Correctness of bbchallenge's deciders

bbchallenge's contributors

Abstract

The Busy Beaver Challenge (or bbchallenge) aims at collaboratively solving the following conjecture: "BB(5) = 47,176,870" [Aaronson, 2020]. This goal amounts to decide whether or not 88,664,064 Turing machines with 5-state halt or not – starting from all-0 tape. In order to decide the behavior of these machines we write *deciders*. A decider is a program that takes as input a Turing machine and outputs `true` if it is able to tell whether the machine halts or not. Each decider is specialised in recognising a particular type of behavior that can be decided.

In this document we are concerned with proving the correctness of these deciders programs. More context and information about this methodology are available at `https://bbchallenge.org`.

Contents

1 Conventions

The set $\mathbb{N}$ denotes $\{0, 1, 2 \dots\}$.

	0	1
A	1RB	1LC
B	1RC	1RB
C	1RD	0LE
D	1LA	1LD
E	- - -	0LA

Table 1: Transition table of the current 5-state busy beaver champion: it halts after 47,176,870 steps. `https://bbchallenge.org/1RB1LC1RC1RB1RD0LE1LA1LD---0LA&status=halt`

Turing machines. The Turing machines that are studied in the context of bbchallenge use a binary alphabet and a single bi-infinite tape. Machine transitions are either undefined (in which case the machine halts) or given by (a) a symbol to write (b) a direction to move (right or left) and (c) a state to go to. Table 1 gives the transition table of the current 5-state busy beaver champion. The machine halts after 47,176,870 steps (starting from all-0 tape) when it reads a 0 in state E, which is undefined.

A *configuration* of a Turing machine is defined by the 3-tuple: (i) state (ii) position of the head (iii) content of the memory tape. In the context of bbchallenge, *the initial configuration* of a machine is always (i) state is 0, i.e. the first state to appear in the machine's description (ii) head's position is 0 (iii) the initial tape is all-0 – i.e. each memory cell is containing 0. We write $c_1 \vdash_{\mathcal{M}} c_2$ if a configuration c_2 is obtained from c_1 in one computation step of machine $\mathcal{M}$. We omit $\mathcal{M}$ if it is clear from context. We let $c_1 \vdash^s c_2$ denote a sequence of s computation steps, and let $c_1 \vdash^* c_2$ denote zero or more computation steps. We write $c_1 \vdash \perp$ if the machine halts after executing one computation step from configuration c_1. In the context of bbchallenge, halting happens when an undefined machine transition is met i.e. no instruction is given for when the machine is in the state, tape position and tape corresponding to configuration c_1.

Space-time diagram. We use space-time diagrams to give a visual representation of the behavior of a given machine. The space-time diagram of machine $\mathcal{M}$ is an image where the i^{th} row of the image gives:

1. The content of the tape after i steps (black is 0 and white is 1).
2. The position of the head is colored to give state information using the following colours for 5-state machines: A, B, C, D, E.

2 Cyclers

The goal of this decider is to recognise Turing machines that cycle through the same configurations for ever. Such machines never halt. The method is simple: remember every configuration seen by a machine and return `true` if one is visited twice. A time limit (maximum number of steps) is also given for running the test in practice: the algorithm recognises any machine whose cycle fits within this limit[1].

Example 1. Figure 1 gives the space-time diagrams of the 30 first iterations of two "Cyclers" machines: bbchallenge's machines #279,081 (left) and #4,239,083 (right). Refer to `https://bbchallenge/279081` and `https://bbchallenge/4239083` for their transition tables. From these space-time diagrams we see that the machines eventually repeat the same configuration.

[1] In practice, for machines with 5 states the decider was run with 1000 steps time limit.

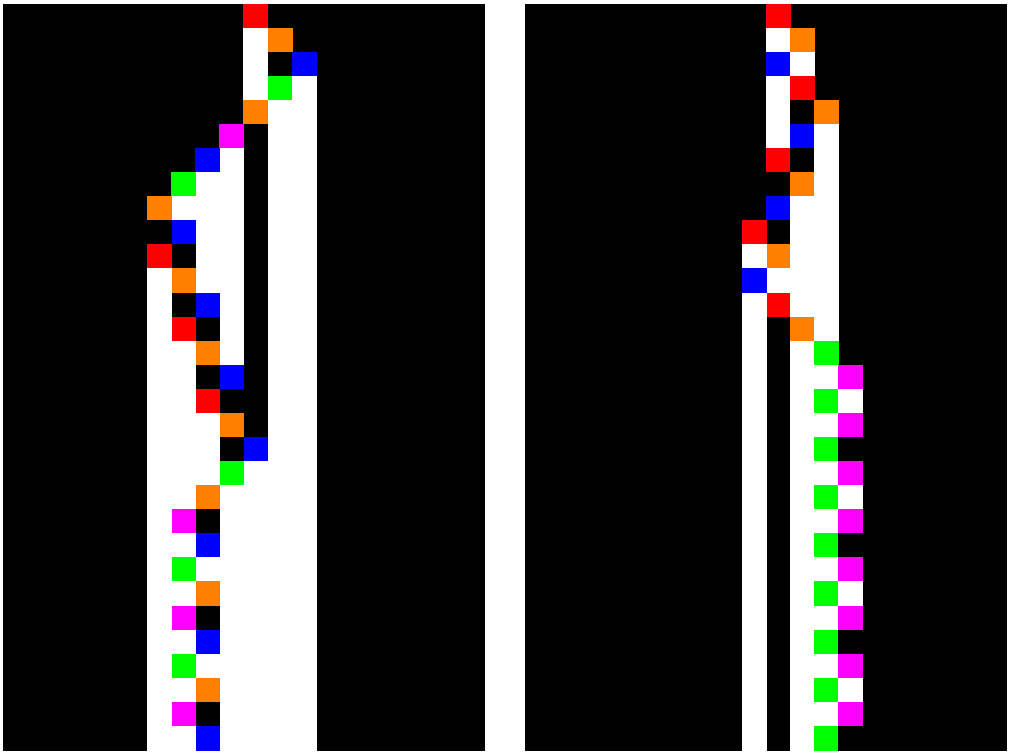


Figure 1: Space-time diagrams of the 30 first steps of bbchallenge's machines #279,081 (left) and #4,239,083 (right) which are both "Cyclers": they eventually repeat the same configuration for ever. Access the machines at `https://bbchallenge/279081` and `https://bbchallenge/4239083`.

2.1 Pseudocode

We assume that we are given a Turing Machine type **TM** that encodes the transition table of a machine as well as a procedure **TuringMachineStep**(machine,configuration) which computes the next configuration of a Turing machine from the given configuration or **nil** if the machine halts at that step.

2.2 Correctness

Theorem 2. Let $\mathcal{M}$ be a Turing machine and $t \in \mathbb{N}$ a time limit. Let c_0 be the initial configuration of the machine. There exists $i \in \mathbb{N}$ and $j \in \mathbb{N}$ such that $c_0 \vdash^i c_i \vdash^j c_i$ with $i + j \leq t$ if and only if DECIDER-CYCLERS($\mathcal{M}$,t) returns `true` (Algorithm 1).

Proof. This follows directly from the behavior of DECIDER-CYCLERS($\mathcal{M}$,t): all intermediate configurations below time t are recorded and the algorithm returns `true` if and only if one is visited twice. This mathematically translates to there exists $i \in \mathbb{N}$ and $j \in \mathbb{N}$ such that $c_0 \vdash^i c_i \vdash^j c_i$ with $i + j \leq t$, which is what we want. Index i corresponds to the first time that c_i is seen (l.13 in Algorithm 1) while index j corresponds to the second time that c_i is seen (l.11 in Algorithm 1). □

Corollary 3. Let $\mathcal{M}$ be a Turing machine and $t \in \mathbb{N}$ a time limit. If DECIDER-CYCLERS($\mathcal{M}$,t) returns `true` then the behavior of $\mathcal{M}$ from all-0 tape has been decided: $\mathcal{M}$ does not halt.

Proof. By Theorem 2, there exists $i \in \mathbb{N}$ and $j \in \mathbb{N}$ such that $c_0 \vdash^i c_i \vdash^j c_i$ with $i + j \leq t$. It follows that for all $k \in \mathbb{N}$, $c_0 \vdash^{i+kj} c_i$. The machine never halts as it will visit c_i infinitely often. □

2.3 Results

The decider was coded in `golang` and is accessible at this link: `https://github.com/bbchallenge/bbchallenge-deciders/tree/main/decider-cyclers`.

The decider found 11,229,238 "Cyclers", out of 88,664,064 machines in the seed database of the Busy Beaver Challenge (c.f. `https://bbchallenge.org/method#seed-database`). Time limit was set to 1000 and an additional memory limit (max number of visited cells) was set to 500. More information about these results are available at: `https://discuss.bbchallenge.org/t/decider-cyclers/33`.

Algorithm 1 DECIDER-CYLERS

```
1: struct Configuration {
2:     int state
3:     int headPosition
4:     int → int tape
5: }
6:
7: procedure bool DECIDER-CYLERS(TM machine,int timeLimit)
8:     Configuration currConfiguration = {.state = 0, .headPosition = 0, .tape = {0:0}}
9:     Set<Configuration> configurationsSeen = {}
10:    int currTime = 0
11:    while currTime < timeLimit do
12:        if currConfiguration in configurationsSeen then
13:            return true
14:        configurationsSeen.insert(currConfiguration)
15:        currConfiguration = TuringMachineStep(machine,currConfiguration)
16:        currTime += 1
17:        if currConfiguration == nil then
18:            return false //machine has halted, it is not a Cycler
19:    return false
```

3 Translated cyclers

The goal of this decider is to recognise Turing machines that translate a bounded pattern for ever. We call such machines "Translated cyclers". They are close to "Cyclers" (Section 2) in the sense that they are only repeating a pattern but there is added complexity as they are able to translate the pattern in space at the same time, hence the decider for Cyclers cannot directly apply here.

The main idea for this decider is illustrated in Figure 2 which gives the space-time diagram of a "Translated cycler': bbchallenge's machine #44,394,115 (c.f. `https://bbchallenge.org/44394115`). The idea is to find two configurations that break a record (i.e. visit a memory cell that was never visited before) in the same state (here state D) such that the content of the memory tape at distance L from the record positions is the same in both record configurations. Distance L is defined as being the maximum distance to record position 1 that was visited between the configuration of record 1 and record 2. In those conditions, we can prove that the machine will never halt.

The translated cycler of Figure 2 features a relatively simple repeating pattern and transient pattern (pattern occurring before the repeating patterns starts). These can get significantly more complex, bbchallenge's machine #59,090,563 is an example see Figure 3 and `https://bbchallenge.org/59090563`. The method for detecting the behavior is the same but more resources are needed.

3.1 Pseudocode

We assume that we are given a Turing Machine type **TM** that encodes the transition table of a machine as well as a procedure **TuringMachineStep**(machine,configuration) which computes the next configuration of a Turing machine from the given configuration or **nil** if the machine halts at that step.

