

Exercise 1 (Positional Policies):**(2 points)**

Let $\mathcal{M}$ be a finite MDP with state space S and initial state s_0 .

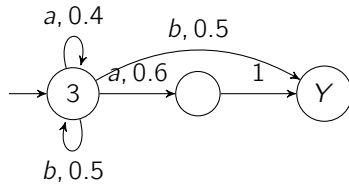
- a) Prove or disprove: For $T \subseteq S$ and for any $k \in \mathbb{N}$ there exists a positional policy such that

$$\Pr_{\sigma}^{\mathcal{M}}(s_0 \models \Diamond^{\leq k} T) = \Pr_{\min}^{\mathcal{M}}(s_0 \models \Diamond^{\leq k} T)$$

Solution: _____

False.

Consider a DTMC and a reachability property with $T = Y$. If only one step remains, the best option is a in the initial state, as T is certainly not reached. Otherwise, b is the safer option, as the probability that we do not reach the target is higher.

**Exercise 2 (Randomized Policies):****(5 points)**

Let $\mathcal{M}$ be a finite MDP with state space S and initial state s_0 . A positional randomized policy is given by a function $\sigma: S \rightarrow \text{Dist}(\text{Act})$.

- a) Define the induced DTMC of a positional randomized policy on an MDP.
- b) Prove or disprove: Let P be an ω -regular property s.t. there exists no deterministic positional policy σ with $\Pr_{\sigma}^{\mathcal{M}}(s_0 \models P) > 0$. Then there exists no randomized positional policy σ' such that $\Pr_{\sigma'}^{\mathcal{M}}(s_0 \models P) > 0$
- c) Prove or disprove: For $T \subseteq S$ and for every randomized positional policy σ exists a deterministic positional policy σ' such that

$$\Pr_{\sigma}^{\mathcal{M}}(s_0 \models \Diamond T) \leq \Pr_{\sigma'}^{\mathcal{M}}(s_0 \models \Diamond T)$$

Solution: _____

- a) Let $\mathcal{M} = (S, \text{Act}, P, s'_0, \text{AP}, L)$ be a finite MDP and $\sigma: S \rightarrow \text{Dist}(\text{Act})$ a positional scheduler. We define $\mathcal{D} = (S', P', s'_0, \text{AP}', L')$ with $S' = S$, $s'_0 = s_0$, $\text{AP}' = \text{AP}$ and $L' = L$.

$$P'(v, v') = \sum_{a \in \text{Act}} \sigma(s)(a) \cdot P(v, a, v') \quad \forall v, v' \in S'$$

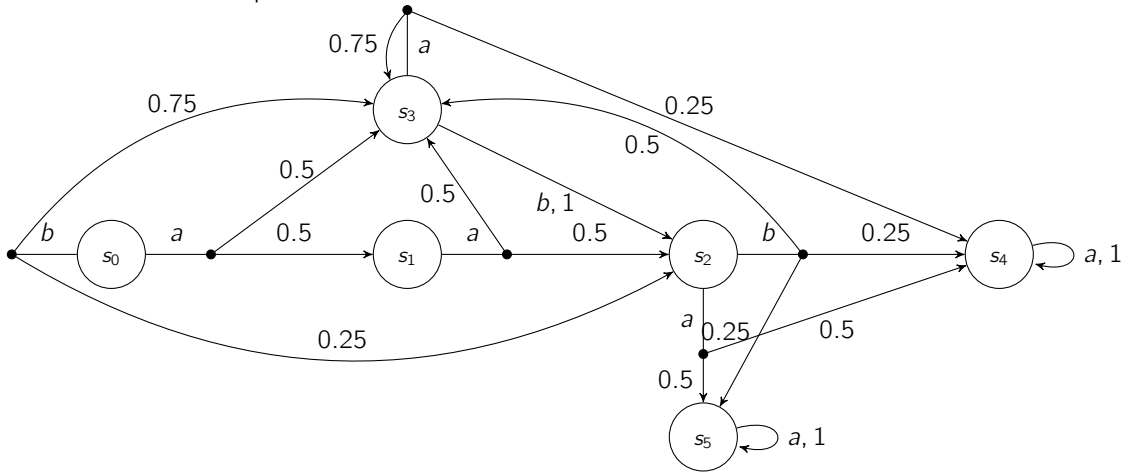
- b) Assume 2 is the initial state and the property infinitely often 1 and infinitely often 3. Any positional scheduler has probability 0, while for a random scheduler with non-zero probability for actions a and b , the probability is 1.

c)

Exercise 3 (Value Iteration):

(3 points)

Consider the MDP depicted below.



