Simulation on One-Counter Machines

Petr Jančar   Faron Moller   Zdeněk Sawa

January 15, 2002

Abstract

We discuss the decidability questions for simulation preorder (and equivalence) for processes generated by one-counter machines. We sketch a proof of decidability in the case when testing for zero is not possible and show the undecidability in the general case.

1 Introduction

By a ***one-counter machine*** M we mean a nondeterministic finite automaton, we refer to its states as *control states*, denoted $p, q, \ldots$, acting on a counter ranging over the set of nonnegative integers $\mathbb{N}$. The term *states* we reserve for configuration, denoted $p(m), \ldots$. There are (finitely many) transition rules; each of them may have a boolean guards of the form 'if the current state is p and counter is positive' or 'if the current state is p and counter is zero' and enables to perform action a out of a finite alphabet Σ, changing the control state to some q and either increment, decrement, or ignore the counter value. (A decrementing transition is only possible if the counter is positive.) We call M 'weak' iff there is no guard of the second form (zero test). Note that, when $m>0$, the initial "moves" of $p(m)$ do not depend on the actual value of m.

A binary relation $\mathcal{S}$ between the states of two such nets is a ***simulation*** if and only if, whenever $\langle p(m), q(n)\rangle \in \mathcal{S}$ and $p(m) \xrightarrow{a} p'(m')$, we have $q(n) \xrightarrow{a} q'(n')$ with $\langle p'(m'), q'(n')\rangle \in \mathcal{S}$. $p(m)$ is ***simulated*** by $q(n)$, written $p(m) \preceq q(n)$, if and only if they are related by some simulation relation $\mathcal{S}$. If two states are related by a symmetric simulation relation, then they are ***bisimilar***.

Such automata are enjoying renewed interest within the automata and process theory communities due to the present active search for the dividing line between decidable and undecidable theories for classes of infinite state systems (see, e.g., [4, 2]).

Recently, Abdulla and Čerāns [1] outlined an extensive and involved proof of the decidability of simulation preorder over one-counter nets. Their 16-page extended abstract is very technical and refers to an unpublished full paper for the proofs of most of the crucial lemmas; it thus is also hard to verify. Here we sketch a short proof, using a 'geometrical' (more 'visible', lucid ?) approach; although short, due to space limitations and other results communicated here, we just sketch it and refer to the full version [**?**].[1]

We then show that simulation preorder between one-counter machines with tests for zero is undecidable, even a very restricted subcase of deterministic machines; it is easily transformed to show undecidability of simulation equivalence in the (very easy subcase of) non-deterministic case, which contrasts with the decidability of bisimulation equivalence [3].

2 Decidability

For any pair of control states $\langle p,q\rangle \in Q_1\times Q_2$ taken from two weak one-counter machines, we can ask for what values $m,n\in\mathbb{N}$ do we have $p(m)\preceq q(n)$. We can picture the "graphs" of the functions $\mathbb{G}_{\langle p,q\rangle}:\mathbb{N}\times\mathbb{N}\to\{\text{black, white}\}$ given by

$$\mathbb{G}_{\langle p,q\rangle}(m,n)=\begin{cases}\text{black}, & \text{if} \quad p(m)\preceq q(n);\\ \text{white}, & \text{if} \quad p(m)\not\preceq q(n)\end{cases}$$

by appropriately colouring (black or white) the integral points in the first quadrant of the plane. Note that if $p(m)\preceq q(n)$ then $p(m')\preceq q(n')$ for all $m'{\leq}m$ and $n'{\geq}n$; this follows immediately from the observation that

$$\Big\{\langle p(m'),q(n')\rangle : p(m)\preceq q(n),\ m'{\leq}m,\ n'{\geq}n\Big\}$$

is a simulation relation. Hence the black points are upwards- and leftwards-closed, and the white points are downwards- and rightwards-closed. For a fixed pair of states $p_0(m_0)$ and $q_0(n_0)$ of these nets, we shall decide the question *"Is $p_0(m_0)\preceq q_0(n_0)$?"* by effectively constructing an initial portion of these $k{=}|Q_1|\times|Q_2|$ graphs which includes the point $\langle m_0,n_0\rangle$, and then look to the colour of $\mathbb{G}_{\langle p_0,q_0\rangle}(m_0,n_0)$.

Define the ***frontier function*** $f_{\langle p,q\rangle}(n)=\max\{m : \mathbb{G}_{\langle p,q\rangle}(m,n)=\text{black}\}$, that is, the greatest value m such that $p(m){\preceq}q(n)$; $f_{\langle p,q\rangle}(n){=}\infty$ if $\mathbb{G}_{\langle p,q\rangle}(m,n)=\text{black}$, that is $p(m)\preceq q(n)$, for all m; and $f_{\langle p,q\rangle}(n){=}-1$ if $\mathbb{G}_{\langle p,q\rangle}(0,n)=\text{white}$, that is $p(0)\not\preceq q(n)$. This function is monotone nondecreasing, and the set of ***frontier points*** $\langle f_{\langle p,q\rangle}(n),n\rangle\in\mathbb{N}\times\mathbb{N}$ defines the ***frontier*** of $\mathbb{G}_{\langle p,q\rangle}$, the collection of the right-most black points from each level. Note that the order of the pairs $\langle f_{\langle p,q\rangle}(n),n\rangle$ in the frontier sets is reverse to the usual way of representing functions as sets; this is an unavoidable confusion. (If we follow the usual convention, then other confusions inevitably arise.) Also, note that we shall abuse notation by using f to refer not just to the frontier function but also to the frontier given by the frontier function.

The next theorem is the clue to our decidability result.[2]

[1] To this submission, we add an appendix for convenience.