One minor complication of the technique described above is that one has to track record-breaking configurations on both sides of the tape: a configuration can break a record on the right or on the left. Also, in order to compute distance L (see above or Definition 5) it is useful to add to memory cells the information of the last time step at which it was visited.

We also assume that we are given a routine GET-EXTREME-POSITION(tape,sideOfTape) which gives us the rightmost or leftmost position of the given tape (well defined as we always manipulate finite tapes).

3.2 Correctness

Definition 4 (record-breaking configurations)**.** Let $\mathcal{M}$ be a Turing machine and c_0 its busy beaver initial configuration (i.e. state is 0, head position is 0 and tape is all-0). Let c be a configuration reachable from c_0, i.e. $c_0 \vdash^* c$. Then c is said to be *record-breaking* if the current head position had never been visited before. Records can be broken to the *right* (positive head position) or to the left (negative head position).

Algorithm 2 DECIDER-TRANSLATED-CYLERS

```
1: const int RIGHT, LEFT = 0, 1
2: struct ValueAndLastTimeVisited {
3:     int value
4:     int lastTimeVisited
5: }
6: struct Configuration {
7:     int state
8:     int headPosition
9:     int → ValueAndLastTimeVisited tape
10: }
11:
12: procedure bool DECIDER-TRANSLATED-CYLERS(TM machine,int timeLimit)
13:     Configuration currConfiguration = {.state = 0, .headPosition = 0, .tape = {0:{.value = 0, .lastTimeVisited = 0}}}
14:     // 0: right records, 1: left records
15:     List<Configuration> recordBreakingConfigurations[2] = [[],[]]
16:     int extremePositions[2] = [0,0]
17:     int currTime = 0
18:     while currTime < timeLimit do
19:         int headPosition = currConfiguration.headPosition
20:         currConfiguration.tape[headPosition].lastTimeVisited = currTime
21:         if headPosition > extremePositions[RIGHT] or headPosition < extremePositions[LEFT] then
22:             int recordSide = (headPosition > extremePositions[RIGHT]) ? RIGHT : LEFT
23:             extremePositions[recordSide] = headPosition
24:             if CHECK-RECORDS(currConfiguration, recordBreakingConfigurations[recordSide], recordSide) then
25:                 return true
26:             recordBreakingConfigurations[recordSide].append(currConfiguration)
27:         currConfiguration = TuringMachineStep(machine,currConfiguration)
28:         currTime += 1
29:         if currConfiguration == nil then
30:             return false //machine has halted, it is not a Translated Cycler
31:     return false
```

Algorithm 3 COMPUTE-DISTANCE-L and AUX-CHECK-RECORDS

```
1: procedure int COMPUTE-DISTANCE-L(Configuration currRecord, Configuration olderRecord, int recordSide)
2:     int olderRecordPos = olderRecord.headPosition
3:     int olderRecordTime = olderRecord.tape[olderRecordPos].lastTimeVisited
4:     int currRecordTime = currRecord.tape[currRecord.headPosition].lastTimeVisited
5:     int distanceL = 0
6:     for int pos in currRecord.tape do
7:         if pos > olderRecordPos and recordSide == RIGHT then continue
8:         if pos < olderRecordPos and recordSide == LEFT then continue
       int lastTimeVisited = currRecord.tape[pos].lastTimeVisited
9:         if lastTimeVisited ≥ olderRecordTime and lastTimeVisited ≤ currRecordTime then
10:            distanceL = max(distanceL,abs(pos-olderRecordPos))
11:    return distanceL
12:
13: procedure bool AUX-CHECK-RECORDS (Configuration currRecord, List<Configuration> olderRecords, int recordSide)
14:    for Configuration olderRecord in olderRecords do
15:        if currRecord.state != olderRecord.state then
16:            continue
17:        int distanceL = COMPUTE-DISTANCE-L(currRecord,olderRecord,recordSide)
18:        int currExtremePos = GET-EXTREME-POSITION(currRecord.tape,recordSide)
19:        int olderExtremePos = GET-EXTREME-POSITION(olderRecord.tape,recordSide)
20:        int step = (recordSide == RIGHT) ? -1 : 1
21:        bool isSameLocalTape = true
22:        for int offset = 0; abs(offset) < distanceL; offset += step do
23:            if currRecord.tape[currExtremePos+offset] != olderRecord.tape[olderExtremePos+offset] then
24:                isSameLocalTape = false
25:                break
26:        if isSameLocalTape then
27:            return true
28:    return false
```

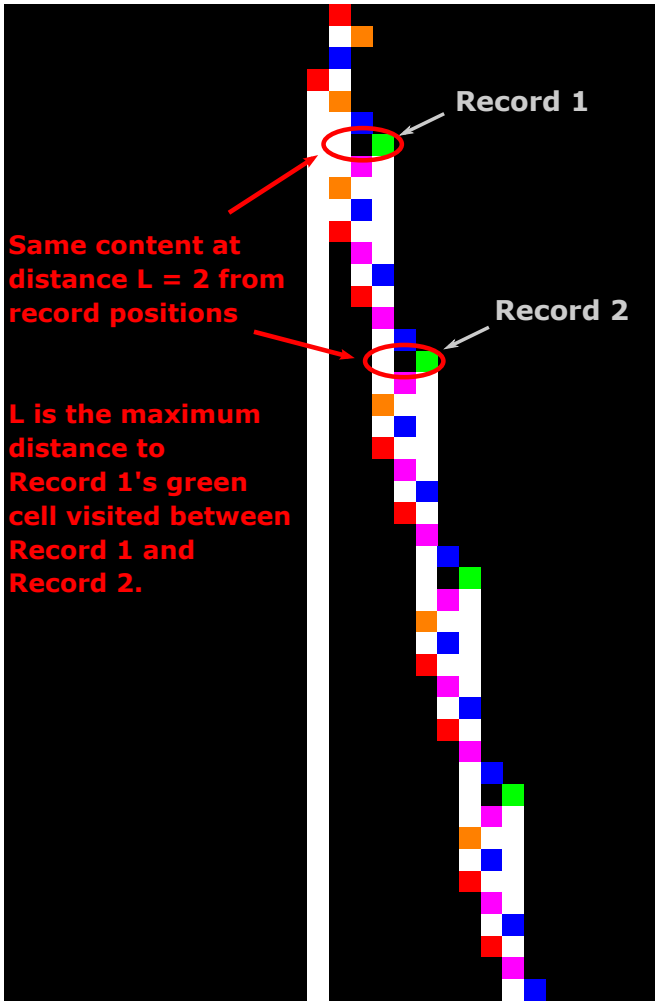


Figure 2: Example "Translated cycler": 45-step space-time diagram of bbchallenge's machine #44,394,115. See `https://bbchallenge.org/44394115`. The same bounded pattern is being translated to the right for ever. The text annotations illustrate the main idea for recognising "Translated Cyclers": find two configurations that break a record (i.e. visit a memory cell that was never visited before) in the same state (here state D) such that the content of the memory tape at distance L from the record positions is the same in both record configurations. Distance L is defined as being the maximum distance to record position 1 that was visited between the configuration of record 1 and record 2.

Definition 5 (Distance L between record-breaking configurations)**.** Let $\mathcal{M}$ be a Turing machine and r_1, r_2 be two record-breaking configurations on the same side of the tape at respective times t_1 and t_2 with $t_1 < t_2$. Let p_1 and p_2 be the tape positions of these records. Then, distance L between r_1 and r_2 is defined as $\max\{|p_1 - p|\}$ with p any position visited by $\mathcal{M}$ between t_1 and t_2 that is not beating record p_1 (i.e. $p \leq p_1$ for a record on the right and $p \geq p_1$ for a record on the left).

Lemma 6. Let $\mathcal{M}$ be a Turing machine. Let r_1 and r_2 be two configurations that broke a record in the same state and on the same side of the tape at respective times t_1 and t_2 with $t_1 < t_2$. Let p_1 and p_2 be the tape positions of these records. Let L be the distance between r_1 and r_2 (Definition 5). If the content of tape in r_1 at distance L of p_1 is the same than the content of the tape in r_2 at distance L of p_2 then $\mathcal{M}$ never halts. Furthermore, by Definition 5, we know that distance L is the maximum distance that $\mathcal{M}$ can travel to the left of p_1 between times t_1 and t_2.

Proof. Let's suppose that the record-breaking configurations are on the right-hand side of the tape. By the hypotheses, we know the machine is in the same state in r_1 and r_2 and that the content of the tape at distance L to the left of p_1 in r_1 is the same as the content of the tape at distance L to the left of p_2 in r_2. Note that the content of the tape to the right of p_1 and p_2 is the same: all-0 since they are record positions. Hence that after r_2, since it will read the same tape content the machine will reproduce the same behavior than it did after r_1 but translated at position p_2: there will a record-breaking configuration

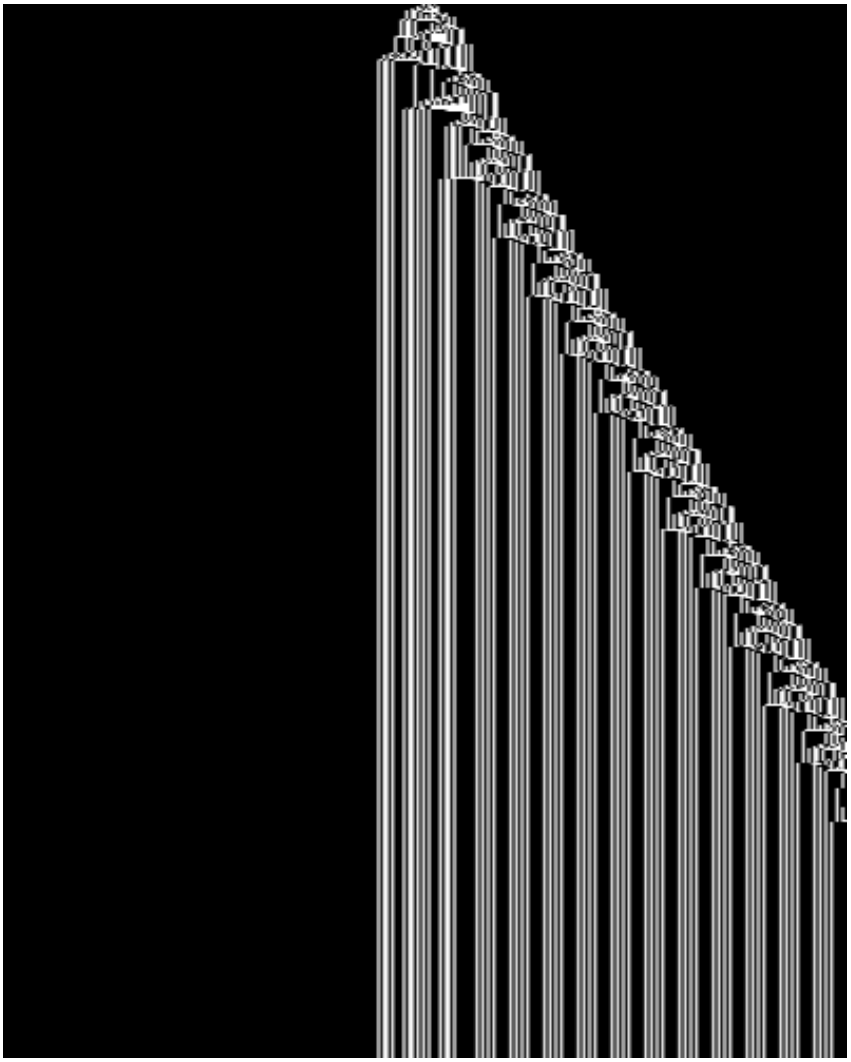


Figure 3: More complex "Translated cycler": 10,000-step space-time diagram (no state colours) of bbchallenge's machine #59,090,563. See `https://bbchallenge.org/59090563`.

r_3 such that the distance between record-breaking configurations r_2 and r_3 is also L (Definition 5). Hence the machine will keep breaking records to the right for ever and will not halt. Analogous proof for records that are broken to the left. □

Theorem 7. Let $\mathcal{M}$ be a Turing machine and t a time limit. The conditions of Lemma 6 are met before time t if and only if DECIDER-TRANSLATED-CYCLERS($\mathcal{M}$,t) outputs `true` (Algorithm 2).

