

# An Automated Approach to the Collatz Conjecture

*(Explorations with Rewriting Systems, Matrix Interpretations, and SAT)*

Emre Yolcu

Computer Science Department  
Carnegie Mellon University

with Scott Aaronson and Marijn Heule

# Automated deduction for mathematics

Examples of conjectures settled with computer assistance:

- ▶ Four color theorem (1976)
- ▶ Robbins conjecture (1996)
- ▶ Kepler conjecture (1998)
- ▶ Boolean Pythagorean triples problem (2016)

**This work:** Approach the Collatz conjecture with the techniques developed for automatically proving termination of string rewriting.

**Available at:** <https://arxiv.org/abs/2105.14697>

# String rewriting

$\Sigma$     alphabet

$l \rightarrow r$     rewrite rule

$R$     rewriting system

$\rightarrow_R$     rewrite relation

# String rewriting

$\Sigma$  alphabet

$l \rightarrow r$  rewrite rule

$R$  rewriting system

$\rightarrow_R$  rewrite relation

**Example:**  $R = \{ab \rightarrow ba\}$

abbab  $\rightarrow_R$  babab  $\rightarrow_R$  bbaab  $\rightarrow_R$  bbaba  $\rightarrow_R$  bbbaa

# Termination

$R$  is terminating if there is no infinite chain

$$X_0 \rightarrow_R X_1 \rightarrow_R X_2 \rightarrow_R \dots$$

where  $X_i \in \Sigma^*$ .

# Termination

## Examples:

▶  $R = \{ab \rightarrow ba\}$  terminating

▶  $S = \{a \rightarrow aa\}$  nonterminating

▶  $T = \{aa \rightarrow a\}$  terminating

▶  $P = \left\{ \begin{array}{l} a \rightarrow b \\ b \rightarrow a \end{array} \right\}$  nonterminating

▶  $Z = \left\{ \begin{array}{l} aa \rightarrow bc \\ bb \rightarrow ac \\ cc \rightarrow ab \end{array} \right\}$  ?

▶  $G = \left\{ \begin{array}{l} aaaa \rightarrow babb \\ baab \rightarrow aaba \end{array} \right\}$  ?

## Proving termination [Baader and Nipkow, 1998]

$R$  is terminating if there exists a well-founded order  $\succ_{\Sigma^*}$  such that for all  $X, Y \in \Sigma^*$

$$X \rightarrow_R Y \quad \implies \quad X \succ_{\Sigma^*} Y.$$

## Proving termination [Baader and Nipkow, 1998]

$R$  is terminating if there exists a well-founded order  $\succ_{\Sigma^*}$  such that for all  $X, Y \in \Sigma^*$

$$X \rightarrow_R Y \quad \Longrightarrow \quad X \succ_{\Sigma^*} Y.$$

Translate  $s \in \Sigma$  into an element  $[s]$  of another domain  $(A, \succ_A)$  and show that  $[\ell] \succ_A [r]$  for all  $\ell \rightarrow r \in R$ .

## Proving termination [Baader and Nipkow, 1998]

$R$  is terminating if there exists a well-founded order  $\succ_{\Sigma^*}$  such that for all  $X, Y \in \Sigma^*$

$$X \rightarrow_R Y \quad \Longrightarrow \quad X \succ_{\Sigma^*} Y.$$

Translate  $s \in \Sigma$  into an element  $[s]$  of another domain  $(A, \succ_A)$  and show that  $[\ell] \succ_A [r]$  for all  $\ell \rightarrow r \in R$ .

### Examples:

- ▶  $T = \{aa \rightarrow a\}$ . Let  $[a] = 1$ , extend to strings additively.

$$[aa] = 2 > 1 = [a]$$

## Proving termination [Baader and Nipkow, 1998]

$R$  is terminating if there exists a well-founded order  $\succ_{\Sigma^*}$  such that for all  $X, Y \in \Sigma^*$

$$X \rightarrow_R Y \quad \Longrightarrow \quad X \succ_{\Sigma^*} Y.$$

Translate  $s \in \Sigma$  into an element  $[s]$  of another domain  $(A, \succ_A)$  and show that  $[\ell] \succ_A [r]$  for all  $\ell \rightarrow r \in R$ .

