# Turn-based stochastic games with finitely and infinitely many states

**Antonín Kučera**

Masaryk University Brno





## Outline

- Preliminaries.
  - Game graphs can be infinite.
  - Strategies may be randomized and history-dependant.
- Stochastic games with reachability objectives.
  - The existence of a value.
  - The (non)existence of optimal strategies.
  - Algorithms for finite-state games.
  - Algorithms for infinite-state games.
- Stochastic games with branching-time objectives.

# Preliminaries

**We start by recalling...**

- Markov chains and the associated probability space.
- Turn-based stochastic games, strategies, and plays.
- Linear-time objectives.

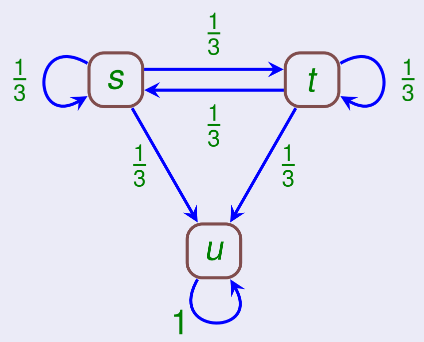


Stochastic games
Antonín Kučera

Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

# Markov chains

**Definition 1 (Markov chain)**




$\mathcal{M} = (S, \rightarrow, Prob)$

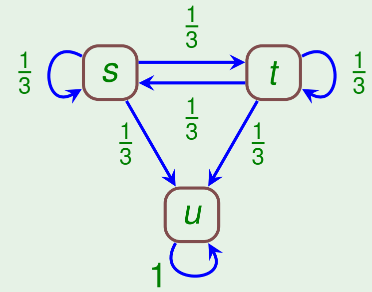
- $S$ is at most countable set of states;
- $\rightarrow \subseteq S \times S$ is a transition relation;
- $Prob$ is a probability assignment which to every transition assigns a positive probability s.t. $\sum_{s \xrightarrow{x} t} x = 1$ for every $s \in S$.

We want to measure the probability of certain subsets of $Run(s)$.

- For every finite path $w$ initiated in $s$, we define the probability of $Run(w)$ in the natural way.
- This assignment can be uniquely extended to the (Borel) $\sigma$-algebra $\mathcal{F}$ generated by all $Run(w)$.
- Thus, we obtain the probability space $(Run(s), \mathcal{F}, \mathcal{P})$.

# Markov chains (2)

## Example 2




- $Reach(u) = \{w \in Run(s) \mid w \text{ visits the vertex } u\}$.
- $Reach^i(u) = \{w \in Run(s) \mid w(i) = u, w(j) \neq u \text{ for every } j < i\}$.
- $Reach^i(u)$ is measurable, hence $Reach(u) = \biguplus_{i\in\mathbb{N}} Reach^i(u)$ is also measurable.

$$\mathcal{P}(Reach(u)) = \sum_{i\in\mathbb{N}} \mathcal{P}(Reach^i(u)) = \sum_{i\in\mathbb{N}} \left(\frac{2}{3}\right)^{i-1}\frac{1}{3} = 1$$

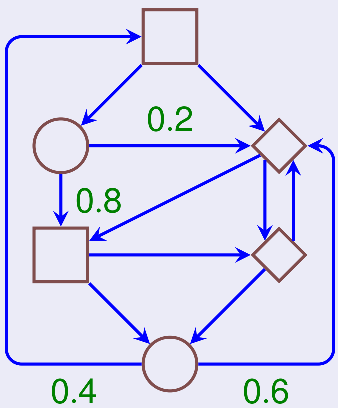
- Note that $Run(s) \setminus Reach(u)$ is not countable, and its probability is $0$.



Stochastic games
Antonín Kučera
Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

# Turn-based stochastic games

## Definition 3 (Turn-based stochastic game)




$G = (V, E, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$

- *the set $V$ is at most countable;*
- *each vertex has a successor;*
- *$Prob$ is positive;*
- *$G$ is a Markov decision process (MDP) if $V_\Diamond = \emptyset$ or $V_\Box = \emptyset$.*

- A (linear-time) winning objective is a Borel set $\mathcal{W}$ of runs in $G$.
- The aim of player $\Box$ is to maximize the probability of $\mathcal{W}$, the aim of player $\Diamond$ is to minimize this probability.

# Strategies

**Definition 4 (Strategy)**

*Let $G = (V, E, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ be a game. A strategy for player $\Box$ is a function $\sigma$ which to every $wv \in V^* V_\Box$ assigns a probability distribution over the set of outgoing edges of $v$.*

- A strategy for player $\Diamond$ is defined analogously.
- We can classify strategies according to
  - memory requirements: history-dependent (H), finite-memory (F), memoryless (M)
  - randomization: randomized (R), deterministic (D)
- Thus, we obtain the classes of MD, MR, FD, FR, HD, and HR strategies.



Stochastic games
Antonín Kučera
Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

# Plays

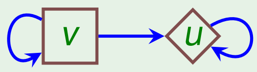
**Definition 5 (Play)**

*Let $G = (V, E, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ be a game. Each pair $(\sigma, \pi)$ of strategies for player $\Box$ and player $\Diamond$ determines a unique play $G^{(\sigma,\pi)}$, which is a Markov chain where $V^+$ is the set of states and transitions are defined accordingly.*

- Plays are infinite trees.
- Each run $w \in Run_{G^{(\sigma,\pi)}}(v)$ determines a unique run $w_G \in Run_G(v)$.
- If $\mathcal{W} \subseteq Run_G(v)$ is Borel, then $\mathcal{W}^{(\sigma,\pi)} = \{w \in Run_{G^{(\sigma,\pi)}}(v) \mid w_G \in \mathcal{W}\}$ is measurable for every pair of strategies $(\sigma, \pi)$.
- For a pair of memoryless strategies $(\sigma, \pi)$, the play $G^{(\sigma,\pi)}$ can be depicted as a Markov chain with the set of states $V$.