Proof. The algorithm consists of a main function DECIDER-TRANSLATED-CYCLERS (Algorithm 2) and two auxiliary functions COMPUTE-DISTANCE-L and AUX-CHECK-RECORDS (Algorithm 3).

The main loop of DECIDER-TRANSLATED-CYCLERS (Algorithm 2 l.17) simulates the machine with the particularity that (a) it keeps track of the last time it visited each memory cell (l.19) and (b) it keeps track of all record-breaking configurations that are met (l.20) before reaching time limit t. When a record-breaking configuration is found, it is compared to all the previous record-breaking configurations on the same side in seek of the conditions of Lemma 6. This is done by auxiliary routine AUX-CHECK-RECORDS (Algorithm 3).

Auxiliary routine AUX-CHECK-RECORDS (Algorithm 3, l.12) loops over all older record-breaking configurations on the same side than the current one (l.13). The routine ignores older record-breaking configurations that were not in the same state than the current one (l.14). If the states are the same, it computes distance L (Definition 5) between the older and the current record-breaking configuration (l.16). This computation is done by auxiliary routine COMPUTE-DISTANCE-L.

Auxiliary routine COMPUTE-DISTANCE-L (Algorithm 3, l.1) uses the "pebbles" that were left on the tape to give the last time a memory cell was seen (field `lastTimeVisited`) in order to compute the farthest position from the old record position that was visited before meeting the new record position (l.10). Note that we discard intermediate positions that beat the old record position (l.7-8) as we know that the part of the tape after the record position in the old record-breaking configuration is all-0, same as the part of the tape after current record position in the current record-breaking position (part of the tape to the right of the red-circled green cell in Figure 2).

Thanks to the computation of COMPUTE-DISTANCE-L the routine AUX-CHECK-RECORDS is able to check whether the tape content at distance L of the record-breaking position in both record-holding configurations is the same or not (Algorithm 3, l.22). The routine returns `true` if they are the same and the function DECIDER-TRANSLATED-CYCLERS will return `true` as well in cascade (Algorithm 2 l.24). That scenario is reached if and only if the algorithm has found two record-breaking configurations on the same side that satisfy the conditions of Lemma 6, which is what we wanted.

□

Corollary 8. Let $\mathcal{M}$ be a Turing machine and $t \in \mathbb{N}$ a time limit. If DECIDER-TRANSLATED-CYCLERS($\mathcal{M}$,t) returns `true` then the behavior of $\mathcal{M}$ from all-0 tape has been decided: $\mathcal{M}$ does not halt.

Proof. Immediate by combining Lemma 6 and Theorem 7. □

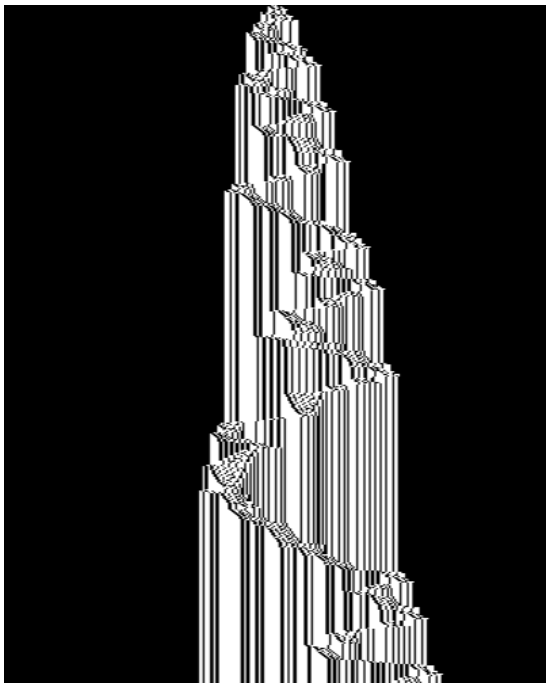
3.3 Results

The decider was coded in `golang` and is accessible at this link: `https://github.com/bbchallenge/bbchallenge-deciders/tree/main/decider-translated-cyclers`.

The decider found 73,860,604 "Translated cyclers", out of 88,664,064 machines in the seed database of the Busy Beaver Challenge (c.f. `https://bbchallenge.org/method#seed-database`). Time limit was set to 1000 in a first run then increased to 10000 for the remaining machines and an additional memory limit (max number of visited cells) was set to 500 then 5000. More information about these results are available at: `https://discuss.bbchallenge.org/t/decider-translated-cyclers/34`.

4 Backward Reasoning

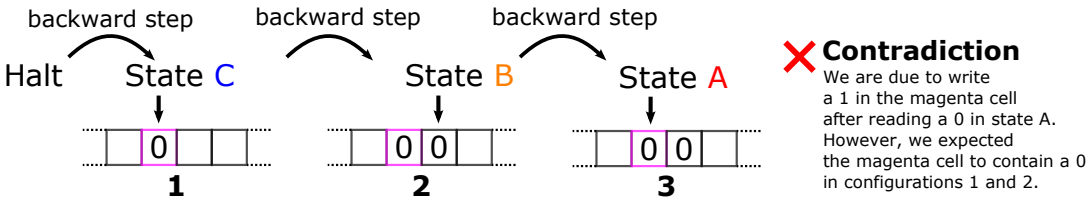
Backward reasoning, as described in [**?**], takes a different approach than what has been done with deciders in Sections 2 and 3. Indeed, instead of trying to recognise a particular kind of machine's behavior, the idea of backward reasoning is to show that, independently of the machine's behavior, the halting configurations are not reachable. In order to do so, the decider simulates the machine *backwards* from halting configurations until it reaches some obvious contradiction.



(a) 10,000-step space-time diagram of bbchallenge's machine #55,897,188. `https://bbchallenge.org/55897188`

	0	1
A	1RB	0LD
B	1LC	0RE
C	- - -	1LD
D	1LA	1LD
E	1RA	0RA

(b) Transition table of machine #55,897,188.



(c) Contradiction reached after 3 backward steps: machine #55,897,188 does cannot reach its halting configuration hence it does not halt.

Figure 4: Applying backward reasoning on bbchallenge's machine #55,897,188. (a) 10,000-step space-time diagram of machine #55,897,188. The *forward* behavior of the machine looks very complex. (b) Transition table. (c) We are able to deduce that the machine will never halt thanks to only 3 backward reasoning steps: because a contradiction is met, it is impossible to reach the halting configuration in more than 3 steps – and, by (a), the machine can do at least 20,000 without halting starting from all-0 tape.

Figure 4 illustrates this idea on bbchallenge's machine #55,897,188. From the space-time diagram, the *forward* behavior of the machine from all-0 tape looks to be extremely complex, Figure 4a. However, by reconstructing the sequence of transitions that would lead to the halting configuration (reading a 0 in state C), we reach a contradiction in only 3 steps, Figure 4c. Indeed, the only way to reach state C is to come from the right in state B where we read a 0. The only way to reach state B is to com from left in state A where we read a 0. However, the transition table (Figure 4b) is instructing us to write a 1 in that case, which is not consistent with the 0 that we assumed was at position in order for the machine to halt.

Backward reasoning in the case of Figure 4 was particularly simple because there was only one possible previous configuration for each backward step – e.g. there is only one transition that can reach state C and same for state B. In general, this is not the case and the structure created by backward reasoning is a tree of configurations instead of just a chain. If all the leaves of a backward reasoning tree of depth D reach a contradiction, we know that if the machine runs for D steps from all-0 tape then the machine cannot reach a halting configuration and thus does not halt.

4.1 Pseudocode

We assume that we are given routine GET-UNDEFINED-TRANSITIONS(machine) which returns the list of (state,readSymbol) pairs of all the undefined transitions in the machine's transition table, for instance [(C,0)] for the machine of Figure 4b. We also assume that we are given routine GET-TRANSITIONS-REACHING-STATE(machine,targetState) which returns the list of all machine's transitions that go to the specified target state, for instance [(A,1,0LD),(C,1,1LD),(D,1,1LD)] for target state D in the machine of Figure 4b. These two routines contain very minimal logic as they only lookup in the description of the machine for the required information.

4.2 Correctness

Theorem 9. Let $\mathcal{M}$ be a Turing machine and $D \in \mathbb{N}$. Then, DECIDER-BACKWARD-REASONING($\mathcal{M}$,D) returns `true` if and only if no undefined transition of $\mathcal{M}$ can be reached in more than D steps.

Proof. The tree of backward configurations is maintained in a DFS fashion through a stack (Algorithm 4, l.24). Initially, the stack is filled with the configurations where only one tape cell is defined and state is set such that the corresponding transition is undefined (i.e. the machine halts after that step), l.25-28.

Then, the main loop runs until either (a) the stack is empty or (b) one leaf exceeded the maximum allowed depth, l.30 and l.32. Note that running the algorithm with increased maximum depth increases its chances to contradict all branches of the backward simulation tree. At each step of loop, we remove the current configuration from the stack and we try to apply all the transitions that leads to its state backwards by calling routine APPLY-TRANSITION-BACKWARDS(configuration, transition).

The only case where it is not possible to apply a transition backwards, i.e. the case where a contradiction is reached is when the tape symbol at the position where the transition comes from (i.e. to the right if transition movement is left and vice-versa) is defined but is not equal to the write instruction of the transition. Indeed, that means that the future (i.e. previous backward steps) is not consistent the current transition's write instruction. This logic is checked l.16. Otherwise, we can construct the previous configuration (i.e. next backward step) and augment depth by 1. We then stack this configuration in the main routine (l.39).

The algorithm returns `true` if and only if the stack ever becomes empty which means that all leaves of the backward simulation tree of depth D have reached a contradiction and thus, no undefined transition of the machine is reachable in more than D steps.

