Mixed Base Rewriting for the Collatz Conjecture

Emre Yolcu ✉
Carnegie Mellon University, Pittsburgh, PA 15213, USA

Scott Aaronson ✉
University of Texas at Austin, Austin, TX 78712, USA

Marijn J. H. Heule ✉
Carnegie Mellon University, Pittsburgh, PA 15213, USA

Abstract

We explore the Collatz conjecture and its variants through the lens of termination of string rewriting. We construct a rewriting system that simulates the iterated application of the Collatz function on strings corresponding to mixed binary–ternary representations of positive integers. We prove that the termination of this rewriting system is equivalent to the Collatz conjecture. We also prove that a previously studied rewriting system that simulates the Collatz function using unary representations does not admit termination proofs via matrix interpretations. To show the feasibility of our approach in proving mathematically interesting statements, we implement a minimal termination prover that uses matrix/arctic interpretations and we find automated proofs of nontrivial weakenings of the Collatz conjecture. Although we do not succeed in proving the Collatz conjecture, we believe that the ideas here represent an interesting new approach.

2012 ACM Subject Classification Theory of computation → Automated reasoning; Theory of computation → Rewrite systems

Keywords and phrases string rewriting, termination, matrix interpretations, SAT solving, Collatz conjecture, computer-assisted mathematics

Related Version This work is a short version of a paper appearing at CADE-28.
Extended preprint: `https://arxiv.org/abs/2105.14697`

Supplementary Material *Code*: `https://github.com/emreyolcu/rewriting-collatz`

1 Introduction

Let $\mathbb{N}$ and $\mathbb{N}^+$ denote the natural numbers and the positive integers, respectively. We define the *Collatz function* $T\colon \mathbb{N}^+ \to \mathbb{N}^+$ as $T(n) = n/2$ if $n \equiv 0 \pmod 2$ and $T(n) = (3n+1)/2$ if $n \equiv 1 \pmod 2$. With T^k denoting the kth iterate of T, the well-known *Collatz conjecture* [4] states that for all $n \in \mathbb{N}^+$, there exists some $k \in \mathbb{N}$ such that $T^k(n) = 1$. More generally, letting X be either $\mathbb{N}$ or $\mathbb{N}^+$, we consider a function $f\colon X \to X$. We call the sequence $x, f(x), f^2(x), \ldots$ the *f-trajectory of x*. If this trajectory contains 1, it is called *convergent*. If all f-trajectories are convergent, we say that f is *convergent*.

In this paper, we describe an approach based on termination of string rewriting to automatically search for a proof of the Collatz conjecture. Although trying to prove the Collatz conjecture via automated deduction is clearly a moonshot goal, there are two technological advances that provide reasons for optimism that at least some interesting variants of the problem might be solvable. First, the invention of the method of matrix interpretations [1] and its variants such as arctic interpretations [3] turns the quest of finding a ranking function to witness termination into a problem that is suitable for systematic search. Second, the progress in satisfiability (SAT) solving makes it possible to solve many seemingly difficult combinatorial problems efficiently in practice. Their combination, i.e., using SAT solvers to find interpretations, has so far been effective in solving challenging termination problems. We make the following contributions:

- We show how a Collatz-like function can be expressed as a rewriting system that is terminating if and only if the function is convergent.
- We prove that no termination proof via matrix interpretations exists for a certain system that simulates the Collatz function using unary representations of numbers.
- We show that translations into rewriting systems that use non-unary representations of numbers are more amenable to automated methods, compared with the previously and commonly studied unary representations.
- We automatically prove various weakenings of the Collatz conjecture. We observe that, for some of these weakenings, the only matrix/arctic interpretations that our termination tool was able to find involved relatively large matrices (of dimension 5). Existing termination tools often limit their default strategies to search for small interpretations as they are tailored for the setting where the task is to quickly solve a large quantity of relatively easy problems. We make the point that, given more resources, the method of matrix/arctic interpretations has the potential to scale.

2 Rewriting the Collatz Function

We start with systems that use unary representations and then demonstrate via examples that mixed base representations can be more suitable for use with automated methods.

Rewriting in Unary. The following system of Zantema [5] simulates the iterated application of the Collatz function to a number represented in unary, and it terminates upon reaching 1.