[2] see Appendix for the proof

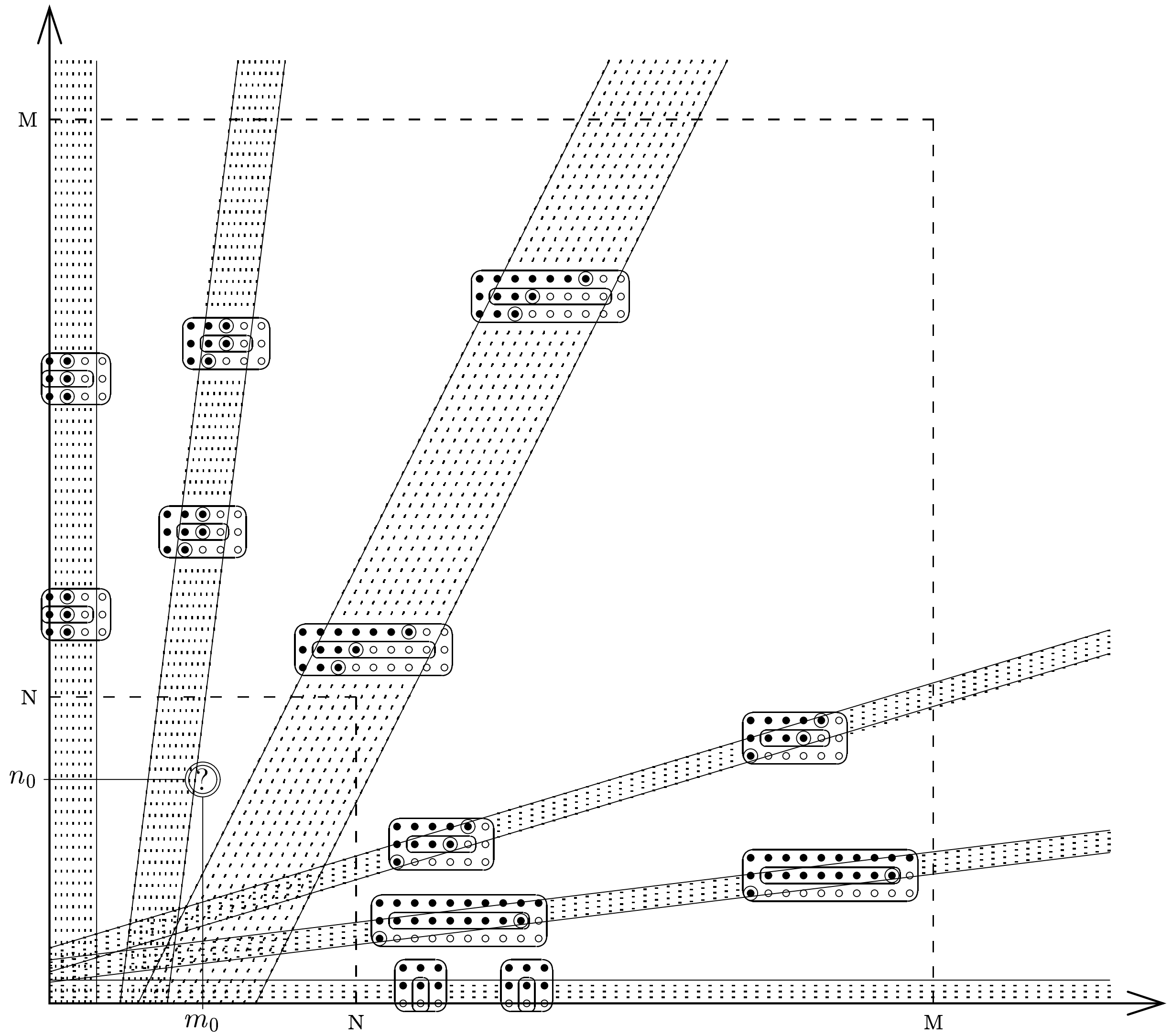


Figure 1: Segments of six graphs $\mathbb{H}_{\langle p,q\rangle}$ superimposed onto each other.

The Belt Theorem *Every frontier lies within a linear belt with rational (or infinite) slope.*

Now we describe the decision procedure which is based on this theorem. Consider constructing the graphs of some functions $\mathbb{H}_{\langle p,q\rangle} : \mathbb{N}\times\mathbb{N} \to \{\text{black, white}\}$ as follows. Start by assuming that all points are black, and by considering ever larger initial squares, make $\mathbb{H}_{\langle p,q\rangle}(m,n)$ white if there is a transition $p(m) \stackrel{a}{\to} p'(m')$ such that for every transition $q(n) \stackrel{a}{\to} q'(n')$ we have previously recoloured $\mathbb{H}_{\langle p',q'\rangle}(m',n')$ white. By induction, if we ever recolour a point $\mathbb{H}_{\langle p,q\rangle}(m,n)$ white, then $p(m) \not\preceq q(n)$. If we carry out this procedure indefinitely, we would in fact construct the graphs $\mathbb{G}_{\langle p,q\rangle}$, since the set of pairs $\langle p(m), q(n)\rangle$ such that $\mathbb{H}_{\langle p,q\rangle}(m,n)$ remains black is readily seen to be a simulation. (No pair corresponding to a black point could contradict the definition of a simulation, as this is the criterion for recolouring the point white.) Thus, every point which should be white (according to the graphs $\mathbb{G}$) would indeed be recoloured white at some point in this construction.