This pseudocode contains a slight optimisation with the use of set `configurationSeen` (l.29). This set racks configurations which would have already been seen in different branches of the tree in order not traverse them twice (l.32-33). While not needed in theory, this optimisation is useful in practice, especially at large depths (e.g. $D = 300$). □

Corollary 10. Let $\mathcal{M}$ be a Turing machine and $D \in \mathbb{N}$. If DECIDER-BACKWARD-REASONING($\mathcal{M}$,D) returns `true` and machine $\mathcal{M}$ can run D steps from all-0 tape without halting then the behavior of $\mathcal{M}$ from all-0 tape has been decided: $\mathcal{M}$ does not halt.

Proof. By Theorem 9 we know that no undefined transition of $\mathcal{M}$ can be reached in more than D steps. Hence, if machine $\mathcal{M}$ can run D steps from all-0 tape without halting, it will be able to run the next $D+1^{\text{th}}$ step. From there, the machine cannot halt or it would contradict the fact that halting trajectories have at most D steps. Hence, $\mathcal{M}$ does not halt from all-0 tape. □

Algorithm 4 DECIDER-BACKWARD-REASONING

```
1: const int RIGHT, LEFT = 0, 1
2: struct Transition {
3:     int state, read, write, move
4: }
5: struct Configuration {
6:     int state
7:     int headPosition
8:     int → int tape
9:     int depth
10: }
11:
12: procedure Configuration APPLY-TRANSITION-BACKWARDS(Configuration conf,Transition t)
13:     int reversedHeadMoveOffset = (t.move == RIGHT) ? -1 : 1
14:     int previousPosition = conf.headPosition+reversedHeadMoveOffset
15:     // Backward contradiction spotted
16:     if previousPosition in conf.tape and conf.tape[previousPosition] != t.write then
17:         return nil
18:     Configuration previousConf = {.state = t.state, .depth = conf.depth + 1, .tape = conf.tape}
19:     previousConf.headPosition = previousPosition
20:     previousConf.tape[previousPosition] = t.read
21:     return previousConf
22:
23: procedure bool DECIDER-BACKWARD-REASONING(TM machine,int maxDepth)
24:     Stack<Configuration> configurationStack
25:     for int (state,read) in GET-UNDEFINED-TRANSITIONS(machine) do
26:         Configuration haltingConfiguration = {.state = state,.depth=0,.headPosition = 0}
27:         haltingConfiguration.tape = {0: read}
28:         configurationStack.push(haltingConfiguration)
29:     Set<Configuration> configurationsSeen = {}
30:     while !configurationStack.empty() do
31:         Configuration currConf = configurationStack.pop()
32:         if currConf.depth > maxDepth then return false
33:         if  currConf in configurationsSeen then continue
34:         configurationsSeen.insert(currConf)
35:         List<Configuration> confList = []
36:         for Transition transition in GET-TRANSITIONS-REACHING-STATE(machine,currConf.state) do
37:             Configuration previousConf = APPLY-TRANSITION-BACKWARDS(currConf, transition)
38:             // If no contradiction
39:             if previousConf != nil then
40:                 configurationStack.push(previousConf)
41:     return true
```

4.3 Results

The decider was coded in `golang` and is accessible at this link: `https://github.com/bbchallenge/bbchallenge-deciders/blob/main/decider-backward-reasoning`. Note that collaborative work allowed to find a bug in the initial algorithm that was implemented[2].

The decider decided 2,035,598 machines, out of 3,574,222 machines that were left after deciders for "Cyclers" and "Translated Cyclers" (Section 2 and Section 7). Maximum depth was set to 300. More information about these results are available at: `https://discuss.bbchallenge.org/t/decider-backward-reasoning/35`.

5 Halting Segment

Acknowledgement. Sincere thanks to bbchallenge's contributor Iijil who initially presented this method and the first implementation[3]. Other contributors have contributed to this method by producing alternative implementations (see Section 5.3) or discussing and writing the formal proof presented here: Mateusz Naściszewski (Mateon1), Nathan Fenner, Tony Guilfoyle, Justin Blanchard and cosmo.

5.1 Overview

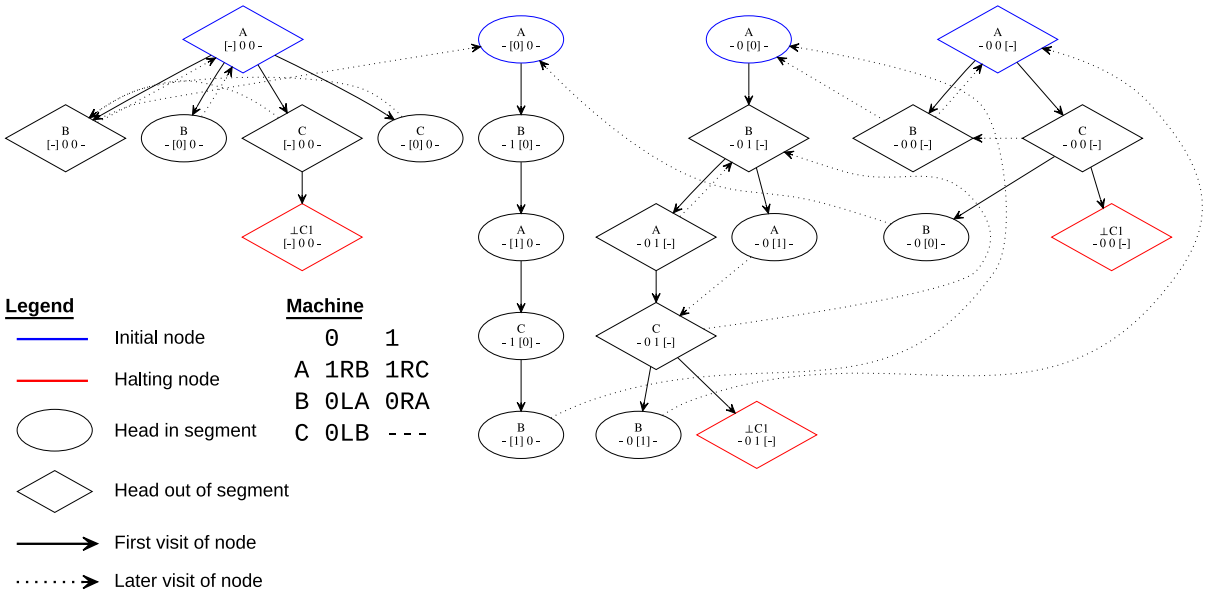


Figure 5: Halting Segment graph for the 3-state machine `https://bbchallenge.org/1RB1RC_0LA0RA_0LB---` and segment size 2, see Definition 13. Nodes of this graph correspond to *segment configurations* (Definition 11), i.e. configurations of the machine on a finite segment (here, of size 2). In a node, the machine's head position is represented between brackets and the symbol - represents the outside of the segment (either to the left or to the right). Nodes where the machine's head is within the segment (circle shape) only one have child corresponding to the next step of the machine and nodes where the head is outside of the segment (diamond shape) may have multiple children corresponding to all the theoretically possible ways (deduced from the machine's transition table) that the machine can enter the segment back or continue to stay out of it. In order to improve readibility, edges that revisit a node are dotted. The machine presented here does not halt because the halting nodes (red outline) that are reachable from the initial nodes (blue outline) do not cover all the positions of the segment (there is no halting node for any of the two internal positions of the segment), by contraposition of Theorem 14.

The idea of the Halting Segment technique is to simulate a Turing machine on a finite segment of tape. When the machine leaves the segment in a certain state, we consider all the possible ways that it can re-enter the segment or stay out of it, based on the machine's transition table. For a given machine and

[2]Thanks to collaborators `https://github.com/aticuscull` and `https://github.com/modderme123`.
[3]See: `https://discuss.bbchallenge.org/t/decider-halting-segment`.

segment size, this method naturally gives rise to a graph, the Halting Segment graph (formally defined in Definition 13).

Figure 5 gives the Halting Segment graph of the 3-state machine[4] `https://bbchallenge.org/1RB1RC_0LA0RA_0LB---` for segment size 2. Let's describe this graph in more details:

- Nodes correspond to *segment configurations* (Definition 11), i.e. the state in which the machine is together with the content of the segment and the position of the head in the segment (or outside of it). For instance, the leftmost node in blue and diamond shape in Figure 5 is `A [-] 0 0 -` which means that the machine is in state A, that the segment currently contains `0 0` and that the machine's head is currently outside of the segment, to the left of it.
- Initial nodes (blue outline) correspond to all sgement configurations that match the initial configuration of the machine (all-0 tape and state A), there are $n+2$ initial nodes with n the size of the segment. Halting nodes (red outline) give the segment configurations where the machine has halted together with the halting transition that was used, for instance, in Figure 5, the leftmost halting node $\perp$ `C1 [-] 0 0 -` signifies that the machine has halted ($\perp$), using halting transition `C1` (reading a 1 in state C), to the left of the segment which contains `0 0`.
- Nodes with a circle shape correspond to segment configurations where the tape's head is **inside** the segment. Such nodes only have one child, which corresponds to the next machine configuration.
- Nodes with a diamond shape correspond to segment configurations where the head is **outside** the segment, these nodes may have several children corresponding to all the ways that the head, in the current state, can stay outside of the segment or enter it back. For instance, the leftmost node in blue and diamond shape in Figure 5, `A [-] 0 0 -`, has 4 children: `B [-] 0 0 -` and `B - [0] 0 -` and `C [-] 0 0 -` and `C - [0] 0 -`. This is because the transitions of the machine in state A are `1RB` and `1RC` and that the move `R` allows either to enter the segment back or to continue being out of it (if the head is far from the segment's left frontier). Note that the write symbol `1` of the transitions are ignored since we do not keep track of the tape outside of the segment.
- In order to increase the readability of Figure 5, only one entrant edge for each node has been drawn with a solid line, corresponding to the first visit of that node in the particular order that the graph was visited. Later visits were drawn with a dotted line.

What is special about the Halting Segment graph? We show in Theorem 14 that if a machine halts, then, for all segment size, its Halting Segment graph contains a set of halting nodes (red outline), for the same halting transition, that covers the entire segment and its outside, i.e. such that there is at least one such node per segment's position and outside of it (left and right). By contraposition, if there is no set of covering halting nodes for a halting transition, the machine does not halt. In Figure 5, we deduce that machine `https://bbchallenge.org/1RB1RC_0LA0RA_0LB---` does not halt since the halting nodes of halting transition `C1` are $\perp$ `C1 [-] 0 0 -`, $\perp$ `C1 - 0 1 [-]` and $\perp$ `C1 - 0 0 [-]` which does not cover the entire segment (both internal segment postions are not covered).

Interestingly, Halting Segment is the method that was used by Newcomb Greenleaf to prove[5] that Marxen & Buntrock's chaotic machine[6] [**?**] does not halt.