### Examples:

- ▶  $T = \{aa \rightarrow a\}$ . Let  $[a] = 1$ , extend to strings additively.

$$[aa] = 2 > 1 = [a]$$

- ▶  $R = \{ab \rightarrow ba\}$ . Let  $[a](x) = x^2$  and  $[b](x) = x + 1$ , extend to strings compositionally.

$$[ab](x) = (x + 1)^2 > x^2 + 1 = [ba](x)$$

## Matrix interpretations [Hofbauer and Waldmann, 2006]

Affine functions  $[s]: \mathbb{N}^d \rightarrow \mathbb{N}^d$  (extended to strings compositionally):

$$[s](x) = M_s x + v_s$$

Well-founded order:

$$x \succ y \iff x_1 > y_1 \wedge x_i \geq y_i \text{ for } i \in \{2, 3, \dots, d\}$$

Look for interpretations satisfying

$$[\ell](x) = M_\ell x + v_\ell \succ M_r x + v_r = [r](x) \text{ for all } x \in \mathbb{N}^d.$$

**Encode all the resulting constraints as a SAT instance.**

## Example proof

$$Z = \left\{ \begin{array}{l} aa \rightarrow bc \\ bb \rightarrow ac \\ cc \rightarrow ab \end{array} \right\}$$

## Example proof

$$Z = \left\{ \begin{array}{l} aa \rightarrow bc \\ bb \rightarrow ac \\ cc \rightarrow ab \end{array} \right\}$$

$$[a](x) = \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 2 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

$$[b](x) = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \end{bmatrix} x + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 2 \end{bmatrix}$$

$$[c](x) = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \\ 3 \\ 0 \end{bmatrix}$$

## Example proof (cont.)

$$x \succ y \iff x_1 > y_1 \wedge x_i \geq y_i \text{ for } i \in \{2, 3, \dots, d\}.$$

It is decidable to check the following.

$$[aa](x) = \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 2 & 0 & 2 \end{bmatrix} x + \begin{bmatrix} 2 \\ 0 \\ 1 \\ 2 \end{bmatrix} \succ \begin{bmatrix} 1 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 2 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 2 \end{bmatrix} = [bc](x)$$

$$[bb](x) = \begin{bmatrix} 1 & 2 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 2 & 0 & 4 \end{bmatrix} x + \begin{bmatrix} 4 \\ 0 \\ 2 \\ 6 \end{bmatrix} \succ \begin{bmatrix} 1 & 1 & 0 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 2 & 0 & 4 \end{bmatrix} x + \begin{bmatrix} 2 \\ 0 \\ 1 \\ 6 \end{bmatrix} = [ac](x)$$

$$[cc](x) = \begin{bmatrix} 1 & 1 & 0 & 2 \\ 0 & 2 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 2 \end{bmatrix} x + \begin{bmatrix} 2 \\ 0 \\ 3 \\ 0 \end{bmatrix} \succ \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 2 \end{bmatrix} x + \begin{bmatrix} 1 \\ 0 \\ 3 \\ 0 \end{bmatrix} = [ab](x)$$

# Collatz conjecture

$$C(n) = \begin{cases} \frac{n}{2} & \text{if } n \equiv 0 \pmod{2} \\ \frac{3n+1}{2} & \text{if } n \equiv 1 \pmod{2} \end{cases}$$

## Conjecture

*For all  $n \in \mathbb{N}^+$ , the trajectory  $C(n), C(C(n)), C(C(C(n))), \dots$  reaches 1.*

## String rewriting for Collatz [Zantema, 2005]

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

### Theorem

*Above system is terminating  $\iff$  Collatz conjecture holds.*

$$\begin{array}{lll} \diamond h1^{2n}\diamond & \rightarrow^* & \diamond h1^n\diamond & \text{for } n > 1 \\ \diamond h1^{2n+1}\diamond & \rightarrow^* & \diamond h1^{3n+2}\diamond & \text{for } n \geq 0 \end{array}$$

## String rewriting for Collatz [Zantema, 2005]

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

### Theorem

*There exists no collection of matrix interpretations of any dimension  $d$  for the above system such that*

- ▶ *for some rule  $\ell \rightarrow r$  we have  $[\ell](x) \succ [r](x)$  for all  $x \in \mathbb{N}^d$ , and*
- ▶ *for remaining  $\ell' \rightarrow r'$  we have  $[\ell'](x) \succeq [r'](x)$  for all  $x \in \mathbb{N}^d$ .*

## Trivial test

$h11 \rightarrow 1h$

$1h\diamond \rightarrow 1t\diamond$

$1t \rightarrow t111$

$\diamond t \rightarrow \diamond h$

Applies  $n \mapsto 3n/2$  as long as 2 divides  $n$ .

No automated termination proof is known.