# Plays (2)

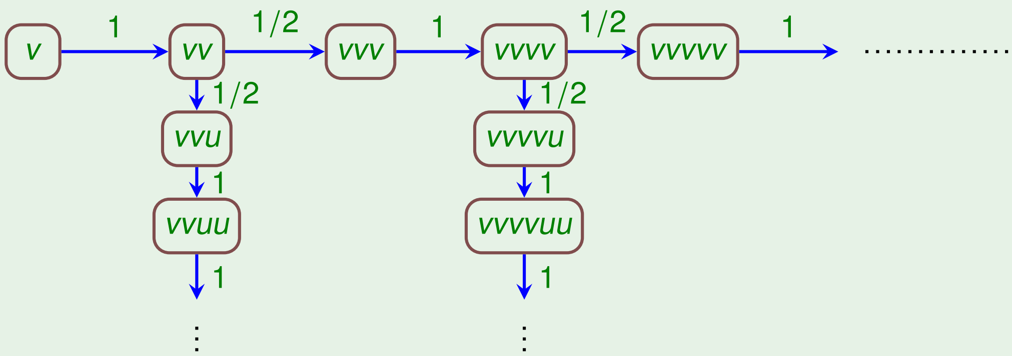
**Example 6 (A game and its play)**

- 

$\mathcal{W} = \{w \in Run(v) \mid w(i) = u \text{ for some } i \in \mathbb{N}\}$

- $\sigma(wv) = \begin{cases} v \xrightarrow{1} v & \text{if } |wv| \text{ is odd;} \\ v \xrightarrow{1/2} v,\ v \xrightarrow{1/2} u & \text{otherwise.} \end{cases} \qquad \pi(wu) = u \xrightarrow{1} u$

- $\mathcal{P}(\mathcal{W}^{(\sigma,\pi)}) = 1$








# Stochastic games have a value

**Theorem 7 (Donald Martin, 1998)**

*Let $G = (V, E, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ be a game, $v \in V$, and $\mathcal{W} \subseteq Run_G(v)$ a Borel set of runs. Then*

$$\sup_\sigma \inf_\pi \mathcal{P}(\mathcal{W}^{(\sigma,\pi)}) \quad = \quad \inf_\pi \sup_\sigma \mathcal{P}(\mathcal{W}^{(\sigma,\pi)})$$

- The equality of Thm. 7 defines the $\mathcal{W}$-value of $v$, denoted $val_{\mathcal{W}}(v)$.
- Thm. 7 does not impose any restrictions on $G$. The set of vertices and the branching degree of $G$ can be infinite.
- References:

  - D.A. Martin. *The Determinacy of Blackwell Games.* The Journal of Symbolic Logic, Vol. 63, No. 4 (Dec., 1998), pp. 1565–1581.
  - A. Maitra and W. Sudderth. *Finitely Additive Stochastic Games with Borel Measurable Payoffs.* International Journal of Game Theory, Vol. 27 (1998), pp. 257–267.

# Optimal strategies

## Definition 8

*Let $G = (V, E, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ be a game, $v \in V$, and $\mathcal{W} \subseteq Run_G(v)$ a Borel set of runs. Let $\varepsilon \in [0,1]$.*

- *An $\varepsilon$-optimal maximizing strategy is a strategy $\sigma$ for player $\Box$ such that for every strategy $\pi$ of player $\Diamond$ we have that $\mathcal{P}(\mathcal{W}^{(\sigma,\pi)}) \geq val_{\mathcal{W}}(v) - \varepsilon$.*
- *An $\varepsilon$-optimal minimizing strategy is a strategy $\pi$ for player $\Diamond$ such that for every strategy $\sigma$ of player $\Box$ we have that $\mathcal{P}(\mathcal{W}^{(\sigma,\pi)}) \leq val_{\mathcal{W}}(v) + \varepsilon$.*

*An optimal maximizing/minimizing strategy is a $0$-optimal maximizing/minimizing strategy.*

- According to Thm. 7, $\varepsilon$-optimal maximizing/minimizing strategies exist for every $\varepsilon > 0$.
- ...and we cannot say much more in the general setting.





# Reachability objectives

**Now we restrict ourselves to reachability objectives and show the following:**

- Properties of minimizing strategies:
  - optimal minimizing strategies do not necessarily exist and $\varepsilon$-optimal strategies may require infinite memory, even for MDPs;
  - in every finitely-branching game, there is an optimal minimizing MD strategy;
- Properties of maximizing strategies:
  - optimal maximizing strategies do not necessarily exist, even for finitely-branching MDPs;
  - in every finite-state game, there is an optimal maximizing MD strategy;

# Reachability games have a value (1)

**Theorem 9**

*Let $G = (V, E, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ be a game, $u \in V$ a target vertex. For every $v \in V$ we have that*

$$\sup_\sigma \inf_\pi \mathcal{P}(Reach(u)^{(\sigma,\pi)}) \quad = \quad \inf_\pi \sup_\sigma \mathcal{P}(Reach(u)^{(\sigma,\pi)})$$



Stochastic games
Antonín Kučera

Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

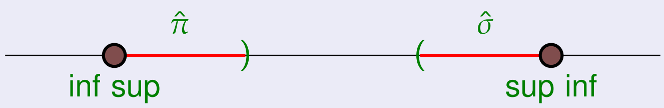
# Reachability games have a value (2)

**Proof sketch.**

- Let $\Gamma : [0,1]^{|V|} \to [0,1]^{|V|}$ be a (monotonic) function defined by

$$\Gamma(\alpha)(v) = \begin{cases} 1 & \text{if } v = u; \\ \sup\{\alpha(v') \mid (v,v') \in E\} & \text{if } v \neq u \text{ and } v \in V_\Box; \\ \inf\{\alpha(v') \mid (v,v') \in E\} & \text{if } v \neq u \text{ and } v \in V_\Diamond; \\ \sum_{(v,v')\in E} Prob(v,v') \cdot \alpha(v') & \text{if } v \neq u \text{ and } v \in V_\bigcirc. \end{cases}$$

- $\mu\Gamma(v) \ \leq \ \sup_\sigma \inf_\pi \mathcal{P}(Reach(u)^{(\sigma,\pi)}) \ \leq \ \inf_\pi \sup_\sigma \mathcal{P}(Reach(u)^{(\sigma,\pi)})$
  - the second inequality holds for all Borel objectives.



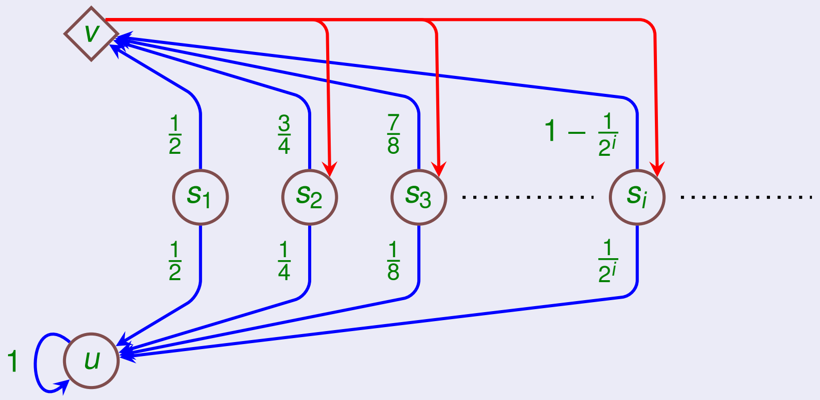

  - the tuple of all $\sup_\sigma \inf_\pi \mathcal{P}(Reach(u)^{(\sigma,\pi)})$ is a fixed-point of $\Gamma$;
- It cannot be that $\mu\Gamma(v) \ < \ \inf_\pi \sup_\sigma \ \mathcal{P}(Reach(u)^{(\sigma,\pi)})$
  - For all $\varepsilon > 0$ and $v \in V$, there is a strategy $\hat{\pi}$ such that $\sup_\sigma \mathcal{P}(Reach(u)^{(\sigma,\hat{\pi})}) \ \leq \ \mu\Gamma(v) + \varepsilon$. $\Box$