5.2 Formal proof

Definition 11 (Segment configurations)**.** Let $n \in \mathbb{N} = \{0, 1, 2 \dots\}$ a natural number called *segment size*. A *segment configuration* is a 3-tuple: (i) state, (ii) $w \in \{0,1\}^n$ which is the segment's content and (iii) the position of the machine's head is an integer $p \in [\![-1, n]\!]$ where positions $[\![0, n[\![$ correspond to the interior of the segment, position -1 for outside to the left and n for outside to the right. *Halting segment configurations* are segment configurations where the state is $\perp$ and with an additional information (iv) of which halting transition of the machine has been used to halt.

Example 12. In Figure 5 we have $n = 2$ and, the leftmost node in blue and diamond shape corresponds to segment configuration `A [-] 0 0 -` (i) state A, (ii) $w =$ `00` and (iii) $p = -1$. The rightmost node in red and diamond shape corresponds to halting segment configuration $\perp$ `C1 - 0 0 [-]` (i) state $\perp$, (ii) $w =$ `00`, (iii) $p = 2$ and (iv) halting transition `C1`.

[4] We chose a 3-state machine in order to have a graph of reasonable size.

[5] `http://turbotm.de/~heiner/BB/TM4-proof.txt`

[6] `https://bbchallenge.org/76708232`

Definition 13 (Halting Segment graph)**.** Let $\mathcal{M}$ be a Turing machine and $n \in \mathbb{N}$ a segment size. The Halting Segment graph for M and n is a directed graph where the nodes are segment configurations (Definition 11). The graph is generated from $n+2$ *initial nodes* (blue outline in Figure 5) that are all in state A with segment content 0^n (n consecutive 0s) but where the head is at each of the $n+2$ possible positions, one per each initial node, see the blue nodes in Figure 5 for an example. Then, edges that go out of a given node r are defined as follows:

- If r's head position is inside the segment (circle nodes in Figure 5), then r only has one child corresponding to the next simulation step for machine $\mathcal{M}$. For instance, in Figure 5, node `A - [0] 0 -` has a unique child `B - 1 [0] -`, following machine's transition `A0` which is `1RB`. That child can be a halting segment configuration if the transition to take is halting.
- If r's head position is outside the segment (diamond nodes in Figure 5), then, we consider each transition of r's state. There are three cases:
  1. If the transition is halting, we add a child to r which is the halting segment configuration node corresponding to this transition. For instance, in Figure 5, `C [-] 0 0 -` has halting child $\perp$ `C1 [-] 0 0 -` corresponding to halting transition `C1`.
  2. If the transition's movement goes further away from the segment (e.g. we are to the left of the segment, $p = -1$, and the transition movement is `L`), we add one child for this transition that only differs from its parent in the new state that it moves into. For instance, in Figure 5, `A - 0 0 [-]` has child $\perp$ `B - 0 0 [-]` for transition `A0` which is `1RB`.
  3. If the transition's movement goes in the direction of the segment (e.g. we are to the left of the segment, $p = -1$, and the transition movement is `R`), we add two children for this transition. One corresponding to the case where that movement is made at the border of the segment and allows to re-enter the segment and the other one corresponding to the case where that movement is made farther away from the border and does not re-enters yet. For instance, in Figure 5, node `A [-] 0 0 -` has children `B [-] 0 0 -` and `B - [0] 0 -` for transition `A0` which is `1RB`.

Halting nodes are nodes corresponding to halting segment configurations (red outline in Figure 5).

Theorem 14 (Halting Segment)**.** Let $\mathcal{M}$ be a Turing machine and $n \in \mathbb{N}$ a segment size. Let G be the Halting Segment graph for $\mathcal{M}$ and n (Definition 13). If $\mathcal{M}$ halts in halting transition T when started from state A and all-0 tape, then G must contain a halting node for transition T for each of the $n+2$ possible values of the head's position $p \in [\![-1, n]\!]$.

Proof. Consider the trace of configurations of $\mathcal{M}$ (full configurations, not segment configurations, as defined in Section 2) from the initial configuration (state A and all-0 tape) to the halting configuration which happens using halting transition T. Starting from the halting configuration, construct the halting segment configuration (with segment size n) for T using any position $p \in [\![-1, n]\!]$ in the segment and fill the segment's content from what is written on the tape around the head in the halting configuration of $\mathcal{M}$. From there, work your way up to the initial configuration: at each step construct the associated segment configuration. This sequence of segment configurations constitute a set of nodes in the Halting Segment graph G of $\mathcal{M}$ for segment size n such that each node points to the next one. At the top of that chain there will be a node matching the initial configuration: state A, all-0 segment and head position somewhere in $[\![-1, n]\!]$, i.e. an initial node.

Hence we have shown that all halting nodes for transition T for each of the $n+2$ possible values of the head's position $p \in [\![-1, n]\!]$ are reachable from some initial node(s). □

Remark 15. By contraposition of Theorem 14, if, for all halting transitions T there is at least one halting node (red outline in Figure 5) for some position in the segment that is not reachable from one of the initial node (blue outline in Figure 5) then the machine does not halt. That way, in Figure 5, we can conclude that machine`https://bbchallenge.org/1RB1RC_0LA0RA_0LB---` does not halt since the halting nodes of halting transition `C1` are $\perp$ `C1 [-] 0 0 -`, $\perp$ `C1 - 0 1 [-]` and $\perp$ `C1 - 0 0 [-]` which does not cover the entire segment (both internal segment postions are not covered).

Note that if all of the segment's positions are covered for some halting transition, we cannot conclude that the machine does not halt, but it does not mean that the machine necessarily halts either.

Remark 16. Some non-halting machines cannot be decided using Halting Segment for any segment size. Such a machine is for instance `https://bbchallenge.org/1RB---_1LC0RB_1LB1LA`.

5.3 Implementations and results

Here are the implementations of the method that were realised, almost all of them construct the Halting Segment graph from the halting nodes (backward implementation) instead than from the initial nodes (forward implementation):

1. Iijil's who originally proposed the method, `https://github.com/bbchallenge/bbchallenge-deciders/tree/main/decider-halting-segment`, and was independently reproduced by cosmo `https://github.com/bbchallenge/bbchallenge-deciders/tree/main/decider-halting-segment-reproduction` (backward implementation)
2. Mateusz Naściszewski (Mateon1)'s: `https://gist.github.com/mateon1/7f5e10169abbb50d1537165c6e71733b` (forward implementation)
3. Nathan Fenner's which has the interesting feature of being written in a language for formal verification (Dafny): `https://github.com/Nathan-Fenner/bbchallenge-dafny-deciders/blob/main/halting-segment.dfy` (backward implementation)
4. Tony Guilfoyle: `https://github.com/TonyGuil/bbchallenge/tree/main/HaltingSegments` (backward implementation)

We will be only discussing the details and results of Iijil's implementation (1) as it was the first implementation to be proposed and that it also was reproduced independently with exactly matching results.

This implementation is a bit different from what is presented in this document because the Halting Segment graph is constructed backward (i.e. from the halting nodes instead than from the initial nodes). Also, the method adopts a lazy strategy consisting in testing only odd segment sizes (up to size $n_{\max}$) and placing the head's position at the center of the tape. Finally, the information of state is not stored for nodes where the head is outside the segment. These implementation choices make the implementation a bit weaker than what was presented here.

Nonetheless, results are impressive, for $n_{\texttt{max}} = 13$, the method decides 1,002,808 machines out of the 1,538,624 remaining after backward reasoning (see Section 4.3). Hence, after Halting Segment, we have 535,816 machines left to be decided[7].

6 Finite automata reduction

6.1 Method overview

The core idea of the method presented in this Section is to find, for a given Turing machine, a regular language that describes the set (or a superset) of the machine's eventually-halting configurations– with finite support[8]. Then, we only have to test that the initial all-0 configuration is not part of the language to deduce that the machine does not halt from it.

This idea has been explored by other authors under the name Closed Tape Languages (CTL) as described in S. Ligocki's blog (`https://www.sligocki.com/2022/06/10/ctl.html`) and credited to H. Marxen in collaboration with J. Buntrock.

Here, we develop an original technique, based on the algebraic description of Nondeterministic Finite Automata (NFA), for finding the regular language associated to a machine's eventually halting configurations.

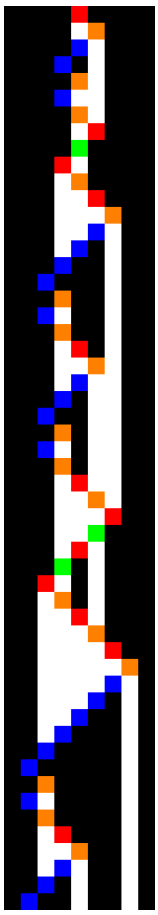
One important aspect of the technique is that, given a Turing machine and its constructed NFA—if found—it is a computationally simple task to verify that the NFA's language does indeed recognise all eventually-halting configurations of the machine.

6.2 Halt-recognizing automata

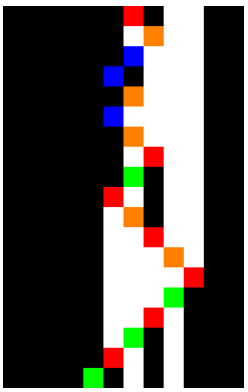
For a given Turing machine, we aim at building an NFA that recognises at least all eventually-halting configurations. That way, if it does not recognise the initial configuration then we know that the Turing machine does not halt from it. ~~as in figure 6~~ . ~~We now seek criteria sufficient to preclude false negatives, i.e. avoiding that the finite-state machine would miss some eventually-halting configurations.~~

[7] In fact 535,801 because 15 additional translated cyclers were decided, including some Skelet's machines.

[8] By finite support, we mean tapes that contain finitely many 1s, i.e. that are prefixed and suffixed by infinitely many 0s.



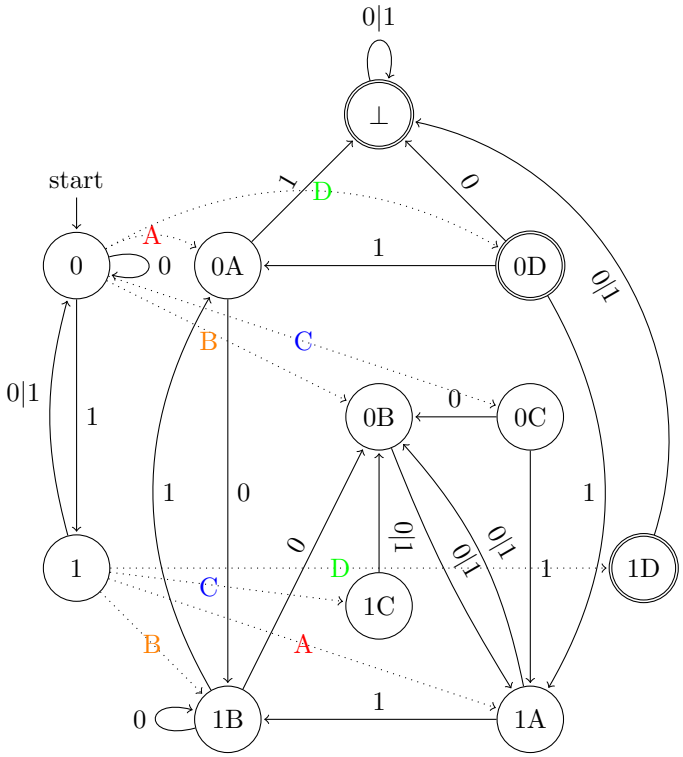
(a) Space-time diagram: the descendants of a 4-state TM's initial configuration.