Because of this, by the Belt Theorem we must eventually be able to lay down a set of linear belts with rational slopes such that (see Figure 1):

- within the currently recoloured square, there is an initial ($\mathrm{M}\times\mathrm{M}$) square inside of which each frontier lies within some belt (we may assume that parallel belts coincide, so that two or more frontiers may appear in the same belt);
- outside of some initial ($\mathrm{N}\times\mathrm{N}$) square ($\mathrm{N}<\mathrm{M}$) containing the point $\langle m_0, n_0\rangle$, the belts are separated by gaps wide enough so that no point has neighbouring points in two belts;
- within the area bounded by the initial ($\mathrm{N}\times\mathrm{N}$) and ($\mathrm{M}\times\mathrm{M}$) squares, looking at each horizontal level within each belt (or each vertical level, in the case of a horizontal belt) we find a pattern which repeat itself—along with all of its neighbouring points—at two different levels. (That is, the colourings of the points and neighbouring points are the same in every graph on these levels within the belt.) Furthermore, the shift from one occurrence of the pattern to the next has a slope equal to that of the belt.

Note that these belts need not *a priori* be the true frontier belts specified in the Belt Theorem; but since (by the pigeonhole principle) the true frontier belts display such a repetitive pattern, the true frontier belts must eventually appear in the above fashion if no other belts appear earlier on in the construction.

Once we recognise such belts, the complete graphs $\mathbb{H}_{\langle p,q\rangle}$ are determined by continuing the colouring of the graphs by periodically repeating the colouring within the belts between the levels at which the patterns repeat, and recolouring points to the right of the belts to maintain the invariant that white points are rightwards-closed.

We can then readily confirm that the set of all pairs $\langle p(m), q(n)\rangle$ such that $\mathbb{H}_{\langle p,q\rangle}(m,n)$ is black is a simulation. (The validity of the simulation condition for a black point is dependent only on its neighbouring points; and to each black point anywhere on the graphs there is a corresponding black point—perhaps the point itself—which has identically-coloured neighbouring points and which has explicitly been shown to locally satisfy the simulation condition in this sense.) Thus, all black points are correct (that is, $\mathbb{G}_{\langle p,q\rangle}(m,n)$ is black whenever $\mathbb{H}_{\langle p,q\rangle}(m,n)$ is black), and all white points within the initial ($\mathrm{N}\times\mathrm{N}$) square are correct, proving that we have correctly constructed the initial ($\mathrm{N}\times\mathrm{N}$) square.

3 Undecidability

A *Minsky machine* C with two nonnegative counters c_1, c_2 is a program

$$1 : COMM_1; 2 : COMM_2; \ldots ; n : COMM_n$$

where $COMM_n$ is a *halt*-command and $COMM_i$ $(i = 1, 2, \ldots, n-1)$ are commands of the following two types (assuming $1 \leq k, k_1, k_2 \leq n, 1 \leq j \leq m$)

(1) $c_j := c_j + 1;\ goto\ k$

(2) $if\ c_j = 0\ then\ goto\ k_1\ else\ (c_j := c_j - 1;\ goto\ k_2)$

The computation of the machine C is deterministic.

Lemma 1 *If $p(m)$ and $q(n)$ are states of one-counter machines with tests for zero M_1 and M_2, then the problem if $p(m) \preceq q(n)$ is undecidable.*

Proof: It is useful to think about simulation in terms of a game. There are two players—Player 1 and Player 2. Each player plays with one one-counter machine (Player 1 with M_1, Player 2 with M_2). Both players take turns selecting transitions of their machines. Player 1 begins with selection of some transition with some label, Player 2 has to respond with selection of some transition with the same label. If Player 2 has no such transition, he loses. If he can always respond, he wins. Player 1 has a winning strategy iff $p(0) \not\preceq q(0)$, Player 2 has defending strategy otherwise.

We can show the undecidability by reduction of the halting problem for a Minsky machine C with 2 counters (which is known to be undecidable) to the problem if $p(m) \preceq q(n)$. We shall construct two one-counter machines M_1 and M_2 corresponding to C, such that $p(0) \not\preceq q(0)$ iff the computation C halts ($p(0)$ and $q(0)$ are start states of M_1 and M_2). There is even possible to construct M_1 and M_2 such that M_1 or M_2 has fixed structure independent on C.

The main idea is same in both cases. Labels of actions of both machines correspond to different kinds of actions of C (increment the value of the counter, decrement it, if it is positive, or ignore it, if it is zero). There are different labels for actions on different counters of C and there is also a special label for the *halt*-command. Both machines simulate the computation of C (M_1 actions on the counter c_1, and M_2 actions on the counter c_2). Player 1 can choose, what action shall be performed next (but he can always choose only actions which are possible in C wrt the value of c_1), but if he chooses an action violating the simulation of C (wrt the value of c_2), Player 2 performs the transition to a special control state, where transitions with any label are possible and M_2 stays there forever. The only way, how Player 1 can win, is to correctly simulate actions of C and to reach the state corresponding to the *halt*-command of C, where no transitions are possible in M_2, but some transition is possible in M_1, and then perform a transition, which Player 2 can not respond to.

Let us first discuss the case where M_2 has fixed structure. M_2 has only two control states one corresponding to the correct simulation of C and one used when Player 1 violates simulation. The control states of M_1 correspond to the commands of C. There is always only one transition possible in all such states, with the only exception of a state corresponding to the decrementation of c_2, where next action depends on the value of c_2 (if it is zero or not). This is the only place, where Player 1 can violate the simulation. In the state corresponding to the halt instruction Player 1 can perform the *halt*-action, which Player 2 can not respond to.

Now let us discuss the case where M_1 has fixed structure. M_1 has only one state and control states of M_2 correspond to the commands of C. Player 1 can choose whatever transition he wants (but can not decrement c_1, if it is zero), but M_2 has information what action is actually possible, so incorrect moves of Player 1 are punished by the same way as described above. When Player 1 chooses the *halt*-action, Player 2 can react (go to the special state, where transitions with all labels are possible) in all states with the only exception of the state corresponding to the *halt*-instruction, where no transitions are possible.