# Minimizing strategies (1)

## Theorem 10

*Optimal minimizing strategies do not necessarily exist, and $\varepsilon$-optimal minimizing strategies may require infinite memory (even for MDPs).*

## Proof.




□





# Minimizing strategies (2)

## Definition 11 (Locally optimal strategy)

*Let $G = (V, E, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ be a game.*

- *An edge $(v, v') \in E$ is value minimizing if*

$$val(v') = \min \left\{ val(\hat{v}) \in V \mid (v, \hat{v}) \in E \right\}$$

- *A locally optimal minimizing strategy is a strategy which in every play selects only value minimizing edges.*
- *Analogously, we define value maximizing edges and locally optimal maximizing strategies.*

## Observation 12

- *Every optimal maximizing/minimizing strategy is also a locally optimal maximizing/minimizing strategy.*
- *For every finitely-branching game, there is a locally optimal maximizing/minimizing MD strategy.*

# Minimizing strategies (3)

**Theorem 13**

*Every locally optimal min. strategy is an optimal min. strategy.*

**Proof.**

Let $v \in V$ be an initial vertex, and $u \in V$ a target vertex.

(1) After playing $k$ rounds according to a locally optimal minimizing strategy, player $\Diamond$ can switch to $\varepsilon$-optimal minimizing strategies in the current vertices of the play. Thus, we always (for every $k$ and $\varepsilon > 0$) obtain an $\varepsilon$-optimal minimizing strategy for $v$.

(2) Let $\pi$ be a locally optimal min. strategy which is not optimal.

- Then there is a strategy $\sigma$ of player $\Box$ such that $\mathcal{P}(Reach(u)^{(\sigma,\pi)}) = val(v) + \delta$, where $\delta > 0$.
- This means that there is $k \in \mathbb{N}$ such that $\mathcal{P}(Reach^k(u)^{(\sigma,\pi)}) > val(v) + \frac{\delta}{2}$.
- Hence, if player $\Diamond$ switches to $\frac{\delta}{2}$-optimal minimizing strategy after playing $k$ rounds according to $\pi$, we do not obtain a $\frac{\delta}{2}$-optimal minimizing strategy for $v$. $\Box$





# Minimizing strategies (4)

**Corollary 14 (Properties of minimizing strategies.)**

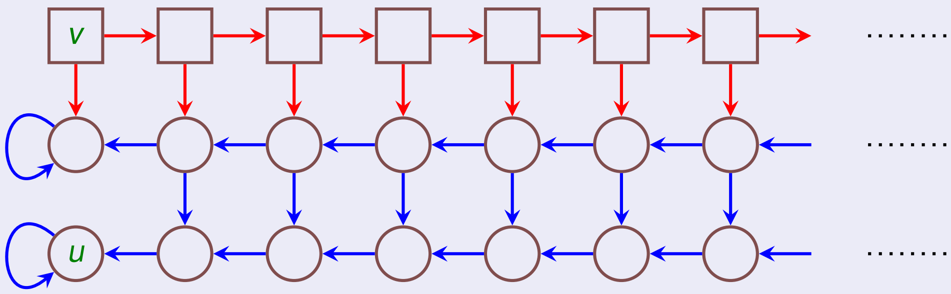
- *Optimal minimizing strategies do not necessarily exist and $\varepsilon$-optimal minimizing strategies may require infinite memory, even for MDPs.*
- *In every finitely-branching game, there is an optimal minimizing MD strategy.*
- *If there is some optimal minimizing strategy, then there is also an optimal minimizing MD strategy.*

# Maximizing strategies (1)

## Theorem 15

*Optimal maximizing strategies do not necessarily exist (even for finitely-branching MDPs).*

## Proof.




□



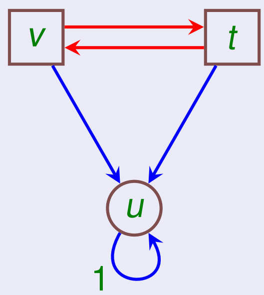
Stochastic games
Antonín Kučera
Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

# Maximizing strategies (2)

## Observation 16

*A locally optimal maximizing strategy is not necessarily an optimal maximizing strategy. This holds even for finite-state MDPs.*

## Proof.



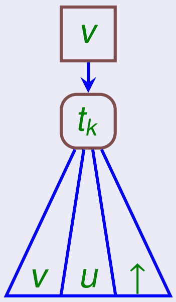

□

# Maximizing strategies (3)

### Theorem 17

*Let $v \in V_\Box$ be a vertex with finitely many successors $t_1, \ldots, t_n$. Then there is $1 \le i \le n$ such that $val(v)$ does not change if all edges $(v, t_j)$, where $i \neq j$, are deleted from the game.*

### Proof.




- $V_{t_k}^{(\sigma,\pi)} = \begin{cases} \frac{\mathcal{P}(u)}{\mathcal{P}(u)+\mathcal{P}(\uparrow)} & \text{if } \mathcal{P}(u)+\mathcal{P}(\uparrow) > 0; \\ 0 & \text{otherwise;} \end{cases}$
- $V_{t_k}^{\sigma} = \inf_\pi V_{t_k}^{(\sigma,\pi)}$
- $V_{t_k} = \sup_\sigma V_{t_k}^{\sigma}$
- There must be some $k$ such that $V_{t_k} = val(v)$.
- We put $i = k$.

$\square$





# Maximizing strategies (4)

### Corollary 18 (Properties of maximizing strategies.)

- *Optimal maximizing strategies do not necessarily exist, even for finitely-branching MDPs.*
- *In every finite-state game, there is an optimal maximizing MD strategy.*

References:


- M.L. Puterman. *Markov Decision Processes*, Wiley, 1994. (Theorem 7.2.11 implies the last claim of Corollary 18 for MDPs.)
- T. Brázdil, V. Brožek, V. Forejt, A. Kučera. *Reachability in recursive Markov decision processes.* Information and Computation, vol. 206, pp. 520–537, 2008.

