Methods and Tools for Solving Infinite-State Stochastic Games

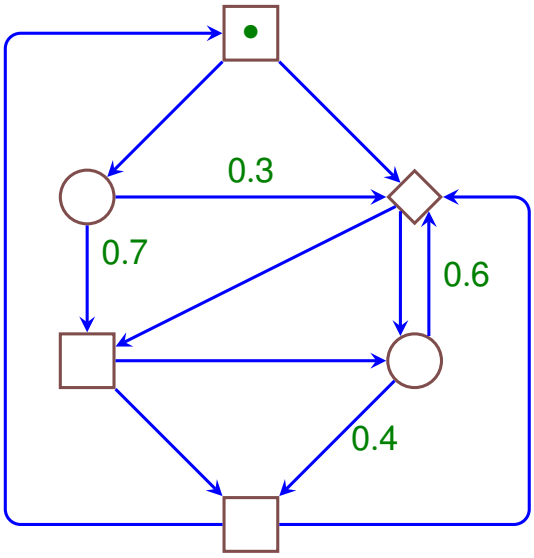
Antonín Kučera

Brussels, September 2016

Outline

- Perfect information turn-based stochastic games
- Linear-time objectives
- Branching-time objectives
  - Basic properties
  - Applicability to AI problems (patrolling)
- Infinite-state stochastic games
- Reachability objectives; differences from finite-state games
- One-counter stochastic games
  - The use of martingales

Turn-Based Stochastic Games



- $\mathcal{G} = (V, (V_\Box, V_\Diamond, V_\bigcirc), E, v_0)$
- Strategies:
  - $\sigma : V^* V_\Box \to \mathcal{D}(V)$
  - $\pi : V^* V_\Diamond \to \mathcal{D}(V)$
- σ, π determine
  - A Markov chain $\mathcal{M}_\mathcal{G}$
  - A prob. measure $\mathcal{P}^{\sigma,\pi}$ over $Run(v_0)$

Win-lose objectives

- Let *good* be a Borel set of runs.
- The aim of Player □/◇ is to maximize/minimize $\mathcal{P}(good)$.
- Examples:
  - reachability/safety
  - Büchi, co-Büchi
  - Muller, Street, Rabin

Linear-time objectives

- Let f be a Borel measurable function assigning a real-valued payoff to every run.
- The aim of Player $\Box/\Diamond$ is to maximize/minimize $\mathbb{E}[f]$.
- Examples:
  - mean-payoff
  - discounted accumulated reward
  - accumulated reward

The existence of the value

- Player $\Box$ (the maximizer) value:
  - For a given σ, we put $val^{\sigma}_{\Box}(v) = \inf_{\pi} \mathbb{E}^{\sigma,\pi}_{v}[f]$
  - $val_{\Box}(v) = \sup_{\sigma} val^{\sigma}_{\Box}(v)$
- Player $\Diamond$ (the minimizer) value:
  - For a given π, we put $val^{\pi}_{\Diamond}(v) = \sup_{\sigma} \mathbb{E}^{\sigma,\pi}_{v}[f]$
  - $val_{\Diamond}(v) = \inf_{\pi} val^{\pi}_{\Diamond}(v)$
- $val_{\Box}(v) \leq val_{\Diamond}(v)$

Theorem 1 (Martin 1998; Maitra & Sudderth 1998)

If f is Borel and bounded, then $val_{\Box}(v) = val_{\Diamond}(v)$.

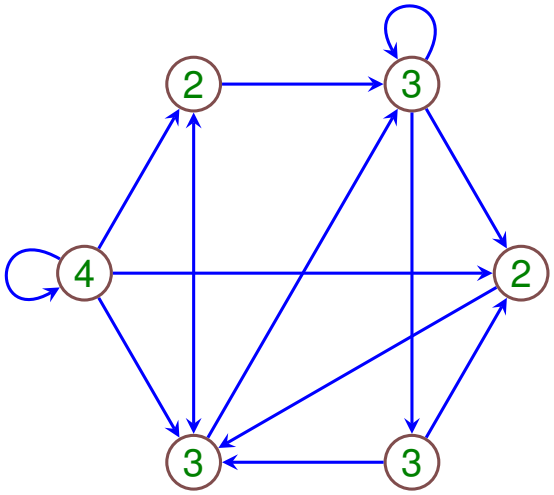
The problems of interest

- Compute/approximate the value of a given vertex.
- Do optimal strategies exist?
- What is the “strategy complexity” of (sub)optimal strategies?
- Compute (sub)optimal strategies.