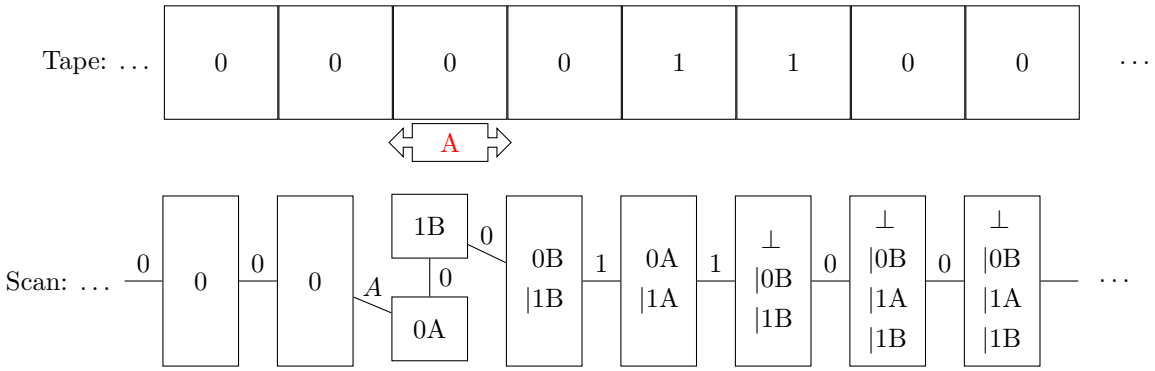
(b) The same TM halts in these configurations.

	0	1
A	1RB	0LD
B	1LC	1RA
C	0RB	0LC
D	- - -	1LA

(c) Transition table.



(d) Diagram of a Nondeterministic Finite Automaton recognizing all rows of (6b) but none of (6a).



(e) An NFA with transitions (6d), scanning the top row of (6b), stays in "accepted" state ⊥|0B|1A|1B.

Figure 6: Reducing a machine (`https://bbchallenge.org/1RB0LD_1LC1RA_0RB0LC_---1LA`) to a finite automata problem. (a) A space-time diagram of the machine, which runs forever as a "counter". Diagram rows depict *non-halting* configurations. (b) With an alternate initial configuration, the machine halts, so these rows depict *halting* configurations. (c) Transition table for the TM. (d) Transition diagram for a tape-scanning NFA. (e) To scan the top row of (6b) with NFA (6d), turn the TM configuration into a finite word (truncating the infinite 0-strings at some point and writing the head's state *before* the bit it's on). If the NFA recognizes all halting configurations (independent of 0-padding) but not the initial configuration, then the TM cannot halt.

Let's first recall how Nondeterministic Finite Automta (NFA) can be described using linear algebra. Let $\mathbf{2}$ denote the Boolean semiring $\{0,1\}$ with operations $+$ and $\cdot$ respectively implemented by OR and AND. Let $M_{m,n}$ be the set of matrices with m rows and n columns over $\mathbf{2}$. We may define a Nondeterministic Finite Automaton (**NFA**) with n states and alphabet $\mathcal{A}$ as a tuple $(q_0, \{T_\gamma\}_\gamma \in \mathcal{A}, a)$ where $q_0 \in M_{1,n}$ and $a \in M_{n,1}$ respectively represent the initial states and accepting states of the NFA. (i.e. if the i^{th} state of the NFA is an initial state then the i^{th} entry of q_0 is set to 1 and the rest are 0, and the i^{th} entry of a is set to 1 if and only if the i^{th} state of the NFA is accepting), and where transitions are matrices $T_\gamma \in M_{n,n}$ for each $\gamma \in \mathcal{A}$. A word $u = \gamma_1 \dots \gamma_\ell \in \mathcal{A}^*$ is accepted by the NFA iff there exists a path from an initial state to an accepting state that is labelled by the symbols of u, which algebraically translates to $q_0 T_u a = 1$.

JEB: Citation needed? LINK

Example 17. *TODO*

Now, we describe how we transform Turing machine configurations that finitely many 1s into finite words that will be read by our NFA. First recall that a Turing machine configuration is defined by the 3-tuple: (i) state in which the machine is (ii) position of the head (iii) content of the memory tape, see Section 1. Then, a word-representation of a configuration is defined by:

Definition 18 (Word-representations of a configuration)**.** Let c be a Turing machine configuration with finite support, i.e. there are finitely many 1s on the memory tape of the configuration. A word-representation of the configuration c is a word $\hat{c}$ constructed by concatenating (from left to right) the symbols of any finite region of the tape that contains all the 1s and, adding the head's state (a letter between A and E in the case of 5-state TMs) right before the tape symbol its currently reading.

Example 19. A word-representation of the configuration depicted in Figure 6e is $\hat{c} = \texttt{00A001100}$.

Note that two word-representations of the same configuration will only differ in the number of leading and trailing 0s that they have. Hence, if $\mathcal{L}$ is the regular language of the NFA that we wish to construct to recognise the eventually-halting configurations of a given TM, it is natural that we ask the following:

$$u \in \mathcal{L} \iff 0u \in \mathcal{L} \qquad \text{(leading zeros ignored)}$$
$$u \in \mathcal{L} \iff u0 \in \mathcal{L} \qquad \text{(trailing zeros ignored)}$$

These are implied by the following, generally stronger, conditions on the transition matrix $T_0 \in M_{n,n}$:

$$q_0 T_0 = q_0 \tag{1}$$
$$T_0 a = a \tag{2}$$

Then, we want our NFA's language $\mathcal{L}$ to include all eventually-halting configurations of a given Turing machine $\mathcal{M}$. Inductively, we want that:

$$c \vdash \perp \implies \hat{c} \in \mathcal{L}$$
$$(c_1 \vdash c_2) \wedge \hat{c}_2 \in \mathcal{L} \implies \hat{c}_1 \in \mathcal{L}$$

With c, c_1, c_2 configurations of the TM (with finite support) and $\hat{c}, \hat{c}_1, \hat{c}_2$ any of their finite word-representations as defined above. Let $f, t \in \{A, B, C, D, E\}$ denote TM states (the "from" and "to" states in a TM transition), and $r, w, b \in \{0, 1\}$ denote bits (a bit "read", a bit "written", and just a bit), then the above conditions turn into:

$$\forall u, z \in \mathbf{2}^* : ufrz \in \mathcal{L}, \text{ if } (f, r) \to \perp \text{ is a halting transition of } \mathcal{M}$$
$$\forall u, z \in \mathbf{2}^*, \forall b \in \mathbf{2} : utbwz \in \mathcal{L} \implies ubfrz \in \mathcal{L}, \text{ if } (f, r) \to (t, w, \text{left}) \text{ is a transition of } \mathcal{M}$$
$$\forall u, z \in \mathbf{2}^*, \forall b \in \mathbf{2} : uwtz \in \mathcal{L} \implies ufrz \in \mathcal{L}, \text{ if } (f, r) \to (t, w, \text{right}) \text{ is a transition of } \mathcal{M}$$

Which algebraically becomes:

$$\forall u, z \in \mathbf{2}^* : q_0 T_u T_f T_r T_z a = 1 \text{ if } (f, r) \to \perp \text{ is a halting transition of } \mathcal{M}$$
$$\forall u, z \in \mathbf{2}^*, \forall b \in \mathbf{2} : q_0 T_u T_t T_b T_w T_z a = 1 \implies q_0 T_u T_b T_f T_r T_z a = 1, \text{ if } (f, r) \to (t, w, \text{left}) \text{ is a transition of } \mathcal{M}$$
$$\forall u, z \in \mathbf{2}^*, \forall b \in \mathbf{2} : q_0 T_u T_w T_t T_z a = 1 \implies q_0 T_u T_f T_r T_z a = 1, \text{ if } (f, r) \to (t, w, \text{right}) \text{ is a transition of } \mathcal{M}$$

These conditions are unwieldy. Let's seek stronger (thus still sufficient) conditions which are simpler.

For machine transitions going left/right, simply require $T_t T_b T_w \preceq T_b T_f T_r$ and $T_w T_t \preceq T_f T_r$, respectively.

To simplify the condition for halting machine transitions: assert that for all $u \in \mathbf{2}^*$, the sub-expression $q_0 T_u$ is not zero (this means that reading 0 and 1 from the initial states must always lead to some state[9]), and ~~(for q a nonzero combination of such vectors and $fr \vdash \perp$ a halt rule)~~ that $q_0 T_u T_f T_r$ is a vector q' satisfying $\forall z \in \mathbf{2}^* : q' T_z a = 1$ (this means that the NFA will accept any word containing a halting transition independently of the bits written after the machine's head).

Going further: define an *accepted steady state* s to be a row vector such that $sa = 1$, $sT_0 \succeq s$, and $sT_1 \succeq s$. Given such s, we have $q' \succeq s \implies \forall z \in \mathbf{2}^* : q' T_z a = 1$.

TS: I have an issue with using the word state for *steady state* here, since s is a set of NFA states. This issue of using the word state to mean a set of states arises at other points of the text.

We have shown:

Theorem 20. Machine $\mathcal{M}$ doesn't halt if there are an NFA $(q_0, \{T_\gamma\}, a)$ and row vector s satisfying the below:

$$q_0 T_0 = q_0 \quad \text{(leading zeros ignored)} \tag{1}$$
$$T_0 a = a \quad \text{(trailing zeros ignored)} \tag{2}$$
$$sa = 1 \quad (s \text{ is accepted}) \tag{3}$$
$$sT_0, sT_1 \succeq s \quad (s \text{ is a steady state}) \tag{4}$$
$$\forall u \in \mathbf{2}^* : q_0 T_u \neq 0 \quad \text{(0-,1- transitions are } \textbf{total} \text{ up to the head)} \tag{5}$$
$$\forall u \in \mathbf{2}^* : q_0 T_u T_f T_r \succeq s \quad \text{if } (f, r) \to \perp \text{ is a halting transition of } \mathcal{M} \tag{6}$$
$$T_b T_f T_r \succeq T_t T_b T_w \quad \text{if } (f, r) \to (t, w, \text{left}) \text{ is a transition of } \mathcal{M} \tag{7}$$
$$T_f T_r \succeq T_w T_t \quad \text{if } (f, r) \to (t, w, \text{right}) \text{ is a transition of } \mathcal{M} \tag{8}$$
$$q_0 T_A a = 0 \quad \text{(initial configuration rejected)} \tag{9}$$

This main theorem opens up a few avenues. The general next step is to impose more structure on the NFA, so that a subset of (1)–(8) is automatically satisfied and it's easier to search for recognizers. Let's take a detour to review *deterministic finite automata* (DFA), which are simpler and satisfy totality (5).

