

Chapter 3 - Finite Markov Decision Processes

Stéphane Liem NGUYEN

August 15, 2021

Some keywords, citations and formulas.

Particularities of the value functions When we do not want to select actions based on the knowledge of the environment dynamics, action-value functions can be used because they "cache the results of all one-step-ahead searches." (Upper part of the page 65).

Bellman (expectation) equations for the four value functions Bellman equation for the state-value function for policy π

$$v_{\pi}(s) = \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a) [r + \gamma v_{\pi}(s')] \quad (1)$$

Bellman equation for the optimal state-value function (Bellman optimal equation for v_*)

$$v_*(s) = \max_a \sum_{s',r} p(s',r|s,a) [r + \gamma v_*(s')] \quad (2)$$

Bellman equation for the action-value function for policy π

$$q_{\pi}(s,a) = \sum_{s',r} p(s',r|s,a) \left[r + \gamma \sum_{a'} \pi(a'|s') q_{\pi}(s',a') \right] \quad (3)$$

Bellman equation for the optimal action-value function (Bellman optimal equation for q_*)

$$q_*(s,a) = \sum_{s',r} p(s',r|s,a) \left[r + \gamma \max_{a'} q_*(s',a') \right] \quad (4)$$

where $a \in \mathcal{A}(s)$, s' and $s \in \mathcal{S}$ and $r \in \mathcal{R}$