Model checking PCTL on MDPs

David N. Jansen

Let's play a better game

Recapitulation

- Markov decision processes combine nondeterministic and probabilistic choices.
- Strategies select how to resolve nondeterministic choices.
- MDP + strategy induce a Markov chain.
- Probabilities on a MDP are defined via the induced Markov chain.

Formal definition

- A Markov decision process consists of:
  - S — finite set of states (often $S = \{1, 2, \ldots n\}$)
  - Act — finite set of actions
  - $\mathbf{P}: S \times Act \times S \rightarrow [0,1]$ — transition probability matrix for every action
  - $L: S \rightarrow 2^{AP}$ — labelling with atomic propositions

Probabilistic CTL

- a logic to describe properties
  of Markov chains and MDPs
- extends CTL
- strictly speaking, not a quantitative logic
  (truth values are Boolean: true or false)

PCTL syntax for MDPs

- state formulas ϕ, ψ
  - a atomic proposition
  - $\neg\varphi$ negation
  - $\varphi \vee \psi$ disjunction
  - $\mathbf{P}_{\leq p}(\mathrm{E}\Pi)$, $\mathbf{P}_{\geq p}(\mathrm{E}\Pi)$, $\mathbf{P}_{\leq p}(\mathrm{A}\Pi)$, $\mathbf{P}_{\geq p}(\mathrm{A}\Pi)$,... probabilistic operator
- path formulas Π
  - $\mathrm{X}\ \phi$ next state
  - $\varphi\ \mathrm{U}\ \psi$ unbounded until
  - $\varphi\ \mathrm{U}^{\leq k}\ \psi$ bounded until

What strategy to choose?

- the best strategy: highest probability to win
- the worst strategy: lowest probability to win

- add operators A, E similar to CTL
  - A = for all strategies
  - E = for some strategy

- often simplified to: implicitly A only

Model checking PCTL for MDPs

- find strategy and probabilities
  at the same time
- three methods:
  - value iteration (similar to Υ)
  - linear programming (inequality system)
  - policy iteration
    (pick a random action and then improve)

What strategy is needed?

- history-dependent strategy? simpler strategy?
- unbounded until: history-independent
- bounded until: step-dependent
- Proofs on the following slides...

Today's programme

- algorithm for model checking
  - general: bottom-up
  - bounded until
    (needs a step-dependent scheduler)
  - unbounded until (history-independent scheduler)
- examples

Model checking PCTL formulas

set of states that satisfy ϕ

Calculate Sat(ϕ) from subformulas:

- Sat(a) = $\{s \in S \mid a \in L(s)\}$
- Sat(¬φ) = $S \setminus$ Sat(φ)
- Sat(φ∨ψ) = Sat(φ) ∪ Sat(ψ)

- Sat($\mathbf{P}_{\geq p}$(AΠ)) = $\{s \mid \inf_{\eta \in strat} \mathrm{Prob}^{\eta}{}_{s}\{\sigma \mid \sigma \vDash \Pi\} \geq p\}$
- Sat($\mathbf{P}_{\geq p}$(EΠ)) = $\{s \mid \sup_{\eta \in strat} \mathrm{Prob}^{\eta}{}_{s}\{\sigma \mid \sigma \vDash \Pi\} \geq p\}$
- Sat($\mathbf{P}_{\leq p}$(AΠ)) = $\{s \mid \sup_{\eta \in strat} \mathrm{Prob}^{\eta}{}_{s}\{\sigma \mid \sigma \vDash \Pi\} \leq p\}$
- Sat($\mathbf{P}_{\leq p}$(EΠ)) = $\{s \mid \inf_{\eta \in strat} \mathrm{Prob}^{\eta}{}_{s}\{\sigma \mid \sigma \vDash \Pi\} \leq p\}$

notatie voor schedulers...