Lemma 2 *If $p(m)$ and $q(n)$ are states of non-deterministic one-counter machines with tests for zero M_1 and M_2, then the problem if $p(m) \equiv q(n)$ is undecidable.*

Proof: We can reduce the problem if one machine is simulated by another (which was shown to be undecidable) to the problem if there is a simulation equivalence between two non-deterministic machines.

Let us have two machines M_1 and M_2. We can restrict to a special case where both M_1 and M_2 have only one possible transition in their start states and both these transitions have the same label. There should not exist any transition to these start states. We can always convert any two machines to this form by adding new start states and transitions.

Now we can construct new non-deterministic machine N, such that M_1 is simulated by M_2 iff there is simulation equivalence of M_2 and N. (We mean the relations of their start states here of course.)

N is constructed as the union of M_1 and M_2 (the union of their control states and transitions with an exception of their start states) with the new common start state and corresponding transitions added. N can non-deterministically choose at the start state if it will act as M_1 or M_2 after the first transition.

It is obvious that M_2 is simulated by N, because M_2 is part part of N's structure. N is simulated by M_2 iff M_1 is simulated by M_2, because if N chooses that it will act as M_2, then M_2 can performs exactly the same transitions as N, but if N chooses that it will act as M_1, then M_2 has to be able to simulate M_1.

References

[1] Abdulla, P., K. Čerāns (1998). Simulation is decidable for one-counter nets (Extended Abstract.) In Proceedings of CONCUR'98, *Lecture Notes in Computer Science* **1466**:253–268.

[2] Burkart, O., J. Esparza (1997). More infinite results. Proceedings of Infinity'97. *Electronic Notes in Theoretical Computer Science* **5**.

[3] Jančar, P. (1997). Decidability of bisimilarity for one-counter processes. In Proceedings of ICALP'97, *Lecture Notes in Computer Science* **1256**:549–559. (Revised version to appear in *The Journal of Information and Computation.*)

[4] Moller, F. (1996). Infinite results. Proceedings of CONCUR'96, *Lecture Notes in Computer Science* **1119**:195–216.

[5] P. Jančar, Undecidability of bisimilarity for Petri nets and some related problems, *Theoretical Computer Science* 148 (1995) 281–301

[6] P. Jančar, Nonprimitive recursive complexity and undecidability for Petri net equivalences (1998)

[7] M. Minsky, *Computation: Finite and Infinite Machines* (Prentice-Hall, Englewood Cliffs, NJ, 1967)

Appendix

Proof of the Belt Theorem

By an ***area*** we mean a set $A \subseteq \mathbb{N}\times\mathbb{N}$. We define its ***interior*** and ***border*** as follows.

$$\begin{aligned}\textbf{interior}(A) &= \Big\{ \langle m,n\rangle \;:\; \{m{-}1,m,m{+}1\}\times\{n{-}1,n,n{+}1\} \;\subseteq\; A \Big\};\\ \textbf{border}(A) &= A - \textbf{interior}(A).\end{aligned}$$

Given an area A and a vector $v \in \mathbb{Z}\times\mathbb{Z}$ (where $\mathbb{Z}$ denotes the set of integers), we shall let $\textbf{shift}(A,v) = (A{+}v)\cap(\mathbb{N}\times\mathbb{N})$ denote the area A shifted by vector v. We say that the shift of an area A by a vector v is ***safe wrt*** $B \subseteq \textbf{shift}(A,v)$ iff for all graphs $\mathbb{G}_{\langle p,q\rangle}$ and all $u{\in}B$ we have that $\mathbb{G}_{\langle p,q\rangle}(u)$ is black whenever $\mathbb{G}_{\langle p,q\rangle}(u{-}v)$ is black. We say that such a shift is ***safe*** iff it is safe wrt $\textbf{shift}(A,v)$, that is, if it never shifts a black point to a white point. We shall use the following fact, as well as its corollary which gives a sufficient condition for when a shift is safe.

Fact 3 *Let A be an area and V a set of vectors which satisfy the following:*

> *if we have some $u \in A$ and $v \in V$ with $u{+}v \in \textbf{border}(\textbf{shift}(A,v))$*
> *such that $\mathbb{G}_{\langle p,q\rangle}(u) =$ black and $\mathbb{G}_{\langle p,q\rangle}(u{+}v) =$ white for some graph $\mathbb{G}_{\langle p,q\rangle}$,*
> *then we also have some $u' \in A$ and $v' \in V$ with $u'{+}v' = u{+}v \in \textbf{interior}(\textbf{shift}(A,v'))$*
> *such that $\mathbb{G}_{\langle p,q\rangle}(u') =$ black.*

Then the shift of A by any vector $v \in V$ is safe.

Proof: It suffices to demonstrate that the following relation is a simulation.

$$\mathcal{S} \;=\; \preceq \cup \Big\{ \langle p(m{+}i), q(n{+}j)\rangle : \langle m,n\rangle \in A,\ \langle i,j\rangle \in V,\ m{+}i, n{+}j \in \mathbb{N},\ p(m) \preceq q(n) \Big\}$$

To do this, we need to verify the simulation condition for each pair of $\mathcal{S}$. This is immediate for pairs in $\preceq$, so consider the pair $\langle p(m{+}i), q(n{+}j)\rangle \notin \preceq$ (that is, $\mathbb{G}_{\langle p,q\rangle}(m{+}i,n{+}j) =$ white) where $\langle m,n\rangle \in A$, $\langle i,j\rangle \in V$, $m{+}i, n{+}j \in \mathbb{N}$ and $p(m) \preceq q(n)$. By the premise in the Lemma, we can assume that $\langle m{+}i, n{+}j\rangle \in \textbf{interior}(\textbf{shift}(A,\langle i,j\rangle))$, and hence $\langle m,n\rangle \in \textbf{interior}(A)$, so $m, m{+}i, n, n{+}j > 0$.