6.3 Deterministic finite automata

Textbooks define *deterministic* finite automata (on the binary alphabet, with acceptance unspecified) as tuples (Q, δ, q_0) of: a finite set Q (states), a $q_0 \in Q$ (initial state), and $\delta : Q \times \mathbf{2} \to Q$ (transition function). Though NFAs generalize DFAs, they can be emulated by (exponentially larger) power-set DFAs.

JEB: cite

To put this definition in the linear-algebraic framework: identify $q_0 \in Q$ with $0 \in [n] := \{0, \ldots, n-1\}$; represent states q with elementary row vectors $\langle q|$; define transition matrices T_b via $\langle q| T_b = \langle \delta(q, b)|$.

As we did for transition matrices, extend δ to words: $\delta(q, \epsilon) = q$, $\delta(q, ub) = \delta(\delta(q, u), b)$.

Given a DFA on $[n]$, call its *transition table* the list $(\delta(0,0), \delta(0,1), \ldots, \delta(n-1,0), \delta(n-1,1))$.

Call $\{\delta(q_0, u) : u \in \mathbf{2}^*\}$ the set of *reachable* states.

When building a larger recognizer, we expect no benefit from considering DFAs which just relabel others or add unreachable states. So motivated, we define a canonical form for DFAs: enumerate the reachable states via breadth-first search from q_0, producing $f : [n] =: Q_{\mathsf{cf}} \to Q$. Explicitly, $f(0) = q_0$ and $f(k)$ is the first of $\delta(f(0), 0), \delta(f(0), 1), \ldots, \delta(f(k-1), 0), \delta(f(k-1), 1)$ not in $f([k])$, valid until $f([k])$ is closed under transitions. This induces $\delta_{\mathsf{cf}}(q, b) \mapsto f^{-1}(f(q), b)$. (Warning: this definition is not standard.)

Lemma 21. *In a DFA with $(Q, q_0) = ([n], 0)$, the following are equivalent:*

1. *it's in canonical form ($Q_{\mathsf{cf}} \to Q$ is the identity) and ignores leading zeros (equation (1) or $\delta(0,0) = 0$);*
2. *its transition table includes each of $0, \ldots, n-1$, whose first appearances occur in order, and with each $q \in (0, n)$ appearing before the $\delta(q, 0)$ position;*
3. *the sequence $\{m_k := \max\{\delta(q, b) : 2q + b \leq k\}\}_{k=0}^{2n-1}$ of cumulative maxima runs from 0 to $n-1$ in steps of 0 or 1, with $m_{2q-1} \geq q$ for $0 < q < n$.*

Proof. 1 $\iff$ 2**:** We prove a partial version by induction: the DFA ignores leading zeros and $f(q) = q$ for $q \leq k$, iff $0, \ldots, k$ have ordered first appearances in the transition table which precede appearances of any $q > k$ and occur before the $\delta(k, 0)$ position if $k > 0$. In case $k = 0$, the DFA ignores leading

[9] This assertion is not limiting since if it is not met by some NFA, adding an additional sink state to the NFA will not change the recognised language and will satisfy the assertion.

zeros iff 0 comes first in the transition table by definition. (The other conditions are vacuous.) In case the claim holds for preceding k, $f(k)$ is by definition the first number outside of $f([k]) = [k]$ in the transition table—if any—and the inductive step follows.

2 $\iff$ 3**:** If the first appearances of $0, \ldots, n-1$ appear in order, any value at its first index is the largest so far, so m_k takes the same values. The sequence m_k is obviously nondecreasing, so to be gap-free it can only grow in steps of 0 or 1. Conversely, if m_k runs from 0 to $n-1$ in steps of 0 or 1, each value $q \in [n]$ must appear in the table at the first index k for which $m_k = q$, and all preceding values in the transition table must be strictly less.

In case these equivalent conditions are true, that last observation shows that q appears before the $\delta(q, 0)$ position iff m_k reaches q by index $k = 2q - 1$, or equivalently $m_{2q-1} \geq q$. □

Corollary 22. *$\{t_k\}_{k=0}^{\ell}$ ($\ell < 2n$) is a prefix of a canonical, leading-zero-ignoring, n-state DFA transition table iff $m_k := \max\{t_j\}_{j=0}^{k}$ runs from 0 to $m_{\ell-1} < n$ in steps of 0 or 1, and $m_{2q-1} \geq q$ where defined.*

Proof. If $\ell = 2n - 1$, $\{m_k\}$ grows to exactly $n - 1$ (since $m_{2(n-1)-1} \geq n - 1$), and lemma 21 applies. Otherwise, we may extend the sequence with $t_{\ell+1} = \min(m_\ell + 1, n - 1)$, the same conditions apply. □

So, Algorithm 5 searches such DFAs incrementally (avoiding partial DFAs already deemed unworkable).

Algorithm 5 SEARCH-DFA

```
1: enum CheckResult {MORE, SKIP, STOP}

2: procedure bool SEARCH-DFA(int n, function<List<int>, CheckResult> check)
Require: check(t) ≠ MORE if t is a complete (length-2n) table
3:     int k = 1, t[2 * n] = [0, ..., 0], m[2 * n] = [0, ..., 0]
4:     CheckResult state = check([0])
5:     loop
6:         if state == MORE then
7:             int q_new = m[k-1] + 1
8:             t[k] = (q_new < n and 2*q_new-1 == k) ? q_new : 0
9:         else if state == SKIP then
10:            repeat
11:                if k ≤ 1 then return false
12:                k -= 1
13:            until t[k] ≤ m[k-1] and t[k] < n-1
14:            t[k] += 1
15:        else return true
16:        m[k] = max(m[k-1], t[k])
17:        k += 1
18:        state = check(length-k prefix of t)
```

6.4 Search algorithm: direct

Let's pick up the thread from §6.2, and add more structure to the NFA $(q_0, \{T_\gamma\}, a)$. First, let's have separate states for scanning each side of the tape: assume state space $\mathbf{2}^l \oplus \mathbf{2}^d$, initial state $\begin{bmatrix} q_0 & 0 \end{bmatrix}$, transitions $T_b = \begin{bmatrix} L_b & 0 \\ 0 & R_b \end{bmatrix}$ ($b \in \{0, 1\}$) and $T_f = \begin{bmatrix} 0 & M_f \\ 0 & 0 \end{bmatrix}$ ($f \in \{A, \ldots, E\}$), and acceptance $\begin{bmatrix} 0 \\ a \end{bmatrix}$.

Next, let's designate a "halt" state $\perp$, and corresponding basis vector $\langle\perp| \in \mathbf{2}^d$. Let $s = 0 \oplus \langle\perp|$.

Also designate independent basis vectors for each $\langle i|M_f$ ($i \in [l]$, $f \in \{A, \ldots, E\}$). Call them $\langle i, f|$.

Finally, let's have $(q_0, \{L_b\})$ come from a DFA that ignores leading zeros. That ensures (1) and (5).

We seek a "direct" algorithm which searches DFAs and solves (1)–(8) for a, R_0, R_1 if possible.

A helpful property of standard basis vectors is: the condition $\langle i|m \succeq v$ is equivalent to $m \succeq |i\rangle v$.

JEB: So many assumptions! Without loss of generality, in fact. Would a pause to prove that distract or clarify?

This lets us re-express conditions (4) and (6)–(8):

$$R_r \succeq |\bot\rangle\langle\bot| \qquad \text{for } r \in \{0,1\} \tag{4'}$$
$$\forall i \in [l] : R_r \succeq |i, f\rangle\langle\bot| \qquad \text{if } \ldots fr \ldots \vdash_{\mathcal{M}} \bot \text{ is a halt rule} \tag{6'}$$
$$\forall i \in [l] : R_r \succeq |\delta(i,b), f\rangle\langle i, t|R_b R_w \qquad \text{if } \ldots bfr \ldots \vdash_{\mathcal{M}} \ldots tbw \ldots \text{ is a left rule} \tag{7'}$$
$$\forall i \in [l] : R_r \succeq |i, f\rangle\langle\delta(i,w), t| \qquad \text{if } \ldots fr \ldots \vdash_{\mathcal{M}} \ldots wt \ldots \text{ is a right rule} \tag{8'}$$

There's a unique minimal (w.r.t $\preceq$) solution to this system of inequalities and a clear way to compute it: initialize R_0, R_1 to zero, then set entries to 1 as these 4 inequalities demand until a fixed point is reached. Note that the R_0 and R_1 values so produced are an increasing function of the matrices $\{L_b\}$, or of δ when it's viewed as a partial function (ordered by extension).

The remaining conditions of Theorem 20 are (2)–(3), the conditions on acceptance. (9) is the "negative" condition $\langle 0, A| \cdot a = 0$. The others translate to $a \succeq |\bot\rangle$ and $a = R_0 a$. From (4'), we see the sequence $|\bot\rangle \preceq R_0|\bot\rangle \preceq R_0^2|\bot\rangle \preceq \ldots$ is increasing, this pair of conditions also has a minimal fixed-point solution. The minimal solution satisfies (9) if anything does. Since the fixed point of $R_0^k|\bot\rangle$ increases with R, the fixed-point solution for R also suffices if anything does. Since that increases under partial-DFA extensions, we can check partial DFAs for violations of (9). And so we obtain a complete decider in Algorithm 6. A slight elaboration allows for checking a specific L instead of all DFAs, and for returning the recognizer.

Algorithm 6 DECIDER-FINITE-AUTOMATA-REDUCTION-DIRECT

```
1:  procedure bool DECIDER-FINITE-AUTOMATA-DIRECT(TM machine, int n, bool left_to_right)
2:      if not left_to_right then replace TM with its mirror-image
3:      Matrix<bool, 5*n+1, 5*n+1> R[2*n+1][2] = [[0,0], ..., [0,0]]
4:      ColVector<bool, 5*n+1> a[2*n+1] = 0
5:      ▷ Note about indexing: for ⟨i,f| use index 5*i+f, and for ⟨⊥|, use index 5*n.
6:      Initialize R[0] using (4') and (6')
7:      Initialize a[0] = |⊥⟩
8:      procedure CheckResult CHECK(List<int> L)
9:          R[L.length()], a[L.length()] = R[L.length()-1], a[L.length()-1]
10:         Increase R[L.length()] using (8'), with (i,w) = divmod(L.length()-1, 2)
11:         repeat
12:             Increase R[L.length()] using (7'), restricted to 2*i+b < L.length()
13:         until R[L.length()] stops changing
14:         repeat
15:             a[L.length()] = R[L.length()][0] * a[L.length()]
16:         until a[L.length()] stops changing
17:         if ⟨0,A|·a ≠ 0 then return SKIP
18:         else if L.length() == 2*n then return STOP
19:         else return MORE
20:     return SEARCH-DFA(check)
```

6.5 Search algorithm: meet-in-the-middle DFA

A symmetric recognizer construction has also shown good results. Again, pass the left half-tape through a DFA with l states. Imagine a DFA with d states scanning the (strict) right half-tape right-to-left.