Branching-time objectives

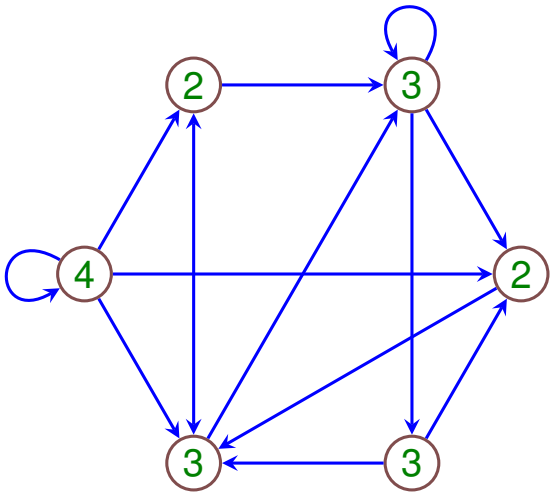
- Let φ be a PCTL or PCTL* formula.
- The aim of Player $\Box/\Diamond$ is to satisfy/falsify φ.
- The winning strategies are defined for both players in the natural way.

Patrolling



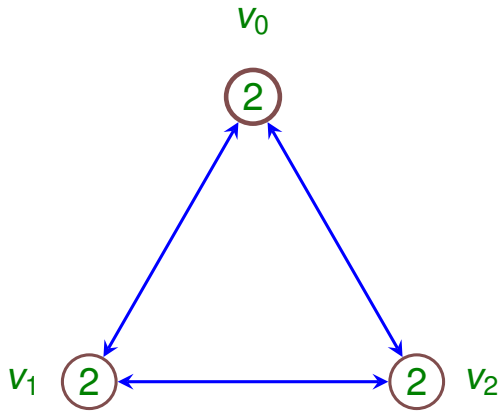
- Defender's strategy: $\sigma : V^+ \to \mathcal{D}(V)$
- Attacker's strategy: $\pi : V^+ \to V \cup \{*\}$ (must be "prefix free")
- $\mathcal{P}^{\sigma,\pi}(DRuns)$
- $val = \sup_\sigma \inf_\pi \mathcal{P}^{\sigma,\pi}(Druns)$

Patrolling (2)



- $Defend(v) \equiv \mathcal{F}_{\langle 1,t(v)\rangle}(v)$
- $\mathcal{G}^{=1} \bigwedge_{v \in V} Defend^{\geq \varrho}(v)$

Patrolling (3)



- $\mathcal{G}^{=1}\left(\mathcal{F}^{\geq\varrho}_{\langle 1,2\rangle}(v_0) \wedge \mathcal{F}^{\geq\varrho}_{\langle 1,2\rangle}(v_1) \wedge \mathcal{F}^{\geq\varrho}_{\langle 1,2\rangle}(v_2)\right)$
- A "uniform" strategy yields $\varrho = 1/2$.
- There is an optimal strategy where $\varrho = (\sqrt{5}-1)/2 = 0.618\ldots$

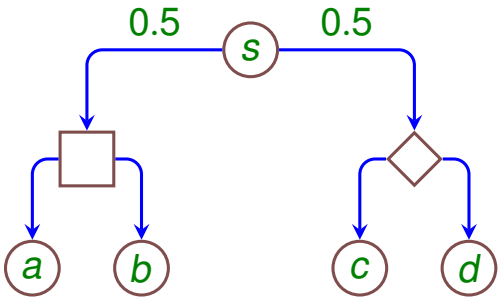
Patrolling (4)

Some recommended reading about patrolling:

- N. Basilico, N. Gatti, and F. Amigoni. *Leader-follower strategies for robotic patrolling in environments with arbitrary topologies.* In Proc. AAMAS 2009, pages 57–64, 2009.
- M. Tambe. *Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned.* Cambridge University Press, 2011.
- Recent results: Proceedings of AAAI, AAMAS, IJCAI.

Properties of branching-time objectives

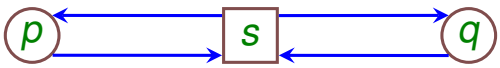
Are games with PCTL/PCTL* objectives determined?



$$\mathcal{F}^{=1}(a \vee c) \ \vee \ \mathcal{F}^{=1}(b \vee d) \ \vee \ \left(\mathcal{F}^{>0} c \wedge \mathcal{F}^{>0} d\right)$$

Properties of branching-time objectives (2)

Does memory/randomization help?



- $\mathcal{X}^{=1} p \wedge \mathcal{F}^{=1} q$. Requires history.
- $\mathcal{X}^{>0} p \wedge \mathcal{X}^{>0} q$. Requires randomization.
- $\mathcal{X}^{>0} p \ \wedge \ \mathcal{X}^{>0} q \ \wedge \ \mathcal{F}^{=1} \mathcal{G}^{=1} (s \vee q)$. Requires history and randomization.
- $\mathcal{G}^{=1} \mathcal{F}^{>0} p \ \wedge \ \mathcal{F}^{<1} p$. Requires infinite memory.