If $p(m{+}i) \xrightarrow{a} p'(m'{+}i)$, then $p(m) \xrightarrow{a} p'(m')$, so $q(n) \xrightarrow{a} q'(n')$ with $p'(m') \preceq q'(n')$, so $q(n{+}j) \xrightarrow{a} q'(n'{+}j)$, and (as $\langle m',n'\rangle \in A$ and $m'{+}i, n'{+}j \in \mathbb{N}$) $\langle p'(m'{+}i), q'(n'{+}j)\rangle \in \mathcal{S}$. $\square$

Corollary 4 *The shift of A by v is safe if it is safe wrt* $\textbf{border}(\textbf{shift}(A,v))$.

By a ***line*** ℓ we mean a line with a finite rational slope $\beta > 0$; however, we shall occasionally refer explicitly to horizontal or vertical lines. We shall also view a line as a function, writing $\ell(y)$ to represent the value x such that the point $\langle x, y\rangle$ is on the line. We shall often refer to areas determined by a horizontal line at level $b \in \mathbb{N}$ and one or two lines. For this, we shall use the following notation: $A[b, \overrightarrow{\ell}, \overleftarrow{\ell'}]$ denotes the set of all points of $\mathbb{N}\times\mathbb{N}$ which lie on or above level b, on or to the right of ℓ, and on or to the left of ℓ'. We shall omit b when $b = 0$. Finally, by a ***belt*** we mean the set of points on or between two parallel lines; here we also allow horizontal and vertical lines. Thus we may have a horizontal belt, or a vertical belt, or a belt of the form $A[\overrightarrow{\ell}, \overleftarrow{\ell'}]$ where ℓ and ℓ' are parallel lines with ℓ' to the right of ℓ.

We can partition the frontiers according to whether or not they lie in a horizontal or a vertical belt. To this end we make the following definitions.

(i) **HF** is the set of frontiers f such that $f(n) = \infty$ for some $n \in \mathbb{N}$. We let $\mathbf{hb} \in \mathbb{N}$ (the "horizontal bound") be the least value such that $f(\mathbf{hb}) = \infty$ for all $f \in$ **HF**. The frontiers of **HF** are those which lie in a horizontal belt.

(ii) **VF** is the set of frontiers f such that $\lim_{n\to\infty} f(n) < \infty$. We let $\mathbf{vb} \in \mathbb{N}$ (the "vertical bound") be the least value such that $f(n) < \mathbf{vb}$ for all $f \in$ **VF** and all $n \in \mathbb{N}$. The frontiers of **VF** are those which lie in a vertical belt.

(iii) **IF** is the set of interior frontiers, those not appearing in **HF** nor in **VF**.

We shall now formalize the notion of a line *separating* frontiers. For this, we need the following notions. We shall refer to a (horizontal) shift of a line ℓ by an amount $i \in \mathbb{Z}$ by $\mathbf{lineshift}(\ell, i)$; this is the line ℓ' such that $\ell'(y) = \ell(y) + i$. Given $\beta{>}0$, we let $\mathbf{step}(\beta) \in \mathbb{N}$ be the least integral horizontal distance which two lines with slope β must be separated so as to fit a unit square between them; this ensures that, given two such lines ℓ and $\ell' = \mathbf{shift}(\ell, \mathbf{step}(\beta))$ we have $A[\overleftarrow{\ell}] \cap \mathbf{interior}(\mathbb{N}\times\mathbb{N}) \subseteq \mathbf{interior}(A[\overleftarrow{\ell'}])$. Note that $\mathbf{step}(\alpha) {\leq} \mathbf{step}(\beta)$ whenever $\alpha {\geq} \beta$.

Definition 1 *A line ℓ with rational slope $\beta > 0$* ***separates frontiers above level*** *$b \in \mathbb{N}$ iff:*

(i) *for all* $f \in \textsf{HF}$, $f(b) = \infty$; *that is,* $b \geq \mathbf{hb}$;

(ii) *for all* f, *if* $f(b) = -1$ *then* $f(n) = -1$ *for all* n;

(iii) *for all* $f \in \textsf{IF}$, $f(b) > \mathbf{vb}$;

(iv) *for all* f, *if* $f(b) \leq \ell(b)$ *then* $f(n) < \ell(n) - \mathbf{step}(\beta)$ *for all* $n \geq b$
(in which case we call f *an* ***ℓ-left frontier****)*;

(v) *for all* f, *if* $f(b) \geq \ell(b)$ *then* $f(n) > \ell(n) + \mathbf{step}(\beta)$ *for all* $n \geq b$
(in which case we call f *an* ***ℓ-right frontier****)*.

Thus the ℓ-left and ℓ-right frontiers are separated by a belt with (horizontal) width $2 \cdot \mathbf{step}(\beta)$ *centered on the line ℓ. We say simply that a line* ***separates frontiers*** *if it separates frontiers above some level.*

The next Lemma shows that there always exists such a separating line.

Lemma 5 *There is a line ℓ (with rational slope $\beta > 0$) which separates frontiers, in which the ℓ-right frontiers are exactly those of* $\textsf{HF}$.

Proof: If $\textsf{IF} = \emptyset$ then we can take ℓ to be any line, e.g., with slope 1/2. Thus assume that $\textsf{IF} \neq \emptyset$. Define $\mathbf{maxjump}(n) = \max\{f(n+1) - f(n) : f \in \textsf{IF}\}$, and let b be chosen so that the first three clauses of Definition 1 are satisfied and so that $\mathbf{maxjump}(b) > 0$.