## Mixed binary–ternary representation

Can view binary representations as compositions of the functions

$$f(x) = 2x$$

$$t(x) = 2x + 1$$

along with  $\triangleleft(x) = 1$ .

**Example:**

$$12 = (1100)_2 = f(f(t(\triangleleft(x))))$$

## Mixed binary–ternary representation

Can view binary representations as compositions of the functions

$$f(x) = 2x$$

$$t(x) = 2x + 1$$

along with  $\triangleleft(x) = 1$ .

**Example:**

$$12 = (1100)_2 = f(f(t(\triangleleft(x))))$$

Division by 2 corresponds to:  $12/2 = 6 = f(t(\triangleleft(x)))$

And then:  $6/2 = 3 = t(\triangleleft(x))$

## Mixed binary–ternary representation

Can view binary representations as compositions of the functions

$$f(x) = 2x$$

$$t(x) = 2x + 1$$

along with  $\triangleleft(x) = 1$ .

**Example:**

$$12 = (1100)_2 = f(f(t(\triangleleft(x))))$$

Division by 2 corresponds to:  $12/2 = 6 = f(t(\triangleleft(x)))$

And then:  $6/2 = 3 = t(\triangleleft(x))$

If we had a function  $h(x) = 3x + 1$ , we could write

$$10 = h(3) = h(t(\triangleleft(x))).$$

## Mixed binary–ternary representation (cont.)

$$\begin{array}{lll} \mathbf{f}(x) = 2x & 0(x) = 3x & \triangleleft(x) = 1 \\ \mathbf{t}(x) = 2x + 1 & 1(x) = 3x + 1 & \triangleright(x) = x \\ & 2(x) = 3x + 2 & \end{array}$$

## Mixed binary–ternary representation (cont.)

$$\begin{array}{lll} \mathbf{f}(x) = 2x & 0(x) = 3x & \triangleleft(x) = 1 \\ \mathbf{t}(x) = 2x + 1 & 1(x) = 3x + 1 & \triangleright(x) = x \\ & 2(x) = 3x + 2 & \end{array}$$

Consider  $19 = \triangleleft 0 \mathbf{f} 1 \triangleright = \triangleright (1(\mathbf{f}(0(\triangleleft(x))))))$ . Ends with ternary, so we will rewrite.

## Mixed binary–ternary representation (cont.)

$$\begin{array}{lll} f(x) = 2x & 0(x) = 3x & \triangleleft(x) = 1 \\ t(x) = 2x + 1 & 1(x) = 3x + 1 & \triangleright(x) = x \\ & 2(x) = 3x + 2 & \end{array}$$

Consider  $19 = \triangleleft 0f1 \triangleright = \triangleright(1(f(0(\triangleleft(x))))))$ . Ends with ternary, so we will rewrite.

We have  $1(f(x)) = 3(2x) + 1 = 6x + 1 = 2(3x) + 1 = t(0(x))$ .

Could also write  $19 = \triangleleft 00t \triangleright$ .

## Mixed binary–ternary representation (cont.)

$$\begin{array}{lll} f(x) = 2x & 0(x) = 3x & \triangleleft(x) = 1 \\ t(x) = 2x + 1 & 1(x) = 3x + 1 & \triangleright(x) = x \\ & 2(x) = 3x + 2 & \end{array}$$

Consider  $19 = \triangleleft 0f1 \triangleright = \triangleright(1(f(0(\triangleleft(x))))))$ . Ends with ternary, so we will rewrite.

We have  $1(f(x)) = 3(2x) + 1 = 6x + 1 = 2(3x) + 1 = t(0(x))$ .  
Could also write  $19 = \triangleleft 00t \triangleright$ .

Apply Collatz map at the rightmost end via

$$C(\triangleright(t(x))) = \frac{3(2x + 1) + 1}{2} = \frac{6x + 4}{2} = 3x + 2 = \triangleright(2(x)),$$

obtaining  $C(19) = 29 = \triangleleft 002 \triangleright$ .