Properties of branching-time objectives (3)

Who wins?

Theorem 2 (Brázdil, Brožek, Forejt, K.; LICS 2008)

The existence of a winning strategy for player □ is

- $\Sigma_2 = \mathbf{NP^{NP}}$ *complete for MD strategies.*
- Σ_2*-hard and in* ***EXPTIME*** *for MR strategies.*
- *undecidable for finite-memory strategies.*
- *highly undecidable* (Σ_1^1 *complete) for HD and HR strategies.*

- The undecidability results hold even for MDPs and 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.

Properties of branching-time objectives (4)

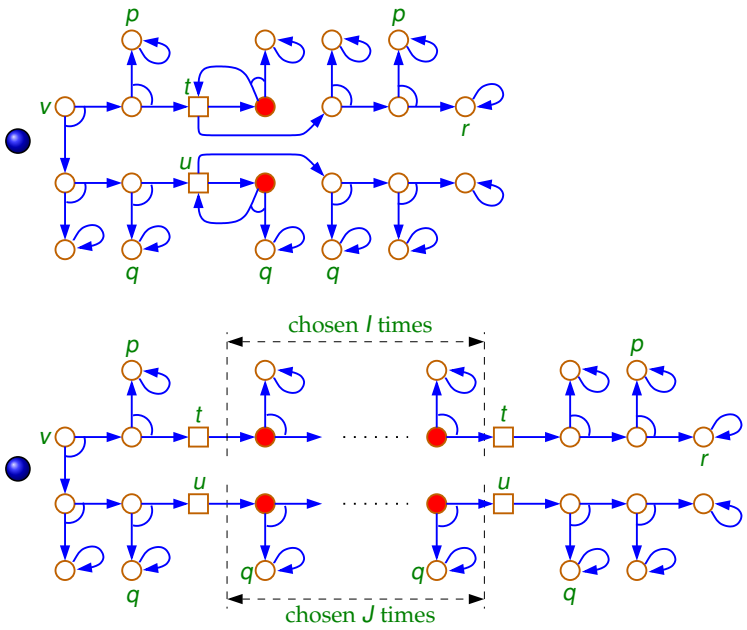
- A non-deterministic Minsky machine $\mathcal{M}$ with two counters c_1, c_2:

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

  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 Σ_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 □ has a winning HD (or HR) strategy for a distinguished vertex v of $G(\mathcal{M})$.

Properties of branching-time objectives (5)



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

Properties of branching-time objectives (6)

Theorem 3 (Brázdil, Forejt, K.; Icalp 2008)

Let $\mathcal{M}$ be a finite MDP. The existence of a winning strategy for player $\Box$ is

- ***EXPTIME**-complete for qualitative PCTL formulae.*
- *2-**EXPTIME**-complete for qualitative PECTL* formulae.*

The complexity stays polynomial in the size of $\mathcal{M}$. The controller (winning strategy) can always be implemented by an effectively constructible one-counter automaton.

Properties of branching-time objectives (7)

Some open problems

- How about qualitative branching-time objectives in (two-player) stochastic games?
- Are there any decidable PCTL fragments involving quantitative connectives?
- What are appropriate logics for AI problems?

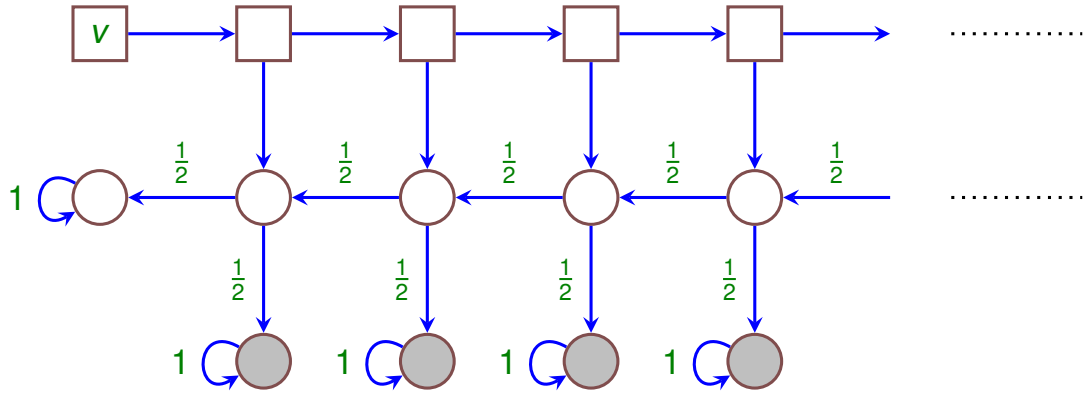
Stochastic games with reachability objectives