- griekse letter, bv. η
- Frakturletter, bv. $\mathfrak{A}$, $\mathfrak{S}$

Path formula: X (next)

$$\sup_{\eta \in strat} \text{Prob}^{\eta}_{s}\{\sigma \mid \sigma \models X\phi\}$$

$$= \max_{\alpha \in Act} \mathbf{P}(s, \alpha, \text{Sat}(\phi))$$

$$= \max_{\alpha \in Act} \sum_{t \in \text{Sat}(\phi)} \mathbf{P}(s, \alpha, t)$$

Path formula: $U^{\leq n}$ (bounded until)

- $\varphi \; U^{\leq k} \; \psi$
- Find
  - S_1 = winning states (Sat(ψ))
  - S_0 = losing states (may add stalemate states) (S \ Sat(ϕ) \ Sat(ψ))
  - $S_?$ = other states (including announce checkmate)

- Which strategy to choose?

Step-bounded strategies

- A step-bounded strategy bases its decision on:
  - current state
  - number of steps taken until now
  - no more information on past states!

- Theorem: **Step-bounded strategies suffice to find the best strategy for bounded until.**

  (Proof idea: number of steps that remain before deadline is the only relevant data.)

Proof: step-bounded strategies suffice

- proof by induction on number of remaining steps
  - base case: 0 steps remain.
    No transition allowed → class of scheduler of no influence
  - induction step: Assume given strategies η_s for each state s that are optimal for $\varphi\ U^{\leq n}\ \psi$.
    To find an optimal strategy from s_0 for $\varphi\ U^{\leq n+1}\ \psi$, choose an action
    that leads to the highest probability under η_s.

Bounded reachability

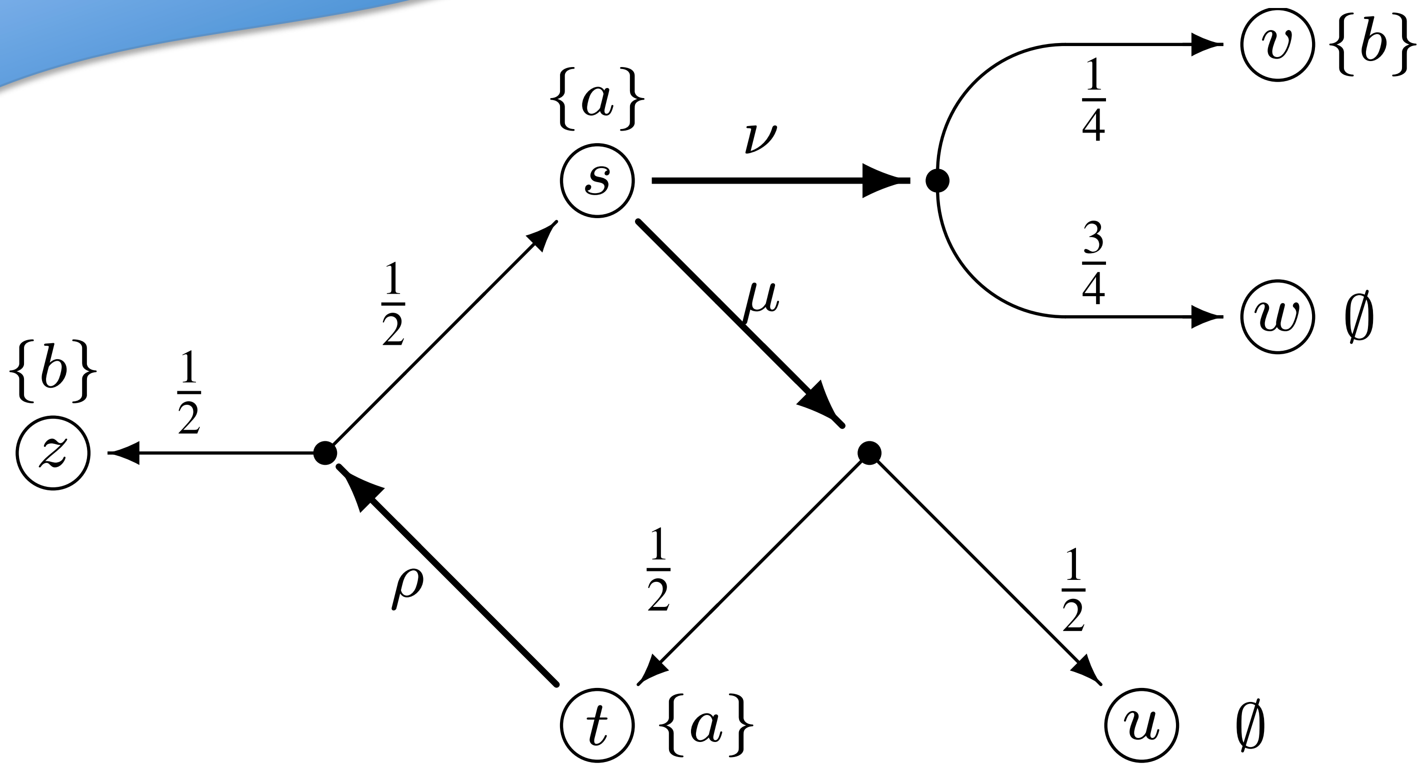
- Assume S_1 = Sat(ψ) (no announce checkmate)
- $x_n(s)$:= maximal probability of $(\varphi\ U^{\leq n}\ \psi)$,
  if s is the start state