# Algorithms for finite-state MDPs and games

**We show how to compute the values and optimal strategies for reachability objectives in finite-state games and MDPs.**

- For finite-state MDPs we have that
  - the values and optimal strategies are computable in polynomial time;
- For finite-state games we have that
  - the values and optimal strategies are computable in polynomial space (for a fixed number of randomized vertices, the problem is in **P**);



Stochastic games
Antonín Kučera
Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

# Finite-state MDPs (1)

**Theorem 19**

*Let $G = (V, E, (V_\square, V_\bigcirc), Prob)$ be a finite-state MDP. Then*

- $\mathcal{V}^{=0} = \{v \in V \mid val(v) = 0\}$
- $\mathcal{V}^{=1} = \{v \in V \mid val(v) = 1\}$

*are computable in polynomial time.*

**Proof.**

It suffices to realize that $\mathcal{V}^{=1}$ is exactly the greatest $S \subseteq V$ satisfying the following conditions:

- If $v \in S$, then there is a finite path from $v$ to the target vertex $u$ which visits only the vertices of $S$.
- If $v \in S \cap V_\bigcirc$, then all successors of $v$ belong to $S$.

Hence, $\mathcal{V}^{=1}$ is computable in polynomial time. The set $\mathcal{V}^{=0}$ can be computed similarly. Note that the sets $\mathcal{V}^{=1}$ and $\mathcal{V}^{=0}$ depend only on the "topology" of $G$. □

# Finite-state MDPs (2)

**Theorem 20**

*Let $G = (V, E, (V_\Box, V_\bigcirc), Prob)$ be a finite-state MDP where $Prob$ is rational. The values $val(v)$, $v \in V$, are rational and computable in polynomial time. An optimal maximizing strategy is also constructible in polynomial time.*

**Proof.**

Let $V = \{v_1, \ldots, v_n\}$, where $v_n$ is the target vertex.

**minimize** $x_1 + \cdots + x_n$
subject to
$x_n = 1$
$x_i \geq x_j$ for all $(v_i, v_j) \in E$ where $v_i \in V_\Box$ and $i < n$
$x_i = \sum_{(v_i,v_j)\in E} Prob(v_i, v_j) \cdot x_j$ for all $v_i \in V_\bigcirc$, $i < n$
$x_i \geq 0$ for all $i \in \{1, \ldots, n\}$

An optimal strategy can be constructed by successively removing the ougoing edges of every $v \in V_\Box$ untill only one such edge is left. □



Stochastic games
Antonín Kučera
Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

# Finite-state MDPs (3)

Similarly, one can prove the following theorem about minimizing MDPs.

**Theorem 21**

*Let $G = (V, E, (V_\Diamond, V_\bigcirc), Prob)$ be a finite-state MDP. The sets*

- $\mathcal{V}^{=0} = \{v \in V \mid val(v) = 0\}$
- $\mathcal{V}^{=1} = \{v \in V \mid val(v) = 1\}$

*are computable in polynomial time. If $Prob$ is rational, then the values $val(v)$, $v \in V$, are rational and computable in polynomial time. An optimal minimizing strategy is also constructible in polynomial time.*

# Finite-state games (1)

### Theorem 22

*Let $G = (V, E, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ be a finite-state game. Then*

- $\mathcal{V}^{=0} = \{v \in V \mid val(v) = 0\}$
- $\mathcal{V}^{=1} = \{v \in V \mid val(v) = 1\}$

*are computable in polynomial time.*

### Proof.

- $\mathcal{V}^{>0} = \mu\Gamma$, where $\Gamma : 2^V \to 2^V$ is defined as follows:

$$\begin{aligned}\Gamma(A) &= \{u\} \cup \{v \in V_\Box \cup V_\bigcirc \mid \exists (v,v') \in E \text{ s.t. } v' \in A\}\\ &\cup \{v \in V_\Diamond \mid \forall (v,v') \in E \text{ we have that } v' \in A\}\end{aligned}$$

- $\mathcal{V}^{=0} = V \setminus \mathcal{V}^{>0}$
- $\mathcal{V}^{<1} = \mu\Gamma$, where $\Gamma : 2^V \to 2^V$ is defined as follows:

$$\begin{aligned}\Gamma(A) &= \mathcal{V}^{=0} \cup \{v \in V_\Diamond \cup V_\bigcirc \mid \exists (v,v') \in E \text{ s.t. } v' \in A\}\\ &\cup \{v \in V_\Box \mid \forall (v,v') \in E \text{ we have that } v' \in A\}\end{aligned}$$

- $\mathcal{V}^{=1} = V \setminus \mathcal{V}^{<1}$

□



Stochastic games

Antonín Kučera

Preliminaries

Reachability objectives

The value

Min strategies

Max strategies

Finite-state games

BPA games

Taxonomy of objectives

Branching-time objectives

Basic properties

Deciding the winner

# Finite-state games (2)

### Theorem 23 (Anne Condon, 1992)

*Let $G = (V, E, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ be a finite-state game. The problem whether $val(v) > \frac{1}{2}$ for a given $v \in V$ is in* ***NP*** $\cap$ ***coNP****.*

### Proof.

Since both players have optimal MD strategies, it suffices to

- “guess” an optimal MD strategy for player □ (or player ◇);
- compute the value in the resulting MDP by solving the associated linear program.

□

Obviously, $val(v)$ and the optimal strategies for both players are computable in polynomial space.

# Finite-state games (3)

The original proof of Condon is somewhat different.

- It is assumed that randomized vertices have two successors, and the associted edges have probability $\frac{1}{2}$.
- It is shown that if $val(v) > \frac{1}{2}$, then $val(v) > \frac{1}{2} + \frac{1}{4^n}$, where $n = |V|$.
- The original game $G$ is efficiently transformed into another stopping game $G'$ such that $val(v) > \frac{1}{2}$ in $G$ iff $val(v) > \frac{1}{2}$ in $G'$.


- The main advantage of stopping games is that optimality equations have a unique solution. Thus, it is shown that both players have optimal MD strategies in stopping games.
- Many (not all) of the existing algorithms which compute the value and an optimal strategy in stochastic games assume that the game is stopping.



Stochastic games
Antonín Kučera
Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

# Finite-state games (4)

**Theorem 24 (Gimbert, Horn, 2008)**

*The values and MD optimal strategies in a finite-state game $G = (V, E, (V_\Box, V_\Diamond, V_\bigcirc), Prob)$ are computable in*

$$\mathcal{O}\Big(|V_\bigcirc|! \cdot (\log(|V|)\,|E| + |p|)\Big)$$

*time, where $|p|$ is the maximal bit-length of an edge probability.*

**Remark 25**

The question whether finite-state stochastic games are solvable in **P** is a longstanding open problem in algorithmic game theory.

References:

- A. Condon. *The Complexity of Stochastic Games.* Information and Computation, 96(2):203–224, 1992.
- L.S. Shapley. *Stochastic games.* Proceedings of the National Academy of Sciences USA, 39:1095–1100, 1953.
- H. Gimbert, F. Horn. *Simple Stochastic Games with Few Random Vertices Are Easy to Solve.* Proc. FoSSaCS 2008, pp. 5–19, LNCS 4962, Springer, 2008.