- The aim of Player $\Box$/$\Diamond$ is to maximize/minimize the probability of reaching a target vertex.
- Recall $\sup_{\sigma} \inf_{\pi} \mathcal{P}^{\sigma,\pi}(\mathit{Reach}) = \inf_{\pi} \sup_{\sigma} \mathcal{P}^{\sigma,\pi}(\mathit{Reach})$
- In finite-state stochastic games, both players have optimal MD strategies (and the value is rational).

$$\sup_{\sigma \in MD} \inf_{\pi \in MD} \mathcal{P}^{\sigma,\pi}(\mathit{Reach}) = \inf_{\pi \in MD} \sup_{\sigma \in MD} \mathcal{P}^{\sigma,\pi}(\mathit{Reach})$$

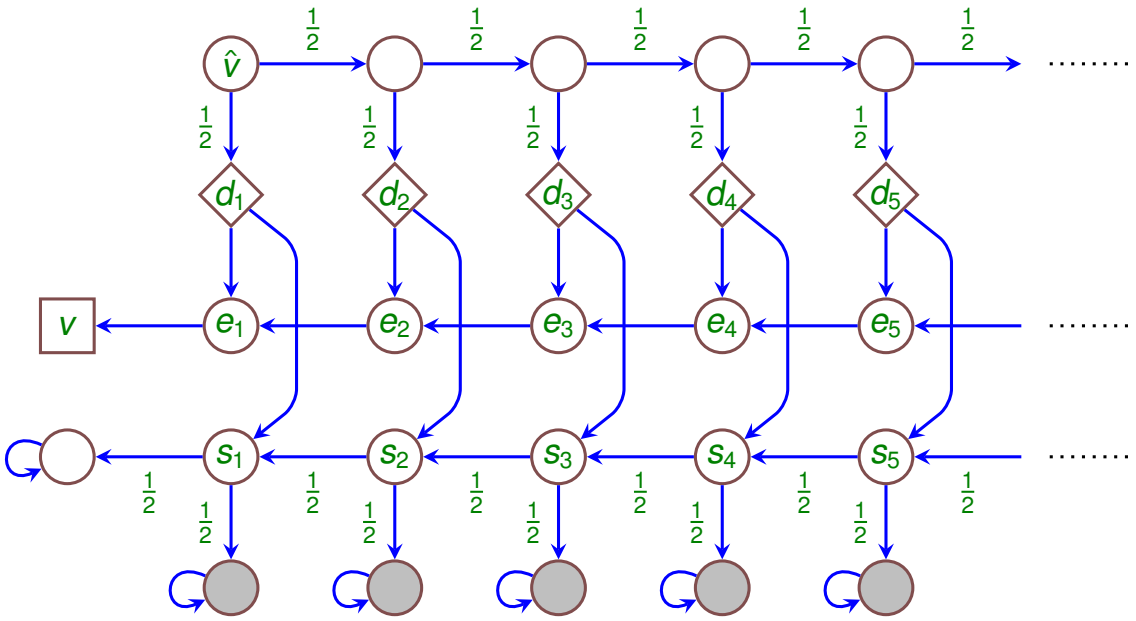
Counterexamples (1)

Player □ (Max) does not necessarily have an optimal strategy, even in finitely-branching MDPs.



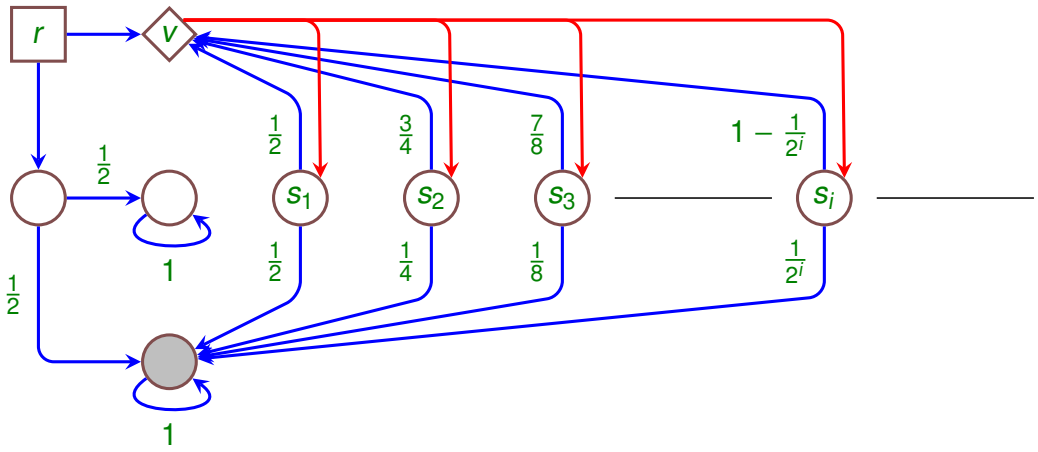
Counterexamples (2)

An optimal strategy for Player □ (Max) may require infinite memory, even in finitely-branching games.