▶ **Example 1.** $\mathcal{Z}$ denotes the following SRS, consisting of 5 symbols and 7 rules.

$$\begin{array}{lll} \mathtt{h11} \to \mathtt{1h} & \mathtt{11h\diamond} \to \mathtt{11s\diamond} & \mathtt{h1\diamond} \to \mathtt{t11\diamond} \\ & \mathtt{1s} \to \mathtt{s1} & \mathtt{1t} \to \mathtt{t111} \\ & \mathtt{\diamond s} \to \mathtt{\diamond h} & \mathtt{\diamond t} \to \mathtt{\diamond h} \end{array}$$

▶ **Theorem 2** ([5, Theorem 16]). *$\mathcal{Z}$ is terminating if and only if the Collatz conjecture holds.*

While the forward direction of the above theorem is easy to see (since $\diamond\mathtt{h1}^{2n}\diamond \to^*_{\mathcal{Z}} \diamond\mathtt{h1}^{n}\diamond$ for $n > 1$ and $\diamond\mathtt{h1}^{2n+1}\diamond \to^*_{\mathcal{Z}} \diamond\mathtt{h1}^{3n+2}\diamond$ for $n \geq 0$), the backward direction is far from obvious because not every string corresponds to a valid configuration of the underlying machine.

As another example, consider the system $\mathcal{W} = \{\mathtt{h11} \to \mathtt{1h}, \mathtt{1h\diamond} \to \mathtt{1t\diamond}, \mathtt{1t} \to \mathtt{t111}, \diamond\mathtt{t} \to \diamond\mathtt{h}\}$ (originally due to Zantema, available at: `https://www.lri.fr/~marche/tpdb/tpdb-2.0/SRS/Zantema/z079.srs`). Termination of this system has yet to be proved via automated methods. Nevertheless, there is a simple reason for its termination: It simulates the iterated application of a Collatz-like function $W : \mathbb{N}^+ \to \mathbb{N}^+$ defined as $W(n) = 3n/2$ if $n \equiv 0 \pmod 2$ and $W(n) = 1$ if $n \equiv 1 \pmod 2$, which is easily seen to be convergent.

Matrix interpretations cannot be used to remove any of the rules from the above kind of unary rewriting systems that simulate certain maps, in particular the Collatz function. We prove the below theorem in the full version of this work. We adopt the notation of [1].

▶ **Theorem 3.** *Let $\Sigma = \{\mathtt{1}, \diamond, \mathtt{h}, \mathtt{s}, \mathtt{t}\}$. There exists no collection $[\cdot]_\Sigma$ of matrix interpretations of any dimension d such that*

- *for at least a rule $\ell \to r \in \mathcal{Z}$ we have $[\ell](\mathbf{x}) > [r](\mathbf{x})$ for all $\mathbf{x} \in \mathbb{N}^d$, and*
- *for the remaining $\ell' \to r' \in \mathcal{Z}$ we have $[\ell'](\mathbf{x}) \gtrsim [r'](\mathbf{x})$ for all $\mathbf{x} \in \mathbb{N}^d$.*

By an argument analogous to above, we can also prove that no such collection of interpretations exists for $\mathcal{W}$. If a proof of the Collatz conjecture is to be produced by some automated method that relies on rewriting, then that method better be able to prove a statement as simple as the convergence of W. With this in mind, we describe an alternative rewriting system that simulates the Collatz function and terminates upon reaching 1. We then provide examples where the alternative system is more suitable for use with termination tools (for instance allowing a matrix interpretations proof of the convergence of W).

Rewriting in Mixed Base. In the mixed base scheme, the overall idea is as follows. Given a number $n \in \mathbb{N}^+$, we write a mixed binary–ternary representation for it (noting that this representation is not unique). With this representation, as long as the least significant digit is binary, the parity of the number can be recognized by checking only this digit, as opposed to scanning the entire string when working in unary. This allows us to easily determine the correct case when applying the Collatz function. If the least significant digit is ternary, then the representation is rewritten (while preserving its value) to make this digit binary. Afterwards, since computing $2n \mapsto n$ corresponds to erasing a trailing binary 0 and computing $2n+1 \mapsto 3n+2$ corresponds to replacing a trailing binary 1 with a ternary 2, applying the Collatz function takes a single rewrite step.