Let $A = A[b+1, \overrightarrow{\ell}]$ where ℓ is the vertical line defined by $\ell(y) = \mathbf{vb} + \mathbf{maxjump}(b)$ and let $v = \langle -\mathbf{maxjump}(b), -1\rangle$. Then the shift of A by v wrt $\mathbf{border}(\mathbf{shift}(A, v))$ is safe: it is certainly safe for the graphs $\mathbb{G}_{\langle p,q\rangle}$ of the frontiers $f_{\langle p,q\rangle} \in \textsf{HF}$ (the relevant points are all black), and for the graphs $\mathbb{G}_{\langle p,q\rangle}$ of the frontiers $f_{\langle p,q\rangle} \in \textsf{VF}$ (the relevant points are all white); and for the graphs $\mathbb{G}_{\langle p,q\rangle}$ of the frontiers $f_{\langle p,q\rangle} \in \textsf{IF}$, the vertical border points are all black, so we could only shift a black point to a white point on the bottom, which would suggest a (contradictory) jump in the frontier greater than $\mathbf{maxjump}(b)$.

Thus by Corollary 4 the shift of A by v is safe, and we can then readily extract a separating line with slope $\beta = 1/\mathbf{maxjump}(b)$, since for any $u \in A$, if $\mathbb{G}_{\langle p,q\rangle}(u) = \text{white}$ then

$$\text{white} \;=\; \mathbb{G}_{\langle p,q\rangle}(u - v) \;=\; \mathbb{G}_{\langle p,q\rangle}(u - 2v) \;=\; \mathbb{G}_{\langle p,q\rangle}(u - 3v) \;=\; \cdots$$

since each point $(u - i \cdot v)$ is in A. For example, we can take a line with slope β which goes through the rightmost frontier point of $\textsf{IF}$ on level $b+1$ and shift this by an amount $\mathbf{step}(\beta)+1$. □

We now prove our Belt Theorem; along the way we shall state and use two technical lemmas, the proofs of which we shall defer until the end.

Proof of The Belt Theorem: Suppose we have a line ℓ with rational slope β which separates frontiers above level b in such a way that ℓ-right frontiers lie in belts and their number cannot be increased by choosing a different ℓ. That such a separating line exists is ensured by Lemma 5.

Let $\mathcal{L}$ be the set of ℓ-left frontiers, and suppose for the sake of contradiction that $\mathcal{L} - \mathsf{VF} \neq \emptyset$ (otherwise we have nothing to prove).

For any $n \geq b$, let $\mathbf{gap_1}(n)$ be the (horizontal) distance from ℓ to the rightmost ℓ-left frontier point on level n; that is, $\mathbf{gap_1}(n) = \min\{\ell(n) - f(n) : f \in \mathcal{L}\}$. Note that, since β is rational, the fractional part of $\mathbf{gap_1}(n)$ ranges over a finite set. Hence we cannot have an infinite sequence of levels $i_1, i_2, i_3, \ldots$ above b such that $\mathbf{gap_1}(i_1) > \mathbf{gap_1}(i_2) > \mathbf{gap_1}(i_3) > \cdots$. We can thus take an infinite sequence $i_1 < i_2 < i_3 < \cdots$ of levels above b such that

1. $\mathbf{gap_1}(i) \leq \mathbf{gap_1}(n)$ for all $i \in \{i_1, i_2, i_3, \ldots\}$ and all $n \geq i$;
2. either $\mathbf{gap_1}(i_1) = \mathbf{gap_1}(i_2) = \mathbf{gap_1}(i_3) = \cdots$
   or $\mathbf{gap_1}(i_1) < \mathbf{gap_1}(i_2) < \mathbf{gap_1}(i_3) < \cdots$;
3. for some fixed ℓ-left frontier $f_{\max} \in \mathcal{L}$: $\mathbf{gap_1}(i) = \ell(i) - f_{\max}(i)$ for all $i \in \{i_1, i_2, i_3, \ldots\}$.

The above conditions can be satisfied by starting with the infinite sequence $b+1, b+2, b+3, \ldots$, and first extracting an infinite subsequence which satisfies the first condition, then extracting from this a further infinite subsequence which satisfies (also) the second condition, and then extracting from this a further infinite subsequence which satisfies (also) the third condition.

For $i \in \{i_1, i_2, i_3, \ldots\}$, we let $\mathbf{offset}_i : \mathcal{L} \to \mathbb{N}$ be defined by $\mathbf{offset}_i(f) = f_{\max}(i) - f(i)$. We can then assume that our infinite sequence further satisfies the following condition.

4. For each $f \in \mathcal{L}$: either $\mathbf{offset}_{i_1}(f) = \mathbf{offset}_{i_2}(f) = \mathbf{offset}_{i_3}(f) = \cdots$
   or $\mathbf{offset}_{i_1}(f) < \mathbf{offset}_{i_2}(f) < \mathbf{offset}_{i_3}(f) < \cdots$.

   In the first case, we call f a ***fixed-offset frontier***; and in the second case, we call f an ***increasing-offset frontier***.

This condition can be satisfied by repeatedly extracting an infinite subsequence to satisfy the condition for each $f \in \mathcal{L}$ in turn. Finally, we assume our sequence satisfies the following condition.

5. We have a maximal number of fixed-offset frontiers; no other sequence satisfying conditions 1–4 can have more ℓ-left frontiers $f \in \mathcal{L}$ with $\mathbf{offset}_{i_1}(f) = \mathbf{offset}_{i_2}(f) = \cdots$.

For technical reasons, we also suppose the next two conditions which can be satisfied by dropping some number of initial levels (that is, sequence elements).