- $x_0(S_1) = 1$ $\qquad$ $x_0(S_0 \cup S_?) = 0$
- $x_{n+1}(S_1) = 1$ $\qquad$ $x_{n+1}(S_0) = 0$

$$x_{n+1}(s) = \max_{\alpha \in Act} \sum_{t \in S} \mathbf{P}(s,\alpha,t) \cdot x_n(t) \qquad \text{for } s \in S_?$$

The example also shows that simple strategies do not suffice!

…ample



Which states satisfy $\mathbf{P}_{>1/4}(a \text{ EU}^{\leq 3} b)$?

Path formula: U (unbounded until)

- φ U ψ
- Find
  - S_1 = winning and announce checkmate states
  - S_0 = losing and stalemate states
  - $S_?$ = other states

- Which strategy to choose?

An equation system...

- $x(s)$:= maximal probability of $(\varphi \cup \psi)$,
  if s is the start state, under a *simple* strategy

- $x(S_1) = 1$ $\qquad$ $x(S_0) = 0$

$$x(s) = \max_{\alpha \in Act} \sum_{t \in S} \mathbf{P}(s,\alpha,t) \cdot x(t) \qquad \text{for } s \in S_?$$

...but what about other strategies?

- Assume
  $y(s)$:= maximal probability of (φ U ψ),
  if s is the start state, under any strategy

- $y(s)$ must be a solution of this equation system!
- uniqueness of solution → $x = y$

Value iteration + convergence test

- $x_n(s)$:= lower bound on maximal prob of (φ U ψ),
  $x^n(s)$:= upper bound on maximal prob of (φ U ψ),
  if s is the start state

- $x_0(S_1) = x^0(S_1 \cup S_?) = 1$, $x_0(S_0 \cup S_?) = x^0(S_0) = 0$
- $x_{n+1}(S_1) = x^{n+1}(S_1) = 1$, $x_{n+1}(S_0) = x^{n+1}(S_0) = 0$

$$x_{n+1}(s) = \max_{\alpha \in Act} \sum_{t \in S} \mathbf{P}(s,\alpha,t) \cdot x_n(t) \qquad \text{for } s \in S_?$$