We will describe an SRS $\mathcal{T}$ over the symbols $\{\mathtt{f}, \mathtt{t}, \mathtt{0}, \mathtt{1}, \mathtt{2}, \triangleleft, \triangleright\}$ that simulates the iterated application of the Collatz function and terminates upon reaching 1. The symbols $\mathtt{f}, \mathtt{t}$ correspond to binary digits $0_2, 1_2$; and $\mathtt{0}, \mathtt{1}, \mathtt{2}$ to ternary digits $0_3, 1_3, 2_3$. The symbol $\triangleleft$ marks the beginning of a string while also standing for the most significant digit (without loss of generality assumed to be 1_0) and $\triangleright$ marks the end of a string while also standing for the redundant trailing digit 0_1. Consider the functional view of these symbols:

$$\begin{aligned} \mathtt{f}(x) &= 2x \\ \mathtt{t}(x) &= 2x+1 \end{aligned} \qquad \begin{aligned} \mathtt{0}(x) &= 3x \\ \mathtt{1}(x) &= 3x+1 \\ \mathtt{2}(x) &= 3x+2 \end{aligned} \qquad \begin{aligned} \triangleleft(x) &= 1 \\ \triangleright(x) &= x \end{aligned} \tag{1}$$

A mixed base representation $N = (n_1)_{b_1}(n_2)_{b_2}\dots(n_k)_{b_k}$ represents the number $\mathrm{Val}(N) := \sum_{i=1}^{k} n_i \prod_{j=i+1}^{k} b_j$. We can see by rearranging this expression that $\mathrm{Val}(N)$ is also given by some composition of the above functions if we view the expression $\triangleleft(x)$ as the constant 1.

▶ **Example 4.** We can write $19 = \mathrm{Val}(\triangleleft\mathtt{0f1}\triangleright) = \triangleright(\mathtt{1}(\mathtt{f}(\mathtt{0}(\triangleleft(x)))))$. The string representation ends with a ternary symbol, so we will rewrite it. With the function view, we have $\mathtt{1}(\mathtt{f}(x)) = 3(2x)+1 = 6x+1 = 2(3x)+1 = \mathtt{t}(\mathtt{0}(x))$. This shows that we could also write $19 = \mathrm{Val}(\triangleleft\mathtt{00t}\triangleright)$, which now ends with the binary digit 1_2. This gives us the rewrite rule $\mathtt{f1} \to \mathtt{0t}$. We can now apply the Collatz function to this representation by rewriting only the rightmost two symbols of the string since $T(\triangleright(\mathtt{t}(x))) = \frac{3(2x+1)+1}{2} = \frac{6x+4}{2} = 3x+2 = (\triangleright(\mathtt{2}(x)))$. This gives us the rewrite rule $\mathtt{t}\triangleright \to \mathtt{2}\triangleright$. After applying this rule to the string $\triangleleft\mathtt{00t}\triangleright$, we indeed obtain $T(19) = 29 = \mathrm{Val}(\triangleleft\mathtt{002}\triangleright)$.

In the manner of the above example, we compute all the necessary transformations and obtain the following 11-rule SRS $\mathcal{T}$.

$$\mathcal{D}_T = \left\{ \begin{aligned} \mathtt{f}\triangleright &\to \triangleright \\ \mathtt{t}\triangleright &\to \mathtt{2}\triangleright \end{aligned} \right\} \qquad \mathcal{A} = \left\{ \begin{aligned} \mathtt{f0} &\to \mathtt{0f} & \mathtt{t0} &\to \mathtt{1t} \\ \mathtt{f1} &\to \mathtt{0t} & \mathtt{t1} &\to \mathtt{2f} \\ \mathtt{f2} &\to \mathtt{1f} & \mathtt{t2} &\to \mathtt{2t} \end{aligned} \right\} \qquad \mathcal{B} = \left\{ \begin{aligned} \triangleleft\mathtt{0} &\to \triangleleft\mathtt{t} \\ \triangleleft\mathtt{1} &\to \triangleleft\mathtt{ff} \\ \triangleleft\mathtt{2} &\to \triangleleft\mathtt{ft} \end{aligned} \right\}$$

This SRS is split into subsystems $\mathcal{D}_T$ (dynamic rules for T) and $\mathcal{X} = \mathcal{A} \cup \mathcal{B}$ (auxiliary rules). The two rules in $\mathcal{D}_T$ encode the application of the Collatz function T, while the rules in $\mathcal{X}$

serve to push binary symbols towards the rightmost end of the string by swapping the bases of adjacent positions without changing the represented value.