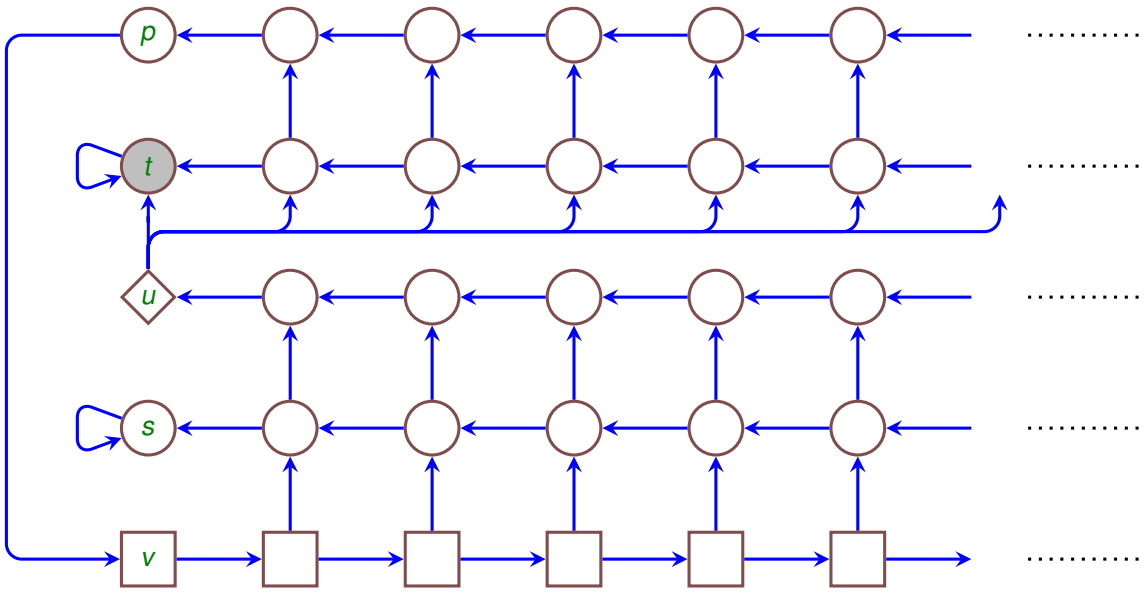
Counterexamples (3)

Optimal minimizing strategies do not necessarily exist, and optimal minimizing strategies may require infinite memory.



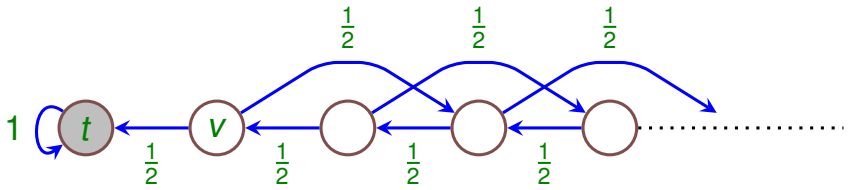
Counterexamples (4)

Infinitely-branching games are not determined for finite-memory strategies.



- $\sup_{\sigma \in MD} \inf_{\pi \in MD} \mathcal{P}^{\sigma,\pi}(Reach) = 0$
- $\inf_{\pi \in MD} \sup_{\sigma \in MD} \mathcal{P}^{\sigma,\pi}(Reach) = 1$