6. $\mathbf{gap_2}(i_1) > |\mathcal{L}| \cdot \mathbf{step}(\beta)$, where $\mathbf{gap_2}(i_j)$ is defined as

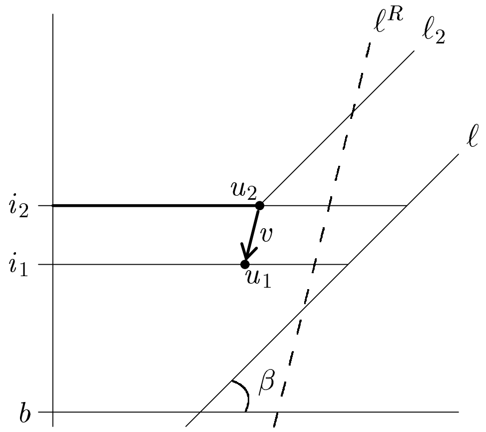
$$\min\{\mathbf{offset}_{i_j}(f) : f \text{ is an increasing-offset frontier}\} \\ - \max\{\mathbf{offset}_{i_j}(f) : f \text{ is a fixed-offset frontier}\}$$

7. $f_{\max}(i_1) < f_{\max}(i_2)$.

The line going through the points $u_1 = \langle f_{\max}(i_1), i_1 \rangle$ and $u_2 = \langle f_{\max}(i_2), i_2 \rangle$ has some slope $\alpha \geq \beta$. If we let **left-of**(u_2) denote the set of points consisting of u_2 along with all points to its left (that is, all (m, i_2) with $m \leq f_{\max}(i_2)$) then the shift of **left-of**(u_2) by $v = u_1 - u_2$ is safe: for the shift of the point onto the y-axis, this is assured by condition (ii) of Definition 1; and for the remaining points, this is assured since frontier offsets cannot shrink (condition 4 above). We can thus invoke the following.

Right Lemma *Consider a line ℓ with slope β which separates frontiers above level b, and take two points $u_1 = \langle m_1, i_1 \rangle$ and $u_2 = \langle m_2, i_2 \rangle$ in $A[b, \overleftarrow{\ell}]$ with $m_1 < m_2$ and $b < i_1 < i_2$ such that the slope α of the vector $v = u_1 - u_2$ is at least β. Let ℓ_2 be the line parallel to ℓ which goes through u_2, and suppose that all ℓ-left frontier points in $A[i_2]$ are in $A[\overleftarrow{\ell_2}]$, and that the shift of* **left-of**(u_2) *by v is safe.*

Then there is a line ℓ^R with slope α which separates the same frontiers as ℓ does above level i_2.

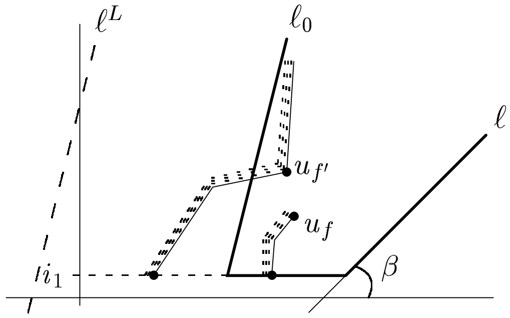


Right Lemma thus gives us a line ℓ^R with slope α separating the same frontiers as ℓ above i_2.

Now, there must then be a line ℓ_0 with slope α to the left of ℓ above level i_1 such that every fixed-offset frontier appears in (that is, intersects with) $A = A[i_1, \overrightarrow{\ell_0}, \overleftarrow{\ell}]$, and such that whenever a frontier f appears in A there is a frontier point $u_f = \langle f(n), n \rangle \in \mathbf{interior}(A)$ such that $f(n) - \ell_0(n) \geq f(i_1) - \ell_0(i_1)$; that is, u_f is at least as far to the right of ℓ_0 as the frontier point of f on level i_1. (We can first consider ℓ_0 to be the line going through the frontier points at levels i_1 and i_2 of the fixed-offset frontier with the greatest offset value; if some frontier f is on the border but not in the interior of A, then we can instead take ℓ_0 to be the shift of this line by $-\mathbf{step}(\alpha)$. If there is now some other frontier which is on the border but not in the interior of A, then we again shift the line by $-\mathbf{step}(\alpha)$. We need only shift (at most) once for each frontier in $\mathcal{L}$ before being guaranteed to arrive at a suitable choice for ℓ_0, so we shift by at most $|\mathcal{L}| \cdot \mathbf{step}(\alpha) \leq |\mathcal{L}| \cdot \mathbf{step}(\beta)$, and hence by condition 5 we don't reach the increasing-offset frontiers on level i_1.) We may then invoke the following.

Left Lemma *Suppose we have a line ℓ with rational slope β which separates frontiers above level i_1, and a line ℓ_0 with rational slope $\alpha \geq \beta$ which is to the left of ℓ above level i_1. Suppose further that whenever a frontier f appears in $A = A[i_1, \overrightarrow{\ell_0}, \overleftarrow{\ell}]$, there is a frontier point $u_f = \langle f(n), n \rangle \in$ **interior**(A) such that $f(n) - \ell_0(n) \geq f(i_1) - \ell_0(i_1)$.*

Then there is a line ℓ^L to the left of ℓ_0 with slope α such that $f \subseteq A[\overrightarrow{\ell^L}]$ for each such frontier f.



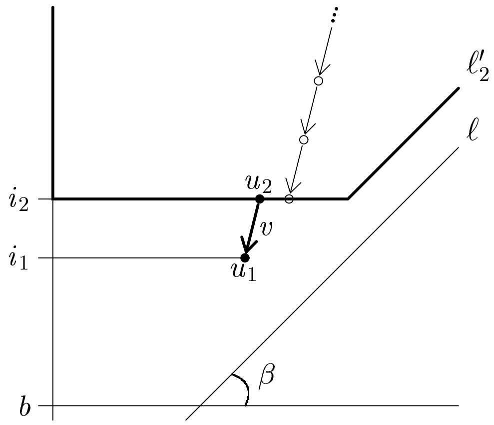
The premise of Left Lemma is thus satisfied, so all frontiers with frontier points in A are in $A[\overrightarrow{\ell^L}]$ for some line ℓ^L with slope α. Hence they are in the belt $A[\overrightarrow{\ell^L}, \overleftarrow{\ell^R}]$ above i_2; and in fact only the fixed-offset frontiers can (and do) have frontier points in $A = A[i_1, \overrightarrow{\ell_0}, \overleftarrow{\ell}]$, for otherwise they would not correspond to increasing-offset frontiers.