▶ **Example 5** (Rewrite sequence of $\mathcal{T}$)**.** Consider the string $s = \triangleleft\mathtt{ff0}\triangleright$ that represents the number 12. Below is a possible rewrite sequence of $\mathcal{T}$ that starts from s, with the corresponding values (under the interpretations from (1)) displayed above the strings. Underlines indicate the parts of the strings where the rules are applied.

$$\begin{array}{ccccccccccccccc} 12 & & 12 & & 6 & & 6 & & 3 & & 3 & & 5 & & 5 \\ \triangleleft\mathtt{f}\underline{\mathtt{f0}}\triangleright & \to_{\mathcal{A}} & \triangleleft\mathtt{f0}\underline{\mathtt{f}\triangleright} & \to_{\mathcal{D}_T} & \triangleleft\underline{\mathtt{f0}}\triangleright & \to_{\mathcal{A}} & \triangleleft\mathtt{0}\underline{\mathtt{f}\triangleright} & \to_{\mathcal{D}_T} & \underline{\triangleleft\mathtt{0}}\triangleright & \to_{\mathcal{B}} & \triangleleft\underline{\mathtt{t}\triangleright} & \to_{\mathcal{D}_T} & \underline{\triangleleft\mathtt{2}}\triangleright & \to_{\mathcal{B}} & \triangleleft\mathtt{f}\underline{\mathtt{t}\triangleright} \end{array}$$

$$\begin{array}{ccccccccccccc} & 8 & & 8 & & 8 & & 4 & & 2 & & 1 \\ \to_{\mathcal{D}_T} & \triangleleft\underline{\mathtt{f2}}\triangleright & \to_{\mathcal{A}} & \underline{\triangleleft\mathtt{1}}\mathtt{f}\triangleright & \to_{\mathcal{B}} & \triangleleft\mathtt{ff}\underline{\mathtt{f}\triangleright} & \to_{\mathcal{D}_T} & \triangleleft\mathtt{f}\underline{\mathtt{f}\triangleright} & \to_{\mathcal{D}_T} & \triangleleft\underline{\mathtt{f}\triangleright} & \to_{\mathcal{D}_T} & \triangleleft\triangleright \end{array}$$

The trajectory of T would continue upon reaching 1; however, in order to be able to formulate the Collatz conjecture as a termination problem, $\mathcal{T}$ is made in such a way that its rewrite sequences stop upon reaching the string representation $\triangleleft\triangleright$ of 1 since no rule is applicable.

Termination of the subsystems of $\mathcal{T}$ with $\mathcal{B}$ or $\mathcal{D}_T$ removed is easily seen. There is also a direct proof via linear polynomial interpretations after reversing the rules.

▶ **Lemma 6.** $\mathrm{SN}(\mathcal{T} \setminus \mathcal{B})$ *and* $\mathrm{SN}(\mathcal{T} \setminus \mathcal{D}_T)$.

When considering the termination of $\mathcal{T}$, it suffices to limit the discussion to initial strings of a specific form that we have been working with so far, e.g., in Examples 4 and 5.

▶ **Lemma 7.** *If $\mathcal{T}$ is terminating on all initial strings of the* canonical form $\triangleleft(\mathtt{f}|\mathtt{t}|\mathtt{0}|\mathtt{1}|\mathtt{2})^*\triangleright$, *then $\mathcal{T}$ is terminating (on all initial strings).*

As a whole, the rewriting system $\mathcal{T}$ simulates the iterated application of T (except at 1). Making use of Lemmas 6 and 7, we prove the following in the full version of this work.

▶ **Theorem 8.** *$\mathcal{T}$ is terminating if and only if T is convergent.*

3 Automated Proofs

We adapt the rewriting system $\mathcal{T}$ for different Collatz-like functions to explore the effectiveness of the mixed base scheme on weakened variants of the Collatz conjecture.