# Infinite-state games

- Interesting classes of infinite-state stochastic games are obtained by extending non-deterministic computational devices with randomized choice. So far, most of the results consider
  - pushdown automata (recursive state machines);
  - lossy channel systems.





# Stochastic BPA games (1)

### Definition 26

*A stochastic BPA game is a tuple* $\Delta = (\Gamma, \hookrightarrow, (\Gamma_\Box, \Gamma_\Diamond, \Gamma_\bigcirc), Prob)$ *where*

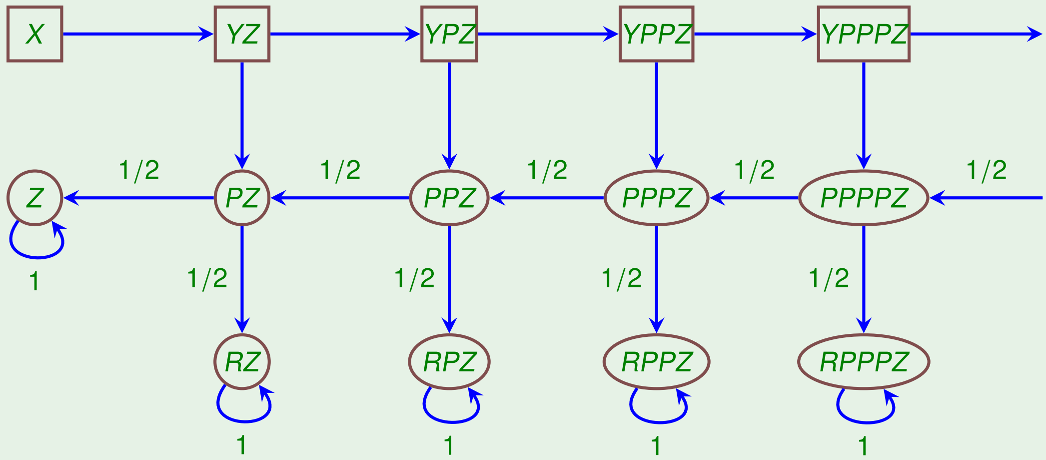
- $\Gamma$ *is a finite stack alphabet,*
- $\hookrightarrow \subseteq \Gamma \times \Gamma^{\leq 2}$ *is a finite set of rules,*
- $(\Gamma_\Box, \Gamma_\Diamond, \Gamma_\bigcirc)$ *is a partition of* $\Gamma$,
- $Prob$ *is a probability assignment which to each* $X \in \Gamma_\bigcirc$ *assigns a rational positive probability distribution on the set of all rules of the form* $X \hookrightarrow \alpha$.

# Stochastic BPA games (2)

## Example 27

Let $\Gamma = \{X, Y, Z, P, R\}$, where $\Gamma_\Box = \{X, Y\}$, $\Gamma_\Diamond = \emptyset$, $\Gamma_\bigcirc = \{P, R\}$, and

$$X \hookrightarrow YZ, \quad Y \hookrightarrow YP, \quad Y \hookrightarrow P, \quad P \overset{1/2}{\hookrightarrow} R, \quad P \overset{1/2}{\hookrightarrow} \varepsilon, \quad R \overset{1}{\hookrightarrow} R$$






Stochastic games
Antonín Kučera
Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

# BPA MDPs with reachability objectives (1)

- Let $\Delta = (\Gamma, \hookrightarrow, (\Gamma_\Box, \Gamma_\bigcirc), Prob)$ be a BPA Markov decision process, and $T \subseteq \Gamma^*$ a regular set of target configurations.
- We can safely assume that $T = \mathcal{R}\Gamma^*$, where $\mathcal{R} \subseteq \Gamma$.
- Consider the sets
  - $\mathcal{W}^{>0} = \{\alpha \in \Gamma^* \mid \exists \sigma : \mathcal{P}(Reach(\alpha \to T)^\sigma) > 0\}$
  - $\mathcal{W}^{=0} = \{\alpha \in \Gamma^* \mid \exists \sigma : \mathcal{P}(Reach(\alpha \to T)^\sigma) = 0\}$
  - $\mathcal{W}^{=1} = \{\alpha \in \Gamma^* \mid \exists \sigma : \mathcal{P}(Reach(\alpha \to T)^\sigma) = 1\}$
  - $\mathcal{W}^{<1} = \{\alpha \in \Gamma^* \mid \exists \sigma : \mathcal{P}(Reach(\alpha \to T)^\sigma) < 1\}$

  We show that these sets are regular and the associated finite-state automata are computable in polynomial time.

# BPA MDPs with reachability objectives (2)

**Theorem 28**

*The set $\mathcal{W}^{>0}$ is effectively regular.*

- Let us consider two sets of stack symbols:
  - $\mathcal{A} = \{X \in \Gamma \mid X \hookrightarrow^* \varepsilon\}$
  - $\mathcal{B} = \{X \in \Gamma \mid X \hookrightarrow^* R\beta, \text{ where } R \in \mathcal{R} \text{ and } \beta \in \Gamma^*\}$
- We have that $\mathcal{W}^{>0} = \mathcal{A}^*\mathcal{B}\Gamma^*$.
- The sets $\mathcal{A}$ and $\mathcal{B}$ are (easily) computable in polynomial time.





# BPA MDPs with reachability objectives (3)

**Theorem 29**

*The set $\mathcal{W}^{=0}$ is effectively regular.*

- Let us consider two sets of stack symbols:
  - $\mathcal{A} = \{X \in \Gamma \mid \exists \sigma : \mathcal{P}(Reach(X \rightarrow T_\varepsilon)^\sigma) = 0\}$
  - $\mathcal{B} = \{X \in \Gamma \mid \exists \sigma : \mathcal{P}(Reach(X \rightarrow T)^\sigma) = 0\}$
- We have that $\mathcal{W}^{=0} = \mathcal{B}^* \cup \mathcal{B}^*\mathcal{A}\Gamma^*$.
- The pair $(\mathcal{A}, \mathcal{B})$ is the greatest fixed-point of a suitably defined $\Theta : 2^\Gamma \times 2^\Gamma \rightarrow 2^\Gamma \times 2^\Gamma$.