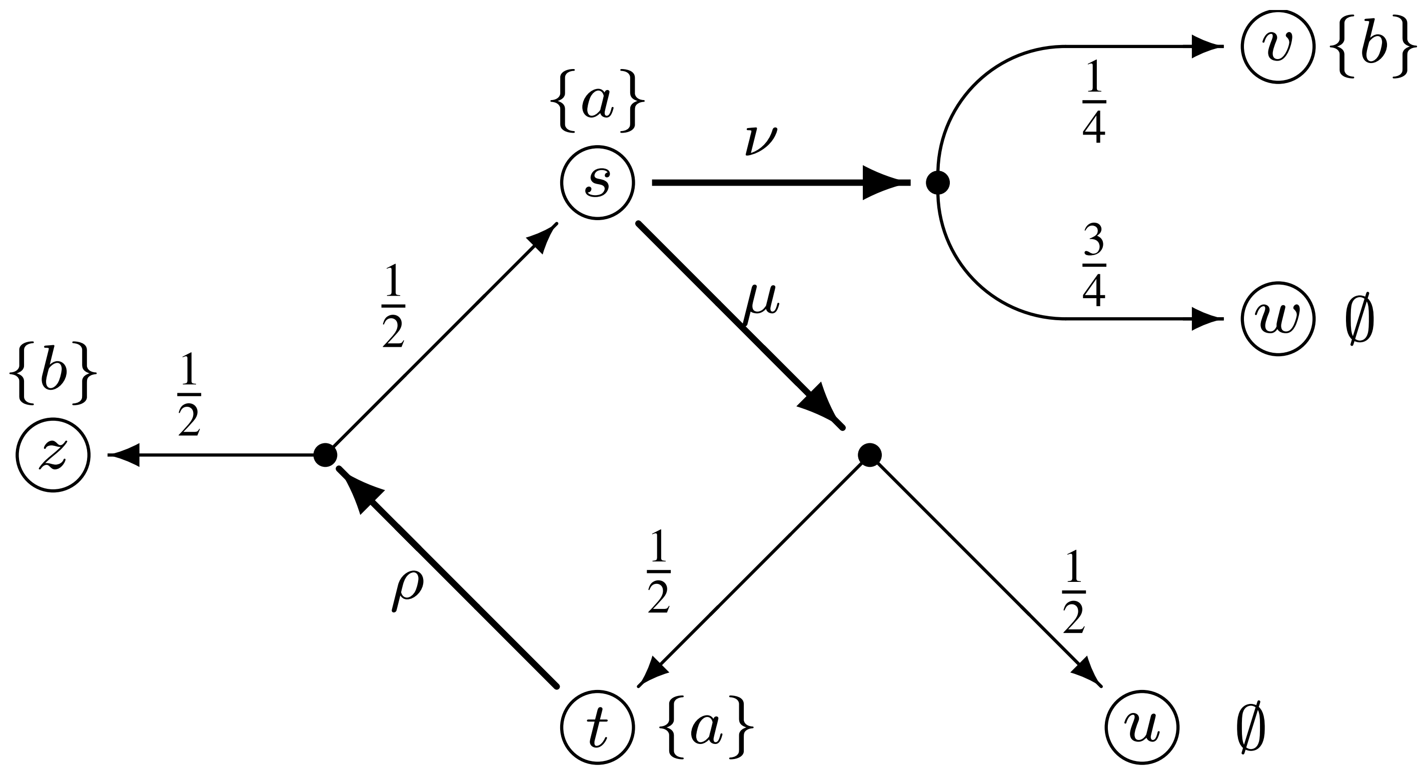
$$x^{n+1}(s) = \max_{\alpha \in Act} \sum_{t \in S} \mathbf{P}(s,\alpha,t) \cdot x^n(t) \qquad \text{for } s \in S_?$$

Policy iteration

- Start with a random strategy ζ_0
- Calculate $x_{\zeta_0}(s) := Prob^{\zeta_0}{}_s(\varphi \text{ U } \psi)$
- For each state, find the optimal action if the other states use ζ_n:

$$\zeta_{n+1}(s) = \arg\max_{\alpha \in Act} \sum_{t \in S} \mathbf{P}(s,\alpha,t) \cdot x_{\zeta_n}(t)$$

Example



Which states satisfy $\mathbf{P}_{\leq ½}$ (*a* AU *b*) ?

Some slides have been corrected
based on remarks made by Remy Viehoff.