The value can be irrational

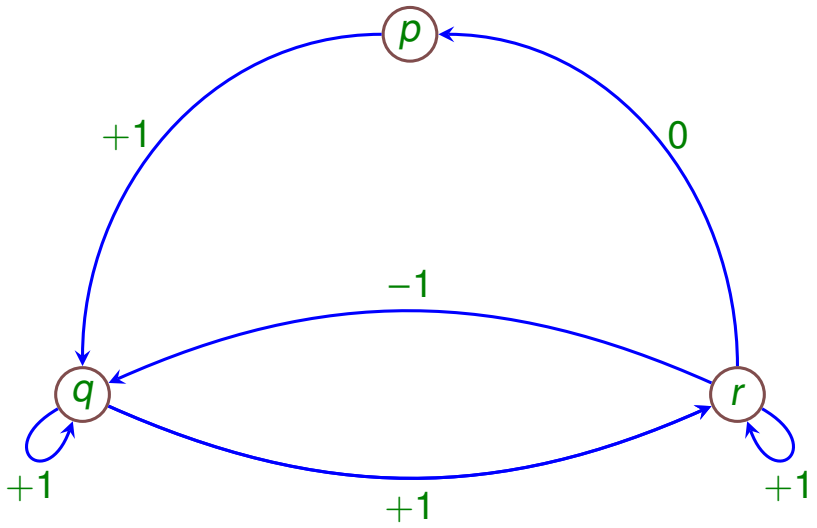


- $val(v)$ is the least solution of $x = \frac{1}{2} + \frac{1}{2}x^3$ in $[0, 1]$, i.e., $\frac{\sqrt{5}-1}{2}$
- v satisfies $\mathcal{G}^{=1}\ \mathcal{F}^{>0}\ t$ but not $\mathcal{F}^{=1}\ t$.

Infinite-state stochastic games

- Stochastic-game extensions of well-known automata classes
  - One-counter automata.
  - Pushdown automata.
  - Multi-counter automata, VASS
  - Lossy channel systems.
- The standard workflow:
  - Understand the existing results about the “standard” non-deterministic model.
  - Consider the fully probabilistic variant of the model.
  - Go on with Markov decision processes (both maximizing and minimizing).
  - Finally, solve the stochastic game extension of the model.

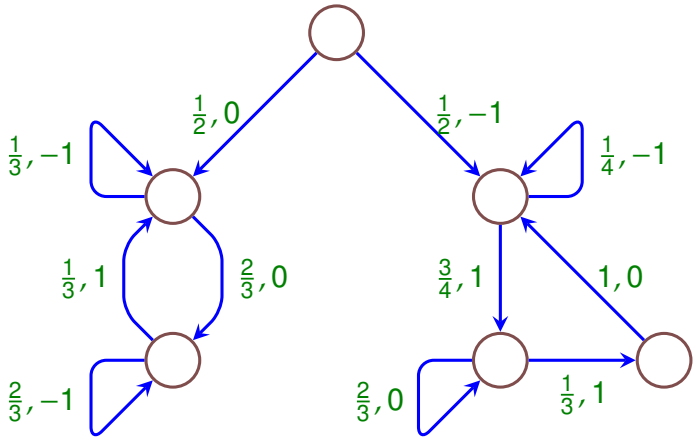
One-counter automata



One-counter automata: the objectives

- Zero reachability: The aim of Player □/◇ is to maximize/minimize the probability of reaching a configuration with zero counter.
- Termination time: The aim of Player □/◇ is to maximize/minimize the expected number of transitions performed before visiting a configuration with zero counter.

One-counter automata: zero reachability objective



- $change(s) = \sum_{s \xrightarrow{x,\delta} t} x \cdot \delta$
- $trend(C) = \sum_{s \in C} \mu(s) \cdot change(s)$

One-counter automata: zero reachability objective (2)

Theorem 4 (Brázdil, Brožek, Etessami, K., Wojtczak; SODA'10)

*Let $\mathcal{G}$ be a maximizing OC MDP, $T = \{p(0) \mid p \in Q\}$. The problem whether $val(q(i)) = 1$ is in **P**. If $val(q(i)) = 1$, then player $\Box$ has an optimal strategy computable in polynomial time.*

Theorem 5 (Brázdil, Brožek, Etessami; FST& TCS'10)

*Let $\mathcal{G}$ be a stochastic OC game, $T = \{p(0) \mid p \in Q\}$. The problem whether $val(q(i)) = 1$ is in **NP** $\cap$ **coNP**. If $val(q(i)) = 1$, then player $\Box$ has an optimal strategy computable in polynomial time using **NP** $\cap$ **coNP** oracle.*

One-counter automata: zero reachability objective (3)

Theorem 6 (Brázdil, Brožek, Etessami, K.; Icalp 2011)

Let $\mathcal{G}$ be a stochastic OC game, $T = \{p(0) \mid p \in Q\}$. Then $val(q(i))$ can be effectively approximated up to an arbitrarily small additive error $\varepsilon > 0$, and ε-optimal strategies for both players are effectively computable. The algorithms run in non-deterministic time which is exponential in the size of $\mathcal{G}$ an polynomial in $\log(1/\varepsilon)$ and $\log(i)$.