# BPA MDPs with reachability objectives (4)

**Theorem 30**

*The set $\mathcal{W}^{=1}$ is effectively regular.*

- Let us consider three sets of stack symbols:
  - $\mathcal{A} = \{X \in \Gamma \mid \exists\sigma : \mathcal{P}(Reach(X \rightarrow T)^\sigma) = 1\}$
  - $\mathcal{B} = \{X \in \Gamma \mid \exists\sigma : \mathcal{P}(Reach(X \rightarrow T_\varepsilon)^\sigma) = 1$ and $\mathcal{P}(Reach(X \rightarrow T)^\sigma) > 0\}$
  - $\mathcal{C} = \{X \in \Gamma \mid \exists\sigma : \mathcal{P}(Reach(X \rightarrow \varepsilon)^\sigma) = 1\}$
- We have that $\mathcal{W}^{=1} = (\mathcal{B} \cup \mathcal{C})^* \mathcal{A} \Gamma^*$.
- Moreover, the set $\mathcal{W}_\varepsilon^{=1} = \{\alpha \in \Gamma^* \mid \exists\sigma : \mathcal{P}(Reach(\alpha \rightarrow T_\varepsilon)^\sigma) = 1\}$ is equal to $(\mathcal{B} \cup \mathcal{C})^* \cup (\mathcal{B} \cup \mathcal{C})^* \mathcal{A} \Gamma^*$.
- The set $\mathcal{C}$ is computable in polynomial time by the results of Etessami & Yannakakis (STACS 2006).
- The sets $\mathcal{A}, \mathcal{B}$ are again computable as the greatest fixed-point of a suitably defined $\Theta : 2^\Gamma \times 2^\Gamma \rightarrow 2^\Gamma \times 2^\Gamma$.







# BPA MDPs with reachability objectives (5)

**Theorem 31**

*The set $\mathcal{W}^{<1}$ is effectively regular.*

- Let us consider two sets of stack symbols:
  - $\mathcal{A} = \{X \in \Gamma \mid \exists\sigma : \mathcal{P}(Reach(X \rightarrow \varepsilon)^\sigma) > 0\}$
  - $\mathcal{B} = \{X \in \Gamma \mid \exists\sigma : \mathcal{P}(Reach(X \rightarrow T_\varepsilon)^\sigma) < 1\}$
- We have that $\mathcal{W}^{<1} = \mathcal{A}^* \cup (\mathcal{A}^* \mathcal{B} \Gamma^*)$.
- The set $\mathcal{A}$ is computable by the previous results.
- The challenge is to compute the set $\mathcal{B}$. The membership $X \in \mathcal{B}$ is witnessed in two ways.
  - There is strategy $\sigma$ such that $\mathcal{P}(Reach(X \rightarrow \mathcal{W}_\varepsilon^{=0})^\sigma) > 0)$.
  - There is a "closed" family of stack symbols disjoint with $\mathcal{R}$ (this family forms a BPA MDP $\Delta'$) such that $X$ can be forced to enter $\Delta'$ with a positive probability, and there is a strategy $\sigma$ in $\Delta'$ such that $\mathcal{P}(Reach(X \rightarrow \varepsilon)^\sigma) < 1$.

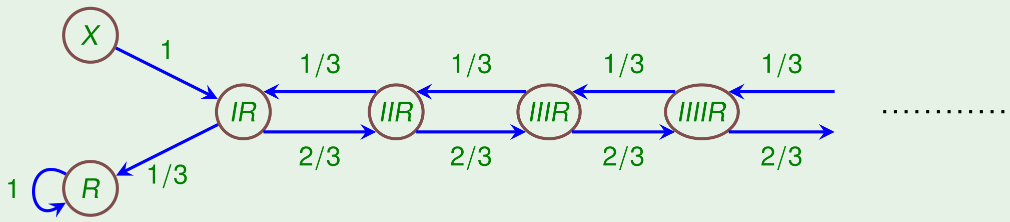
# BPA MDPs with reachability objectives (6)

## Example 32

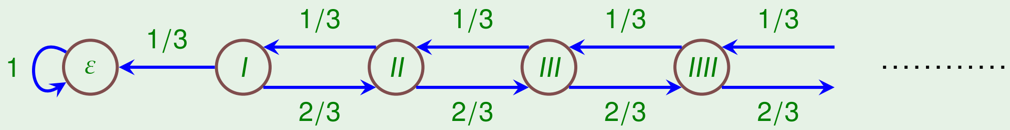
- Let $\Gamma = \{X, R, I\}$, where $\Gamma_\Box = \emptyset$, $\Gamma_\bigcirc = \{X, R, I\}$, and

$$X \overset{1}{\hookrightarrow} IR,\ R \overset{1}{\hookrightarrow} R,\ I \overset{2/3}{\hookrightarrow} II,\ I \overset{1/3}{\hookrightarrow} \varepsilon$$

- Let $\mathcal{R} = \{R\}$. Then $\mathcal{A} = \emptyset$, $\mathcal{B} = \{X, I\}$, hence $\mathcal{W}^{<1} = \{X, I\}\Gamma^*$.




- The membership $I \in \mathcal{B}$ is witnesses by a BPA MDP $\Delta'$ where $\Gamma' = \Gamma'_\bigcirc = \{I\}$ and $I \overset{2/3}{\hookrightarrow} II$, $I \overset{1/3}{\hookrightarrow} \varepsilon$.








# BPA games with termination objectives

## Theorem 33 (Etessami, Yannakakis, 2006)

*Let $\Delta = (\Gamma, \hookrightarrow, (\Gamma_\Box, \Gamma_\Diamond, \Gamma_\bigcirc), Prob)$ be a BPA game. For every $\alpha \in \Gamma^*$, let $val(\alpha) = val_{Reach(\varepsilon)}(\alpha)$. Then*

- *for all $\beta, \gamma \in \Gamma^*$ we have that $val(\beta\gamma) = val(\beta) \cdot val(\gamma)$;*
- *the tuple of all $val(X)$, where $X \in \Gamma$, forms the least solution of a system of recursive equations constructed as follows:*
  - *if $X \hookrightarrow \varepsilon$ and $X \in \Gamma_\Box$, we put $V_X = 1$;*
  - *otherwise, we put*
    - $V_X = \max_{X \hookrightarrow Y,\ X \hookrightarrow YZ} \{V_Y, V_Y \cdot V_Z\}$
    - $V_X = \min_{X \hookrightarrow Y,\ X \hookrightarrow YZ} \{V_Y, V_Y \cdot V_Z\}$
    - $V_X = \sum_{X \overset{p}{\hookrightarrow} \varepsilon} p + \sum_{X \overset{p}{\hookrightarrow} Y} p \cdot V_Y + \sum_{X \overset{p}{\hookrightarrow} YZ} p \cdot V_Y \cdot V_Z$

    *depending on whether $X \in \Gamma_\Box$, $X \in \Gamma_\Diamond$, or $X \in \Gamma_\bigcirc$, respectively.*
- *both players have optimal SMD strategies constructible in polynomial space;*
- *the problem whether $val(X) = 1$ and $val(X) \leq \varrho$ is in **NP** $\cap$ **coNP** and **PSPACE**, respectively.*