Convergence of W. Earlier we mentioned a Collatz-like function W as a simple example that could serve as a sanity check for an automated method aiming to solve Collatz-like problems. With the mixed binary–ternary scheme, this function can be seen to be simulated by the system $\mathcal{W}' = \{\mathtt{f}\triangleright \to \mathtt{0}\triangleright\} \cup \mathcal{X}$. A small matrix interpretations proof is found for this system in less than a second, in contrast to its variant $\mathcal{W}$ that uses unary representations for which no automated proof is known.

Farkas' Variant. Farkas [2] studied a slight modification of the Collatz function for which it becomes possible to prove convergence via induction. We consider automatically proving the convergence of this function as another test case for the mixed base scheme that is easier than the Collatz conjecture without being entirely trivial. Below, we define a function $F \colon \mathbb{N} \to \mathbb{N}$ that is equivalent to Farkas' definition in terms of convergence while resembling the Collatz function even more closely (with respect to the definitions of the cases). This variant is obtained by introducing an additional case in the Collatz function for $n \equiv 1 \pmod 3$ and applying T otherwise. Its definition and a set $\mathcal{D}_F$ of dynamic rules are shown below.

$$F(n) = \begin{cases} \frac{n-1}{3} & \text{if } n \equiv 1 \pmod 3 \\ \frac{n}{2} & \text{if } n \equiv 0 \text{ or } n \equiv 2 \pmod 6 \\ \frac{3n+1}{2} & \text{if } n \equiv 3 \text{ or } n \equiv 5 \pmod 6 \end{cases} \qquad \mathcal{D}_F = \left\{ \begin{array}{rcl} \mathtt{1}\triangleright & \to & \triangleright \\ \mathtt{0f}\triangleright & \to & \mathtt{0}\triangleright \\ \mathtt{1f}\triangleright & \to & \mathtt{1}\triangleright \\ \mathtt{1t}\triangleright & \to & \mathtt{12}\triangleright \\ \mathtt{2t}\triangleright & \to & \mathtt{22}\triangleright \end{array} \right\}$$

Termination of the rewriting system $\mathcal{F} = \mathcal{D}_F \cup \mathcal{X}$ is equivalent to the convergence of F. The proof of the equivalence is similar to that of Theorem 8, with the difference that when constructing a nonterminating rewrite sequence from a nonconvergent trajectory we write the first number in the trajectory in ternary (except for the most significant digit) and always perform the rightmost possible rewrite.

Farkas gave an inductive proof of convergence for (a variant of) F via case analysis. We found an automated proof that $\mathcal{F}$ is terminating via arctic interpretations (where the proof appears to require matrices of dimension 5 for certain steps). It is worth mentioning that the default configurations of the existing termination tools (e.g., AProVE, Matchbox) are too conservative to prove the termination of this system, but after their authors tweaked the strategies they were also able to find automated proofs via arctic interpretations.

Subsets of $\mathcal{T}$. It is also interesting to consider whether we can automatically prove the terminations of proper subsets of $\mathcal{T}$. Specifically, we considered the 11 subsystems obtained by leaving out a single rewriting rule from $\mathcal{T}$, and we found termination proofs via matrix/arctic interpretations for all of the 11 subproblems. Our interest in these problems is threefold:

1. Termination of $\mathcal{T}$ implies the terminations of all of its subsystems, so proving its termination is at least as difficult a task as proving the terminations of the 11 subsystems. Therefore, the subproblems serve as additional sanity checks that an automated approach aspiring to succeed for the Collatz conjecture ought to be able to pass.
2. Having proved the terminations of all 11 subsystems is a partial solution to the full problem, since it implies that for any single rule $\ell \to r \in \mathcal{T}$, proving that $\ell \to r$ is terminating relative to $\mathcal{T}$ settles the Collatz conjecture.
3. After the removal of a rule, the termination of the remaining system still encodes a valid mathematical question about the Collatz trajectories, i.e., the system does not become terminating for a trivial reason.

Table 1 shows the parameters of the matrix/arctic interpretations proofs that we found for the termination of each subsystem. For each rule $\ell \to r$ that is left out, we searched for a stepwise proof to show that $\mathcal{T} \setminus \{\ell \to r\}$ is terminating. On the table, we report the smallest parameters (in terms of matrix dimension) that work for all of the proof steps. In the experiments we searched (with a timeout of 30 seconds) for matrices of up to 7 dimensions, with the coefficients taking at most 8 different values.