One-counter automata: termination time objective

- The aim of Player $\Box/\Diamond$ is to maximize/minimize the expected number of transitions performed before visiting a configuration with zero counter.
- We start by developing a method for computing $E[p{\downarrow}q]$ in fully probabilistic (and strongly connected) one-counter automaton.
- The family of all conditional expectations of the form $E[p{\downarrow}q]$ is the (unique) solution of an effectively constructible system of linear equations. Let $[r{\downarrow}t]$ be the probability $\mathcal{P}(r(1) \rightarrow^*_{>0} t(0))$. Then

$$E(p{\downarrow}q) = \sum_{p \xrightarrow{x,-1} q} \frac{x}{[p{\downarrow}q]} + \sum_{p \xrightarrow{x,0} t} \frac{x \cdot [t{\downarrow}q]}{[p{\downarrow}q]} \cdot \Big(1 + E(t{\downarrow}q)\Big)$$

$$+ \sum_{p \xrightarrow{x,1} t} \sum_{r \in Q} \frac{x \cdot [t{\downarrow}r] \cdot [r{\downarrow}q]}{[p{\downarrow}q]} \cdot \Big(1 + E(t{\downarrow}r) + E(r{\downarrow}q)\Big)$$

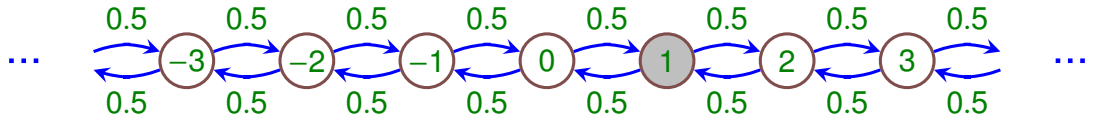
One-counter automata: termination time objective (2)

- Idea: approximate the coefficients up to a sufficiently small precision $\varepsilon > 0$ and solve the approximated linear system.
- The main problem: Determine a sufficient precision ε.
- Key technical ingredient: Obtain an upper bound on $E[p{\downarrow}q]$.
- $E[p{\downarrow}q] = \frac{1}{[p{\downarrow}q]} \sum_{i=1}^{\infty} [p{\downarrow}q, i] \cdot i$.
- We show that if $t \neq 0$, then $[p{\downarrow}q, i] \leq a^i$, where $a < 1$ can be effectively bounded.
- This is achieved by designing an appropriate martingale and applying Azuma's inequality.

Martingales

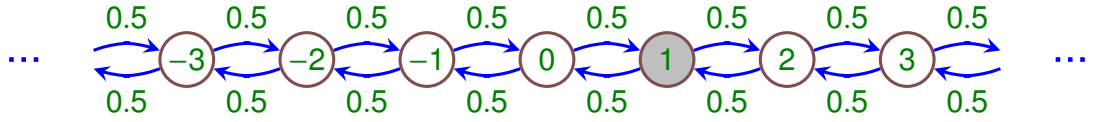
- A martingale is a stochastic process $m^{(0)}, m^{(1)}, m^{(2)} \ldots$ such that
  - $\mathbb{E}(|m^{(i)}|) < \infty$ for all $i \geq 0$;
  - $\mathbb{E}(m^{(i+1)} \mid m^{(0)}, \ldots, m^{(i)}) = m^{(i)}$ almost surely.
- We have that $\mathbb{E}(m^{(i)}) = \mathbb{E}(m^0)$.
- Assume $|m^{(i)} - m^{(i-1)}| \leq c$ for some fixed $c \in \mathbb{N}$. Then
  - Let $X : \Omega \to \mathbb{N}_0$ be a stopping time. Then $\mathbb{E}(m^{(X)}) = \mathbb{E}(m^{(0)})$.
  - $\mathcal{P}(|m^{(i)} - m^{(0)}| \geq t) \leq 2 \exp\left(\frac{-t^2}{i \cdot c^2}\right)$

Martingales, first example.



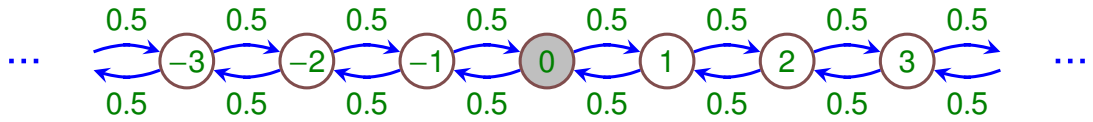
- $m^{(i)} : \mathit{Run}(1) \to \mathbb{N}_0$ assigns to each $w \in \mathit{Run}(1)$ the value $w(i)$.
- $\mathbb{E}(m^{i+1} \mid m^{(i)}{=}k) \;=\; \frac{1}{2}(k+1) + \frac{1}{2}(k-1) = k$.