# BPA games with reachability objectives

**Theorem 34 (Brázdil, Brožek, K., Obdržálek, 2009)**

*Let $\Delta = (\Gamma, \hookrightarrow, (\Gamma_\Box, \Gamma_\Diamond, \Gamma_\bigcirc), Prob)$ be a BPA game, and $T \subseteq \Gamma^*$ a regular set of target configurations. For every $\alpha \in \Gamma^*$, let $val(\alpha) = val_{Reach(T)}(\alpha)$. Then*

- *The sets*
  - $\mathcal{V}^{=0} = \{v \in V \mid val(v) = 0\}$
  - $\mathcal{W}^{=1} = \{v \in V \mid val(v) = 1 \text{ and player } \Box \text{ has an optimal max. strategy}\}$

  *are regular. The associated finite-state automata are computable by a deterministic polynomial-time algorithm with* **NP** $\cap$ **coNP** *oracle.*
- *The membership to $\mathcal{V}^{=0}$ and $\mathcal{W}^{=1}$ is in* **NP** $\cap$ **coNP**.
- *Optimal strategies for both players are not necessarily SMD and it does not hold that $val(\beta\gamma) = val(\beta) \cdot val(\gamma)$.*

References:


- K. Etessami, M. Yannakakis. *Efficient Qualitative Analysis of Classes of Recursive Markov Decision Processes and Simple Stochastic Games.* Proc. STACS 2006, pp. 634–645, LNCS 3884, Springer 2006.
- T. Brázdil, V. Brožek, A. Kučera, and J. Obdržálek. *Qualitative Reachability in Stochastic BPA Games.* Proc. STACS 2009, pp. 207–218, 2009.






# A taxonomy of objectives in stoch. games

- Linear-time objectives.
  - Specified by Borel sets of runs in stochastic games.
  - Büchi, parity, Rabin, Street, Muller winning objectives.
  - In finite-state games, optimal strategies exist and are either memoryless (Büchi, parity) or require a finite memory.
- Long-run average objectives.
  - Specified as certain “limit” random variables defined over runs.
  - Mean payoff: $MP(w) = \lim_{n\to\infty} \frac{\sum_{i=0}^{n} rew(w(i))}{n}$
  - Discounted payoff: $DP(w) = \sum_{i=0}^{\infty} \lambda^i \cdot rew(w(i))$
  - The players aim at maximizing/minimizing the expected value of $MP$, $DP$.
- Multi-objectives (studied mainly for MDPs).
  - Specified by a vector of characteristics that are to be maximized/minimized. Generally, there is no “best” vector of values. The algorithms aim at computing/approximating the associated Pareto curve.
- Branching-time objectives.
  - Specified by formulae of branching-time logics that are interpreted over Markov chains (such as PCTL or PCTL*).

# Branching-time winning objectives

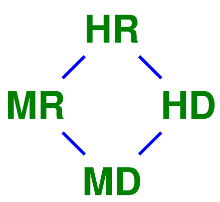
- Specified by formulae of branching-time logics that are interpreted over Markov chains (such as PCTL or PCTL*).
- $\mathcal{G}^{=1}(p \Rightarrow \mathcal{F}^{\geq 0.1} q)$
- The aim of player □ and player ◇ is to satisfy and falsify a given formula, respectively.
- Properties of stochastic games with branching-time objectives are quite different from the ones with linear-time objectives.



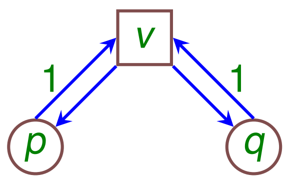
Stochastic games
Antonín Kučera
Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

# Properties of games with b.-t. objectives (I)

- Memory and randomization help:



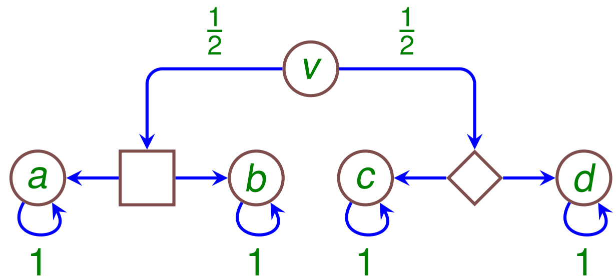

- Consider the following game:




- $\mathcal{X}^{=1} p \wedge \mathcal{F}^{=1} q$. Requires memory.
- $\mathcal{X}^{>0} p \wedge \mathcal{X}^{>0} q$. Requires randomization.
- $\mathcal{X}^{>0} p \wedge \mathcal{X}^{>0} q \wedge \mathcal{F}^{=1}\mathcal{G}^{=1} q$. Requires both memory and randomization.
- In some cases, infinite memory is required.

# Properties of games with b.-t. objectives (II)

- The games are not determined (for any strategy type).
- $\mathcal{F}^{=1}(a \vee c) \;\vee\; \mathcal{F}^{=1}(b \vee d) \;\vee\; \left(\mathcal{F}^{>0}c \wedge \mathcal{F}^{>0}d\right)$








# Who wins the game (MD strategies) ?

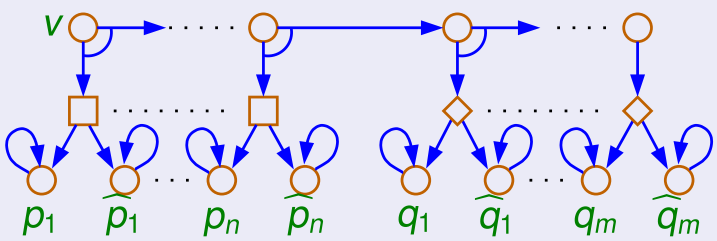
**Theorem 35 (Brázdil, Brožek, Forejt, K., 2006)**

*The existence of a winning MD strategy for player $\Box$ is $\Sigma_2 = \mathbf{NP}^{\mathbf{NP}}$ complete.*

**Proof.**

The membership to $\Sigma_2$ follows easily. The $\Sigma_2$-hardness can be established as follows:

- Let $\exists x_1, \cdots, x_n \forall y_1, \cdots, y_m\, B$ be a $\Sigma_2$ formula.
- Consider the following game:




- Let $\varphi$ be the PCTL formula obtained from $B$ by substituting each occurrence of $x_i$, $\neg x_i$, $y_j$, and $\neg y_j$ with $\mathcal{F}^{>0}p_i$, $\mathcal{F}^{>0}\widehat{p}_i$, $\mathcal{F}^{>0}q_j$, and $\mathcal{F}^{>0}\widehat{q}_j$, respectively. $\Box$

# Who wins the game (MR strategies) ?

**Theorem 36 (Brázdil, Brožek, Forejt, K., 2006)**

*The existence of a winning MR strategy for player $\Box$ is $\Sigma_2$-hard and in **EXPTIME**. For the qualitative fragment of PCTL, the problem is $\Sigma_2$-complete.*

**Proof.**

- The $\Sigma_2$-hardness is established similarly as for MD strategies.
- The membership to **EXPTIME** is obtained by encoding the condition into Tarski algebra.
- The membership to $\Sigma_2$ for the qualitative PCTL follows easily.

