mathbf{u}01) - f(\mathbf{u}01),$$
$$\frac{3g(\mathbf{u}) + 2^{k-1}}{2^k - 3^x} - \frac{3g(\mathbf{u}) + 2^{k-2}}{2^k - 3^x},$$
$$\frac{2^{k-1} - 2^{k-2}}{2^k - 3^x}, \text{ and}$$
$$\frac{2^{k-2}}{2^k - 3^x}.$$

However, 2^{k-2} is not divisible by any odd number (when $k > 2$), so we have a contradiction, implying the high cycle contains at least one non-integral member. $\square$

Example 5.5. *For the* $(8, 5)$*-high-cycle, Figure 1 gives* $f(\mathbf{u}01) - f(\mathbf{u}01) = \frac{485}{13} - \frac{421}{13} = \frac{2^6}{13}$.

6. Summary

There are three notable features of this proof versus the no-circuit proof.

- Instead of assuming, by way of contradiction, that $f(\mathbf{v_c})$ is an integer, we instead assume that both $f(\mathbf{v_h})$ and $f(\mathbf{v_h^R})$ are both integers.

- Unlike the no-circuit proof, we do not require deep lower bounds [1, 7] on the size of $2^k - 3^x$.

- We require no closed-form expression for any high cycle member; by contrast, the no-circuit proof relies on the expression $f(\mathbf{v_c}) = \frac{\mathbf{3^x - 2^x}}{\mathbf{2^k - 3^x}}$.

References

[1] Alan Baker, *Transcendental number theory*, Cambridge University Press, 1975.

[2] David Barina, *Convergence verification of the Collatz problem*, The Journal of Supercomputing **77** (2021), 2681–2688.

[3] Jean Berstel, Aaron Lauve, Christophe Reutenauer, and Franco V. Saliola, *Combinatorics on words: Christoffel words and repetitions in words*, American Mathematical Society, Providence, RI, 2009.

[4] Corrado Böhm and Giovanna Sontacchi, *On the existence of cycles of given length in integer sequences like $x_{n+1} = x_n/2$ when x is even and $x_{n+1} = 3x_n + 1$ otherwise*, Atti della Accademia Nazionale dei Lincei. Classe di Scienze Fisiche, Matematiche e Naturali **64** (1978), 260–264.

[5] Harvey Cohn, *Markoff forms and primitive words*, Mathematische Annalen **196** (1972), 8–22.

[6] Shalom Eliahou, *The $3x + 1$ problem: new lower bounds on nontrivial cycle lengths*, Discrete Mathematics **118** (1993), 45–56.

[7] W. J. Ellison, *On a theorem of S. Sivasankaranarayana Pillai*, Séminaire de théorie des nombres de Bordeaux (1971), 1–10.

[8] Lorenz Halbeisen and Norbert Hungerbühler, *Optimal bounds for the length of rational Collatz cycles*, Acta Arithmetica **LXXVIII** (1997), 227–239.

[9] Jeffrey C. Lagarias, *The 3x+1 problem and its generalizations*, The 3x+1 Problem: The Ultimate Challenge (Jeffrey C. Lagarias, ed.), American Mathematical Society, 2010, pp. 31–54.

[10] O. Rozier, *Démonstration de l'absence de cycles d'une certaine forme pour le problème de syracuse*, Singularitié **3** (1990), 9–10.

[11] R. P. Steiner, *A theorem on the Syracuse problem*, Proc. 7th Manitoba Conference on Numerical Mathematics, 1977, pp. 553–559.