It remains to demonstrate that we can choose ℓ^L so that it separates frontiers. This can only fail if an increasing-offset frontier appears infinitely often in $A[\overrightarrow{\ell'}]$ where $\ell' =$ **shift**$(\ell^L, -2 \cdot$ **step**$(\alpha))$. But then there would be two levels i_{j_1} and i_{j_2} where $f_{\max}(i_{j_1}) - \ell^L(i_{j_1}) = f_{\max}(i_{j_2}) - \ell^L(i_{j_2}) = d$ and $\mathbf{gap_2}(i_{j_1}) > d + |\mathcal{L}| \cdot$ **step**(α). We could then find a contradiction using Left Lemma, by considering now the area $A[i_{j_1}, \overrightarrow{\ell^L}, \overleftarrow{\ell^R}]$. $\square$

Proof of Right Lemma: It suffices to show that the shift of $A = A[i_2, \overleftarrow{\ell'_2}]$ by v is safe, where $\ell'_2 = \mathbf{shift}(\ell_2, \mathbf{step}(\beta))$, for then we get that for any $u \in A$, if $\mathbb{G}_{\langle p,q\rangle}(u) = \text{white}$ then

$$\begin{aligned}\text{white} &= \mathbb{G}_{\langle p,q\rangle}(u - v)\\ &= \mathbb{G}_{\langle p,q\rangle}(u - 2v)\\ &= \mathbb{G}_{\langle p,q\rangle}(u - 3v)\\ &= \cdots\end{aligned}$$

since each point $(u - i{\cdot}v)$ is in A, from which the result readily follows.

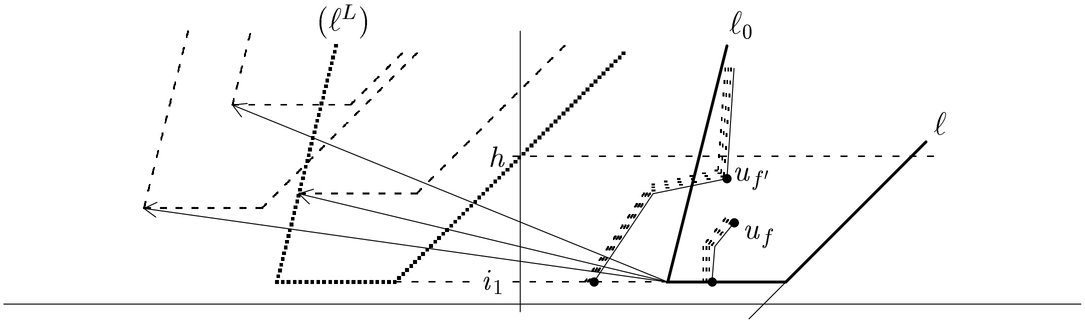


To show that the shift of A by v is safe, by Corollary 4 we need only check that the shift of A by v wrt $\mathbf{border}(\mathbf{shift}(A, v))$ is safe.

- The safety of the shift to the y-axis border points is assured since if $\mathbb{G}_{\langle p,q\rangle}(0, j) = \text{white}$ for $j \geq b$, then by the second clause of Definition 1, $\mathbb{G}_{\langle p,q\rangle}(i, j) = \text{white}$ for all $i, j \in \mathbb{N}$.
- The safety of the shift of **left-of**(u_2) is assumed in the premise of the Lemma.
- The safety of the remaining border points is assured, as the only black points which are shifted to these border points appear in the graphs of ℓ-right frontiers; and by the final clause of Definition 1, the border points in this case must all be black. □

Proof of Left Lemma: Let $h > i_1$ be a level such that the occurrences of the frontier points u_f appearing in A as described in the Lemma appear below level h.

Take $d \in \mathbb{N}$ so that shifting A to the left by d units moves the portion of A below level h to the left of the y-axis. Then let V be the set of vectors describing shifts to the left by a (rational) amount $d' \geq d$ followed by a (possibly null) shift upwards at an angle of α.



Suppose $u \in A$ and $v \in V$ with $u{+}v \in \mathbf{border}(\mathbf{shift}(A, v))$ such that for some graph $\mathbb{G}_{\langle p,q\rangle}$, $\mathbb{G}_{\langle p,q\rangle}(u) = \text{black}$ and $\mathbb{G}_{\langle p,q\rangle}(u{+}v) = \text{white}$. Then $u{+}v$ must be above level h in $\mathbf{interior}(\mathbb{N}\times\mathbb{N})$, $f_{\langle p,q\rangle}$ must appear in A, and u must be on the bottom- or ℓ_0-border of A. Let $u_{f_{\langle p,q\rangle}}$ be the black point in $\mathbf{interior}(A)$ given in the statement of the Lemma. As $u_{f_{\langle p,q\rangle}}$ must be at least as far to the right of ℓ_0 as u is, the vector $v' = v + (u - u_{f_{\langle p,q\rangle}})$ is readily seen to be in V.

The premise of Fact 3 is thus satisfied, so the shift of A by any vector $v \in V$ is safe. The existence of the proposed line ℓ^L then follows immediately (in fact, we can take $\ell^L = \mathbf{shift}(\ell_0, -d)$). $\square$