Collatz Trajectories Modulo 8. Let m be a power of 2. Given $k \in \{0, 1, \ldots, m-1\}$, is it the case that all nonconvergent Collatz trajectories contain some $n \equiv k \pmod m$? For several values of k this can be proved to hold by inspecting the transitions of the iterates in the Collatz trajectories across residue classes modulo m. These questions can also be formulated as the terminations of some rewriting systems. With this approach we found automated proofs for several cases, which are also not difficult to prove by hand.

Table 1 Smallest proofs found for the terminations of subsystems of $\mathcal{T}$. The columns show the matrix dimension D and the maximum number V of distinct coefficients that appear in the matrices, along with the median time to find an entire proof across 25 repetitions for the fixed D and V.

	Matrix			Arctic		
Rule removed	D	V	Time	D	V	Time
$\mathtt{f}\triangleright \to \triangleright$	3	4	1.42s	3	5	15.95s
$\mathtt{t}\triangleright \to \mathtt{2}\triangleright$	1	2	0.27s	1	3	0.28s
$\mathtt{f0} \to \mathtt{0f}$	4	2	0.92s	3	4	2.46s
$\mathtt{f1} \to \mathtt{0t}$	1	3	0.50s	1	4	0.51s
$\mathtt{f2} \to \mathtt{1f}$	1	2	0.38s	1	3	0.39s
$\mathtt{t0} \to \mathtt{1t}$	4	3	1.20s	3	4	0.87s
$\mathtt{t1} \to \mathtt{2f}$	5	2	0.89s	4	3	0.84s
$\mathtt{t2} \to \mathtt{2t}$	4	4	10.00s	2	5	0.62s
$\triangleleft\mathtt{0} \to \triangleleft\mathtt{t}$	2	2	0.40s	2	3	0.42s
$\triangleleft\mathtt{1} \to \triangleleft\mathtt{ff}$	3	3	0.53s	3	4	0.57s
$\triangleleft\mathtt{2} \to \triangleleft\mathtt{ft}$	4	4	7.51s	4	3	4.04s

▶ **Theorem 9.** *If there exists a nonconvergent Collatz trajectory, it cannot avoid the residue classes of* 2, 3, 4, 6 *modulo* 8.

It remains open whether the above holds for the residue classes of 0, 1, 5, 7 modulo 8.

4 Future Work

Several extensions to this work can further our understanding of the potential of rewriting techniques for answering mathematical questions. For instance, it is of interest to study the efficacy of different termination proving techniques on the problems that we considered. We found matrix/arctic interpretations to be the most successful for our purposes despite experimenting with existing tools that implement newer techniques developed for automatically proving the terminations of a few select challenging instances. It might also be possible to prove that there exist no matrix/arctic interpretations to establish the termination of the Collatz system $\mathcal{T}$. This would be an interesting result in itself. Another issue is the matter of representation; specifically, it is worth exploring whether there exists a suitable translation of the Collatz conjecture into the termination of a term, instead of string, rewriting system since many automated termination proving techniques are generalized to term rewriting. Finally, injecting problem-specific knowledge into the rewriting systems or the termination techniques would be helpful as there exists a wealth of information about the Collatz conjecture that could simplify the search for a termination proof.

References

1 Jörg Endrullis, Johannes Waldmann, and Hans Zantema. Matrix interpretations for proving termination of term rewriting. *Journal of Automated Reasoning*, 40(2):195–220, 2008.

2 Hershel M. Farkas. Variants of the $3N + 1$ conjecture and multiplicative semigroups. In *Geometry, Spectral Theory, Groups, and Dynamics*, pages 121–127. 2005.

3 Adam Koprowski and Johannes Waldmann. Max/plus tree automata for termination of term rewriting. *Acta Cybernetica*, 19(2):357–392, 2009.

4 Jeffrey C. Lagarias. *The Ultimate Challenge: The* $3x + 1$ *Problem*. American Mathematical Society, 2010.

5 Hans Zantema. Termination of string rewriting proved automatically. *Journal of Automated Reasoning*, 34(2):105–139, 2005.