$\Box$



Stochastic games
Antonín Kučera

Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

# Who wins the game (HD, HR, FD, FR) ?

**Theorem 37 (Brázdil, Brožek, Forejt, K., 2006)**

*The existence of a winning HD (or HR) strategy for player $\Box$ in MDPs is highly undecidable (and $\Sigma_1^1$-complete). Moreover, the existence of a winning FD (or FR) strategy is also undecidable.*

- The result holds for the $\mathcal{L}(\mathcal{F}^{=1/2}, \mathcal{F}^{=1}, \mathcal{F}^{>0}, \mathcal{G}^{=1})$ fragment of PCTL (the role of $\mathcal{F}^{=1/2}$ is crucial).
- The proof is obtained by reduction of the problem whether a given non-deterministic Minsky machine has an infinite recurrent computation.

# The undecidability proof

- A non-deterministic Minsky machine $\mathcal{M}$ with two counters $c_1, c_2$:

$$1 : ins_1, \cdots, n : ins_n$$

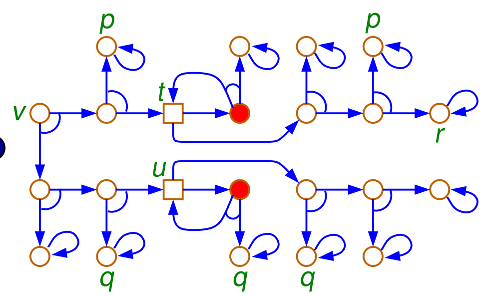
  where each $ins_i$ takes one of the following forms:
  - $c_j := c_j + 1;\ goto\ k$
  - $if\ c_j{=}0\ then\ goto\ k\ else\ c_j := c_j - 1;\ goto\ m$
  - $goto\ \{k\ or\ m\}$
- The problem whether a given non-deterministic Minsky machine with two counters initialized to zero has an infinite computation that executes $ins_1$ infinitely often is $\Sigma_1^1$-complete.
- For a given machine $\mathcal{M}$, we construct a finite-state MDP $G(\mathcal{M})$ and a formula $\varphi \in \mathcal{L}(\mathcal{F}^{=1/2}, \mathcal{F}^{=1}, \mathcal{F}^{>0}, \mathcal{G}^{=1})$ such that $\mathcal{M}$ has an infinite recurrent computation iff player $\Box$ has a winning HD (or HR) strategy for $\varphi$ in a distinguished vertex $v$ of $G(\mathcal{M})$.

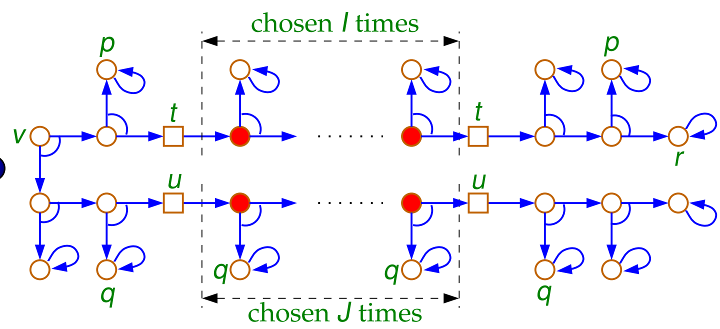


Stochastic games
Antonín Kučera

Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

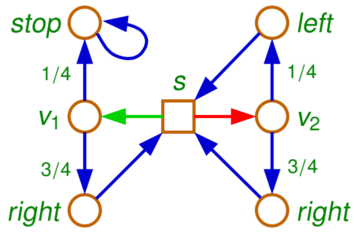
# The construction of $G(\mathcal{M})$ and $\varphi$

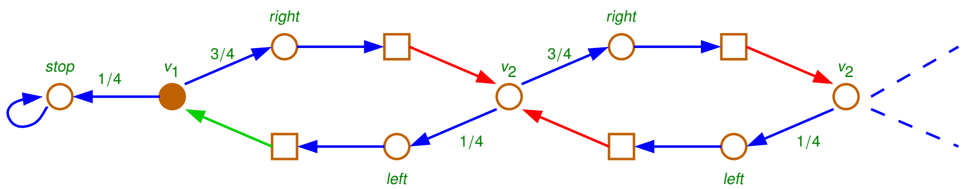
- 


- 


- $I = J < \omega$ iff $v \models \mathcal{F}^{>0} r \wedge \mathcal{F}^{=1/2}(p \vee q)$
- The probability of $\mathcal{F}(p \vee q)$: $0.01\underbrace{0\cdots0}_{I}01 + 0.001\underbrace{1\cdots1}_{J}1$

# Positive results (1)

- We restrict ourselves to qualitative fragments of probabilistic branching time logics.
- Even MDPs with qualitative PCTL objectives may require infinite memory.
- 


  $\mathcal{G}^{>0}(\neg stop \wedge \mathcal{F}^{>0} stop)$
  $\wedge\ \mathcal{G}^{=1}(s \Rightarrow (X^{=1} v_1 \vee X^{=1} v_2))$
- A winning strategy: if $\#left < \#right$ use the red transition, otherwise use the green one.
- 




Stochastic games
Antonín Kučera

Preliminaries
Reachability objectives
The value
Min strategies
Max strategies
Finite-state games
BPA games
Taxonomy of objectives
Branching-time objectives
Basic properties
Deciding the winner

# Positive result (2)

### Theorem 38 (Brázdil, Forejt, K., 2008)

- *The existence of a winning HD (or HR) strategy for player □ in MDPs with qualitative PECTL* objectives is decidable in time which is polynomial in the size of MDP and doubly exponential in the size of the formula. The problem is* **2-EXPTIME***-hard.*
- *Moreover, iff there is a winning HD (or HR) strategy, there is also a one-counter winning strategy and one can effectively construct a one-counter automaton which implements this strategy (the associated complexity bounds are the same as above).*

References:

- T. Brázdil, V. Brožek, V. Forejt, and A. Kučera. *Stochastic Games with Branching-Time Winning Objectives.* Proc. of LICS 2006, pp. 349-358, 2006.
- T. Brázdil, V. Forejt, and A. Kučera. *Controller Synthesis and Verification for Markov Decision Processes with Qualitative Branching Time Objectives.* Proc. of ICALP 2008, pp. 148-159, volume 5126 of LNCS. Springer, 2008.