## Alternative rewriting system for Collatz

$f\triangleright \rightarrow \triangleright$	$f0 \rightarrow 0f$	$t0 \rightarrow 1t$	$\triangleleft 0 \rightarrow \triangleleft t$
$t\triangleright \rightarrow 2\triangleright$	$f1 \rightarrow 0t$	$t1 \rightarrow 2f$	$\triangleleft 1 \rightarrow \triangleleft ff$
	$f2 \rightarrow 1f$	$t2 \rightarrow 2t$	$\triangleleft 2 \rightarrow \triangleleft ft$

### Theorem

*Above system is terminating  $\iff$  Collatz conjecture holds.*

# Alternative rewriting system for Collatz

$f\triangleright \rightarrow \triangleright$	$f0 \rightarrow 0f$	$t0 \rightarrow 1t$	$\triangleleft 0 \rightarrow \triangleleft t$
$t\triangleright \rightarrow 2\triangleright$	$f1 \rightarrow 0t$	$t1 \rightarrow 2f$	$\triangleleft 1 \rightarrow \triangleleft ff$
	$f2 \rightarrow 1f$	$t2 \rightarrow 2t$	$\triangleleft 2 \rightarrow \triangleleft ft$

## Theorem

*Above system is terminating  $\iff$  Collatz conjecture holds.*

## Example sequence:

12	12	6	6	3	3	5
$\triangleleft \underline{ff0} \triangleright$	$\rightarrow \triangleleft \underline{f0f} \triangleright$	$\rightarrow \triangleleft \underline{f0} \triangleright$	$\rightarrow \triangleleft \underline{0f} \triangleright$	$\rightarrow \triangleleft \underline{0} \triangleright$	$\rightarrow \triangleleft \underline{t} \triangleright$	$\rightarrow \triangleleft \underline{2} \triangleright$
5	8	8	8	4	2	1
$\rightarrow \triangleleft \underline{ft} \triangleright$	$\rightarrow \triangleleft \underline{f2} \triangleright$	$\rightarrow \triangleleft \underline{1f} \triangleright$	$\rightarrow \triangleleft \underline{fff} \triangleright$	$\rightarrow \triangleleft \underline{ff} \triangleright$	$\rightarrow \triangleleft \underline{f} \triangleright$	$\rightarrow \triangleleft \triangleright$

## Nontrivial test: Farkas' map [Farkas, 2005]

$$F(n) = \begin{cases} \frac{n-1}{3} & \text{if } n \equiv 1 \pmod{3} \\ \frac{n}{2} & \text{if } n \equiv 0 \text{ or } n \equiv 2 \pmod{6} \\ \frac{3n+1}{2} & \text{if } n \equiv 3 \text{ or } n \equiv 5 \pmod{6} \end{cases}$$

1▷ → ▷	f0 → 0f	t0 → 1t	◁0 → ◁t
0f▷ → 0▷	f1 → 0t	t1 → 2f	◁1 → ◁ff
1f▷ → 1▷	f2 → 1f	t2 → 2t	◁2 → ◁ft
1t▷ → 12▷			
2t▷ → 22▷			

## Farkas' map (cont.)

$$[f](x) = \begin{bmatrix} - & - & - & 2 & - \\ & 2 & 0 & - & - \\ 2 & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix} x \oplus \begin{bmatrix} 0 \\ - \\ - \\ - \\ - \end{bmatrix} \quad [t](x) = \begin{bmatrix} - & - & - & - & 2 \\ 0 & 2 & 0 & - & 0 \\ 2 & - & 2 & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix} x \oplus \begin{bmatrix} 0 \\ - \\ - \\ - \\ - \end{bmatrix}$$

$$[\triangleleft](x) = \begin{bmatrix} 0 \\ 2 \\ - \\ - \\ 4 \end{bmatrix} \quad [\triangleright](x) = \begin{bmatrix} 0 & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \\ - & - & - & - & - \end{bmatrix} x$$

$$[0](x) = \begin{bmatrix} 0 & 4 & 0 & - & - \\ - & 4 & - & - & - \\ - & 4 & 0 & - & - \\ 0 & 3 & 0 & - & - \\ - & - & - & - & - \end{bmatrix} x \quad [1](x) = \begin{bmatrix} 1 & - & - & - & - \\ - & 4 & 0 & - & - \\ - & 4 & 0 & - & - \\ 0 & - & - & - & - \\ 0 & 3 & 0 & - & - \end{bmatrix} x \quad [2](x) = \begin{bmatrix} 0 & - & 0 & - & - \\ - & 4 & - & - & - \\ 0 & - & 1 & - & 0 \\ - & - & - & - & - \\ 0 & - & 0 & - & 0 \end{bmatrix} x$$