Remark 23. *Our definitions require a left-to-right scan direction. Any NFA $(\langle 0|, \{R_b\})$ can be transposed. (Transposing transition matrices reverses the arrows in the diagram, as with graph adjacency matrices.) We can shoehorn this into the preceding framework by making an accept state from R's transposed initial state $|0\rangle$, defining middle transitions M_{fr} for the configuration's head state/bit, superposing all states of R to get our s vector, and trying to satisfy conditions like $\langle 0|M_{A0}|0\rangle = 0$, $M_{fr} = \sum_L |q\rangle\langle s|$ (for halt rules), $L_b M_{fr} \succeq M_{tb} R_w^T$ (for left rules), $M_{fr} R_b^T \succeq L_w M_{tb}$ (for right rules). What follows is more intuitive.*

JEB: Is this remark needed?

As in Figure 7, let's consider the DFAs on their own terms. Each one partitions its input into a family of regular languages (one per state). Accounting for the head state/bit and right half-tape, we obtain $l \cdot 5 \cdot 2 \cdot d$ classes of TM configuration. Propose a recognizer which distils this classification into a result. We'll work out conditions for a good "accepted" set $A \subseteq [l] \times \{A, \ldots, E\} \times \mathbf{2} \times [d]$. If they're satisfiable, even if we don't prove the scheme sound, we can feed the left DFA into Algorithm 6 to check the result.

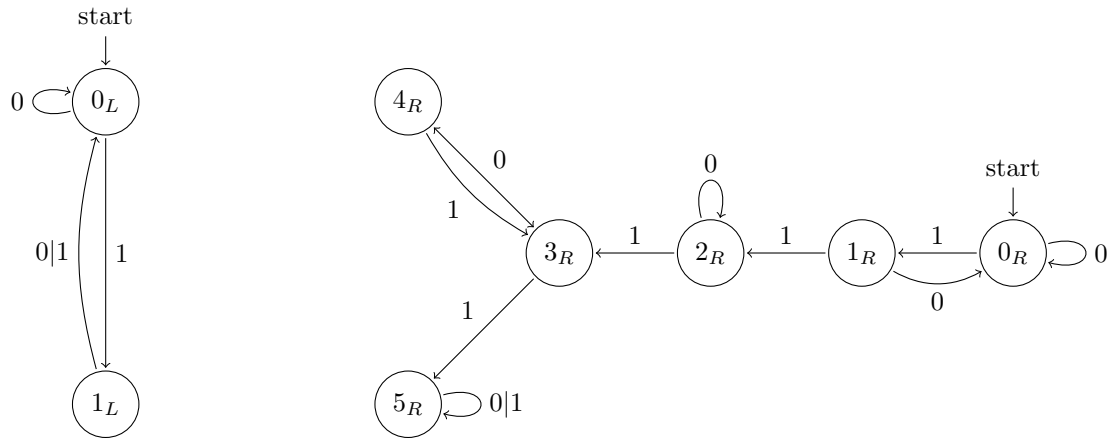


Figure 7: This pair of DFAs can also recognize halting configurations for the TM of figure 6. Configurations are classified by their head state, head bit, and the two half-tapes (processed outside-in by the DFAs.)

Despite the new setting, we can write out closure conditions analogous to §6.2's, each $\forall i \in [l], j \in [d]$:

$$(i, f, r, j) \in A \qquad \text{if } \ldots fr \ldots \vdash_{\mathcal{M}} \perp \text{ is a halt rule} \qquad (6")$$

$$(i, t, b, \delta_R(j, w)) \in A \implies (\delta_L(i, b), f, r, j) \in A \qquad \text{if } \ldots bfr \ldots \vdash_{\mathcal{M}} \ldots tbw \ldots \text{ is a left rule} \qquad (7")$$

$$(\delta_L(i, w), t, b, j) \in A \implies (i, f, r, \delta_R(j, b)) \in A \qquad \text{if } \ldots frb \ldots \vdash_{\mathcal{M}} \ldots wtb \ldots \text{ is a right rule} \qquad (8")$$

The goal, analogous to (9), is $(0, A, 0, 0) \notin A$.

We could now search all DFA pairs, checking if the smallest A closed under (6")–(8") rejects $(0, A, 0, 0)$. However, to get decent performance, we must express the above as a boolean satisfiability (SAT) problem.

Other lessons learned in practice: it was most effective to use the same state count on both sides ($l = d = n$), and it was decisively faster to impose the canonical form restrictions of Lemma 21.

Algorithm 7 shows how this works. Here especially, actual code can vary from the given pseudocode:

- If SAT solvers use integers for literals (variables and their negations), one needn't "allocate variables".
- It may be possible to simplify by adding propositional variables for more edge cases.
- The "outcomes are mutually exclusive" condition may be represented differently.
- Checking a solution is valid needn't involve Algorithm 6, if the author proves more.

6.6 Correctness

Our implementation takes a brutally simple approach to correctness: independent of the proof-search algorithm, it checks the recognizer it returns satisfies (1)–(8). By theorem 20, this prevents false proofs.

6.7 Results

JEB: TBD. If accepted/merged, update repo link, state the parameters, summarize the progress.

The decider was coded in `Rust` and is accessible at this link: `https://github.com/UncombedCoconut/bbchallenge-deciders/tree/finite-automata-reduction/decider-finite-automata-reduction`.

It uses Kissat[**?**] as its SAT solver.

More information about these results are available at: `https://discuss.bbchallenge.org/t/decider-finite-automata-reduction/123`.

6.8 Related research

The "direct" algorithm of §6.4 was first found by applying [**?**]. Rough sketch: treat a TM as a two-stack machine (one per side of the tape). Fix a DFA classifying left half-tape configurations. Form a "quotient" of $\mathcal{M}$ in which stack values are replaced by corresponding DFA states. The many-to-one mapping of configurations induces a mapping of transitions, leaving us with a nondeterministic one-stack machine. The algorithm of [**?**] solves reachability for its halt-configurations. This perspective was dropped in favour of a more direct approach, but it can be useful when probing the limits of this technique.

Algorithm 7 DECIDER-FINITE-AUTOMATA-REDUCTION-MITM-DFA

procedure bool DECIDER-FINITE-MITM-DFA(**TM** machine, **int** n)
  ▷ Allocate variables.
  Map<tuple, int> tk_eq, tk_le, mk_eq, A
  for all $(\mathrm{lr}, \mathrm{k}, \mathrm{y}) \in [2] \times [n] \times [2 * n] \times [n + 1]$ **do**
    if (k, y) == (0, 0) **then** tk_eq[lr, k, y] = true
    else if $0 \leq y \leq \min(k, n - 1)$ **then** tk_eq[lr, k, y] = NEW-VARIABLE
    else tk_eq[lr, k, y] = false
    if $y \leq 0$ **then** tk_le[lr, k, y] = tk_eq[lr, k, y]
    else if $0 \leq y \leq \min(k - 1, n - 2)$ **then** tk_le[lr, k, y] = NEW-VARIABLE
    else tk_le[lr, k, y] = true
    if (k, y) == (2*n-1, n-1) **then** mk_eq[lr, k, y] = true
    else if not $\lceil \frac{k}{2} \rceil \leq y < \min(n, k + 1)$ **then** mk_eq[lr, k, y] = false
    else if $\min(n, k + 1) - ((k + 1)/2) \leq 1$ **then** mk_eq[lr, k, y] = true
    else mk_eq[lr, k, y] = NEW-VARIABLE
  for all $(\mathrm{i}, \mathrm{f}, \mathrm{r}, \mathrm{j}) \in [n] \times [5] \times [2] \times [n]$ **do**
    if (k, y) == (2*n-1, n-1) **then** A[i, f, r, j] = false
    else A[i, f, r, j] = NEW-VARIABLE
  ▷ Transition validity: outcomes are mutually exclusive.
  for all $(\mathrm{lr}, k, \mathrm{y}) \in [2] \times [2 * n] \times [n]$ **do**
    NEW-CLAUSE$(\mathrm{tk_eq}(\mathrm{lr}, k, y) \implies \mathrm{tk_le}(\mathrm{lr}, k, y))$
    NEW-CLAUSE$(\mathrm{tk_le}(\mathrm{lr}, k, y) \implies \mathrm{tk_le}(\mathrm{lr}, k, y + 1))$
    NEW-CLAUSE$(\mathrm{tk_eq}(\mathrm{lr}, k, y + 1) \implies \neg \mathrm{tk_le}(\mathrm{lr}, k, y))$
  ▷ Transition validity: an outcome occurs.
  for all $(\mathrm{lr}, k) \in [2] \times \{1, \ldots, 2 * n - 1\}$ **do** NEW-CLAUSE$(\bigvee_{y=0}^{\min(k, n-1)} \mathrm{tk_eq}(\mathrm{lr}, k, y))$
  ▷ Closure conditions.
  for all $(i, j, (f, r)) \in [n]^2 \times$HALT-RULES(machine) **do**
    NEW-CLAUSE$(A[i, f, r, j])$
  for all $(i, j, \mathrm{ib}, \mathrm{jw}, (f, r, w, L, t)) \in [n]^4 \times$LEFT-RULES(machine) **do**
    NEW-CLAUSE$(\mathrm{tk_eq}[L, i, b, \mathrm{ib}] \wedge \mathrm{tk_eq}[R, j, w, \mathrm{jw}] \wedge A[i, t, b, \mathrm{jw}] \implies A[\mathrm{ib}, f, r, j])$
  for all $(i, j, \mathrm{iw}, \mathrm{jb}, (f, r, w, R, t)) \in [n]^4 \times$RIGHT-RULES(machine) **do**
    NEW-CLAUSE$(\mathrm{tk_eq}[R, j, b, \mathrm{jb}] \wedge \mathrm{tk_eq}[L, i, w, \mathrm{iw}] \wedge A[\mathrm{iw}, t, b, j] \implies A[i, f, r, \mathrm{jb}])$
  ▷ DFA is in canonical form (Lemma 21).
  for all $(\mathrm{lr}, k) \in [2] \times \{1, \ldots, 2 * n - 1\}$ **do**
    for $m = \lfloor k/2 \rfloor, \ldots, \min(n, k)$ **do**
      NEW-CLAUSE$(\mathrm{mk_eq}(\mathrm{lr}, k - 1, m) \implies \mathrm{tk_le}(\mathrm{lr}, k, m + 1))$
      NEW-CLAUSE$(\mathrm{mk_eq}(\mathrm{lr}, k - 1, m) \wedge \mathrm{tk_le}(\mathrm{lr}, k, m) \implies \mathrm{mk_eq}(\mathrm{lr}, k, m))$
      NEW-CLAUSE$(\mathrm{mk_eq}(\mathrm{lr}, k - 1, m) \wedge \mathrm{tk_eq}(\mathrm{lr}, k, m + 1) \implies \mathrm{mk_eq}(\mathrm{lr}, k, m + 1))$
  if CHECK-SAT **then**
    Assert L DFA from the model proves the machine using Algorithm 6.
    return true
  else return false