Use value iteration to compute the probabilities $\Pr_{max}^M(s \models \Diamond s_4)$ for all $s \in S$. You may abort the process after 6 iteration. Provide the final policy as described in the lecture.

Solution:

Initial step:

$$(0, 0, 0, 0, 1, 0)$$

First step:

$$(0, 0, \alpha, \beta, 1, 0)$$

$$\alpha = \max(0.5 \cdot 0 + 0.5 \cdot 1, 0.25 \cdot 1 + 0.5 \cdot 0 + 0.25 \cdot 0) = 0.5$$

$$\beta = \max(0.25 \cdot 1 + 0.75 \cdot 0, 1 \cdot 0) = 0.25$$

Second step:

$$(\alpha, \beta, \gamma, \delta, 1, 0)$$

$$\alpha = \max(0.5 \cdot 0.25 + 0.5 \cdot 0, 0.75 \cdot 0.25 + 0.25 \cdot 0.5) = \frac{3}{16} + \frac{2}{16} = \frac{5}{16}$$

$$\beta = 0.5 \cdot 0.5 + 0.5 \cdot 0.25 = \frac{3}{8}$$

$$\gamma = \max(0.5, 0.25 + 0.5 \cdot 0.25) = \frac{3}{8}$$

$$\delta = \max(0.25 + 0.75 \cdot 0.25, 1 \cdot 0.5) = 0.5$$

Third step:

$$(\alpha, \beta, \gamma, \delta, 1, 0)$$

$$\alpha = \max(0.5 \cdot \frac{3}{8} + 0.5 \cdot 0.5, 0.75 \cdot 0.5 + 0.25 \cdot \frac{3}{8}) = \frac{3}{8} + \frac{3}{32} = \frac{15}{32}$$

$$\beta = 0.5 \cdot \frac{3}{8} + 0.5 \cdot 0.5 = \frac{7}{16}$$

$$\gamma = \max(0.5, 0.25 + 0.5 \cdot 0.5) = \frac{1}{2}$$

$$\delta = \max(0.25 + 0.75 \cdot 0.5, 1 \cdot \frac{3}{8}) = \frac{5}{8}$$

Fourth step:

$$\begin{aligned}
 & (\alpha, \beta, \gamma, \delta, 1, 0) \\
 \alpha &= \max(0.5 \cdot \frac{7}{16} + 0.5 \cdot \frac{5}{8}, 0.75 \cdot \frac{5}{8} + 0.25 \cdot 0.5) = \frac{15}{32} + \frac{1}{8} = \frac{19}{32} \\
 \beta &= 0.5 \cdot \frac{5}{8} + 0.5 \cdot 0.5 = \frac{9}{16} \\
 \gamma &= \max(0.5, 0.25 + 0.5 \cdot \frac{5}{8}) = \frac{9}{16} \\
 \delta &= \max(0.25 + 0.75 \cdot \frac{5}{8}, 1 \cdot 0.5) = \frac{19}{32}
 \end{aligned}$$

Fifth step:

$$\begin{aligned}
 & (\alpha, \beta, \gamma, \delta, 1, 0) \\
 \alpha &= \max(0.5 \cdot \frac{9}{16} + 0.5 \cdot \frac{19}{32}, 0.75 \cdot \frac{19}{32} + 0.25 \cdot 0.5) = \frac{15}{32} + \frac{1}{8} = \frac{19}{32} \\
 \beta &= 0.5 \cdot \frac{9}{16} + 0.5 \cdot \frac{19}{32} = \frac{37}{64} \\
 \gamma &= \max(0.5, 0.25 + 0.5 \cdot \frac{19}{32}) = \frac{35}{64} \\
 \delta &= \max(0.25 + 0.75 \cdot \frac{9}{16}, 1 \cdot 0.5) = \frac{43}{64}
 \end{aligned}$$

Sixth step:

$$\begin{aligned}
 & (\alpha, \beta, \gamma, \delta, 1, 0) \\
 \alpha &= \max(0.5 \cdot \frac{19}{32} + 0.5 \cdot \frac{43}{64}, 0.75 \cdot \frac{43}{64} + 0.25 \cdot 0.5) = \frac{129}{256} + \frac{1}{8} = \frac{161}{256} \\
 \beta &= 0.5 \cdot \frac{35}{64} + 0.5 \cdot \frac{43}{64} = \frac{39}{64} \\
 \gamma &= \max(0.5, 0.25 + 0.5 \cdot \frac{43}{64}) = \frac{75}{128} \\
 \delta &= \max(0.25 + 0.75 \cdot \frac{43}{64}, 1 \cdot 0.5) = \frac{193}{256}
 \end{aligned}$$