Martingales, first example (2).



- What is the probability p of visiting the state $k > 1$ without visiting the state 0 before?
- For every $w \in \mathit{Run}(1)$, let $X(w)$ be the least index ℓ such that $w(\ell) = 0$ or $w(\ell) = k$.
- By optional stopping theorem, $\mathbb{E}(m^{(X)}) = 1$. Now observe $\mathbb{E}(m^{(X)}) = p \cdot k + (1-p) \cdot 0$. Hence, $p = 1/k$.

Martingales, first example (3).

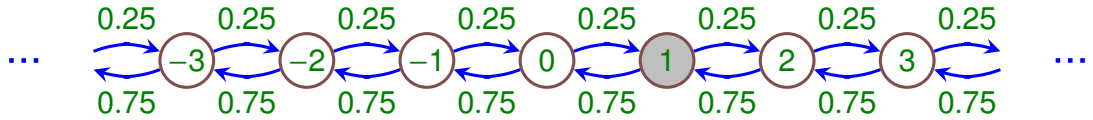


- What is the probability p that after performing $2k$ steps we end up outside the interval $(-k, k)$?
- Clearly, $|m^{(i)} - m^{(i-1)}| \leq 1$
- Azuma's inequality:

$$\mathcal{P}(|m^{(2k)} - m^{(0)}| \geq k) \;\leq\; 2\exp\left(\frac{-k^2}{2k}\right) \;=\; 2a^k$$

where $a = \exp(\frac{-1}{2}) < 1$.

Martingales, second example.



- $m^{(i)} : Run(1) \to \mathbb{N}_0$ assigns to each $w \in Run(1)$ the value $w(i)$.
- $\mathbb{E}(m^{i+1} \mid m^{(i)}{=}k) \;=\; \frac{3}{4}(k-1) + \frac{1}{4}(k+1) = k - \frac{1}{2} \neq k$.

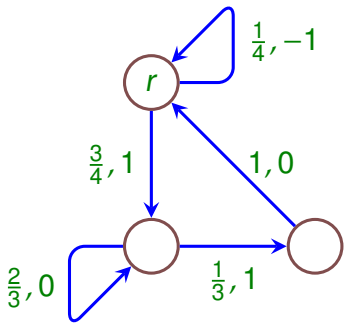
A slight modification helps:

- $m^{(i)} : Run(1) \to \mathbb{N}_0$ assigns to each $w \in Run(1)$ the value $w(i) + i \cdot \frac{1}{2}$.
- $\mathbb{E}(m^{i+1} \mid m^{(i)}{=}k) \;=\; \frac{3}{4}(k - \frac{i}{2} - 1 + \frac{i+1}{2}) + \frac{1}{4}(k - \frac{i}{2} + 1 + \frac{i+1}{2}) = k$

How can we bound $[p{\downarrow}p, i]$ from above?

- If we enter $p(0)$ from $p(1)$ after exactly i steps, we have that $m^{(i)} = i \cdot \frac{1}{2}$
- Hence, $[p{\downarrow}p, i] \;\leq\; \mathcal{P}(m^{(i)} = i \cdot \frac{1}{2}) \;\leq\; \mathcal{P}(|m^{(i)} - m^{(0)}| \geq i \cdot \frac{1}{2} - 1) \;\leq\; a^i$.

Martingales, third example.



- If $change(s) = trend(C)$ for every $s \in C$, we can setup a martingale over $Run(r(\ell))$ as follows:

$$m(i) = counter^{(i)} - i \cdot trend(C)$$

- Otherwise, the difference among the individual control states can be "compensated" by suitable constants v_s, where $s \in C$.

$$m(i) = counter^{(i)} - i \cdot trend(C) + v_{s^{(i)}}$$

Martingales, other examples.

- For details, see *Efficient Analysis of Probabilistic Programs with an Unbounded Counter*, by Brázdil, Kiefer, K. JACM. 61(6), 2014.
- Other useful martingales were defined for
  - Probabilistic BPA and PDA systems.
    - pPDA dichotomy: the expected termination time is either infinite, or it is finite and the probability of terminating runs of length n decays exponentially in n.
  - Two-counter probabilistic VASS
    - the limit-average properties of runs in (strongly connected) pVASS may assume more than one value and may even be undefined with probability 1. This contradicts the "classical" results about stochastic Petri nets established in the 80s.

Some directions for future research.

- Improve the understanding of non-deterministic and fully stochastic VASS.
  - reachability for VASS;
  - revisit the ergodicity questions for probabilistic VASS considered in the 80s;
  - simplify the existing proofs.
- Extend the results achieved for turn-based games to concurrent games.