## Partial solution

Leave out a rule from

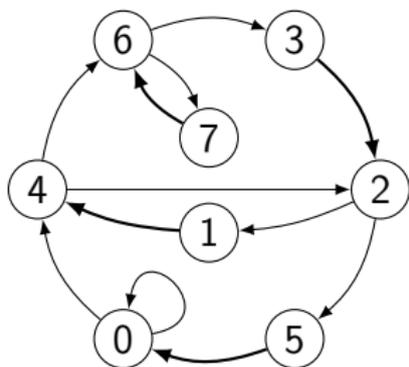
$$\begin{array}{llll} f \triangleright \rightarrow \triangleright & f0 \rightarrow 0f & t0 \rightarrow 1t & \triangleleft 0 \rightarrow \triangleleft t \\ t \triangleright \rightarrow 2 \triangleright & f1 \rightarrow 0t & t1 \rightarrow 2f & \triangleleft 1 \rightarrow \triangleleft ff \\ & f2 \rightarrow 1f & t2 \rightarrow 2t & \triangleleft 2 \rightarrow \triangleleft ft \end{array}$$

and try to prove termination.

All 11 subproblems are solvable.

$\implies$  An interpretation that decreases strictly wrt any single rule (nonstrictly wrt others) settles the Collatz conjecture.

## Collatz trajectories modulo 8



Nonconvergent trajectories cannot avoid  $\{2, 3, 4, 6\}$ .

It remains unknown whether any of  $\{1, 5, 7\}$  has to be encountered in all trajectories.

# SAT solving considerations

## Phase-saving heuristic:

- ▶ Phase-saving uses cached assignments to variables when choosing a branch to explore.
- ▶ For small values, order encoding assigns a large fraction of the variables to false.
- ▶ Disable phase-saving and use negative branching: always explore the “false” branch first.

# SAT solving considerations

## Phase-saving heuristic:

- ▶ Phase-saving uses cached assignments to variables when choosing a branch to explore.
- ▶ For small values, order encoding assigns a large fraction of the variables to false.
- ▶ Disable phase-saving and use negative branching: always explore the “false” branch first.

## Heavy-tailed behavior:

- ▶ There is a large variance in runtime across different initial conditions of the search. Some runs have a nonnegligible chance of finishing early.
- ▶ To take advantage, run multiple instances of the SAT solver with varying seeds and different shufflings of the formula.

## Experiments with solver configurations

Interpretation	$D$	$V$	Phase-saving		Negative branching	
			Single	Parallel	Single	Parallel
Arctic	5	8	240.00s	240.00s	44.45s	<b>9.22s</b>
Arctic	3	5	1.52s	<b>0.13s</b>	29.95s	13.12s
Arctic	3	4	3.75s	<b>0.83s</b>	3.27s	1.71s
Natural	4	4	75.78s	19.12s	29.62s	<b>8.13s</b>
Natural	4	4	75.05s	<b>5.22s</b>	24.31s	6.43s
Arctic	4	3	3.33s	<b>0.52s</b>	11.55s	3.84s
Natural	3	11	240.00s	240.00s	240.00s	<b>79.05s</b>
Arctic	3	12	1.94s	<b>0.33s</b>	3.28s	0.38s

## More problems

**Mahler's 3/2 problem [1968].** Is there  $\xi \in \mathbb{R}_{>0}$  such that  $\text{frac}\left(\xi\left(\frac{3}{2}\right)^k\right) < \frac{1}{2}$  for all  $k \in \mathbb{N}$ ?

f▷ → 0▷	f0 → 0f	t0 → 1t	◁0 → ◁t
ft▷ → 10▷	f1 → 0t	t1 → 2f	◁1 → ◁ff
	f2 → 1f	t2 → 2t	◁2 → ◁ft

## More problems

**Mahler's 3/2 problem [1968].** Is there  $\xi \in \mathbb{R}_{>0}$  such that  $\text{frac} \left( \xi \left( \frac{3}{2} \right)^k \right) < \frac{1}{2}$  for all  $k \in \mathbb{N}$ ?

f▷ → 0▷	f0 → 0f	t0 → 1t	◁0 → ◁t
ft▷ → 10▷	f1 → 0t	t1 → 2f	◁1 → ◁ff
	f2 → 1f	t2 → 2t	◁2 → ◁ft

**Erdős on powers of 2 [1979].** When does the ternary expansion of  $2^n$  omit the digit 2?

0▷ → ▷	0f → f0	1t → t0	◁t → ◁0
1▷ → ▷	0t → f1	2f → t1	◁ff → ◁1
◁▷ → ◁▷	1f → f2	2t → t2	◁ft → ◁2

## Closing remarks

- ▶ Efficacy of techniques other than matrix interpretations
- ▶ Proving the *nonexistence* of matrix interpretations of any dimension for the Collatz system
- ▶ Moving from string to term rewriting
- ▶ Injecting specific knowledge about the problem into the rewriting system/termination proofs