Optimization and Uncertainty

Summer term 2021

Prof. Dr. Martin Hoefer
Tim Koglin, Lisa Wilhelmi

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Institute of Computer Science
Algorithms and Complexity

Assignment 8

Issued: 22.06.2021
Due: 29.06.2021, **10:00h**

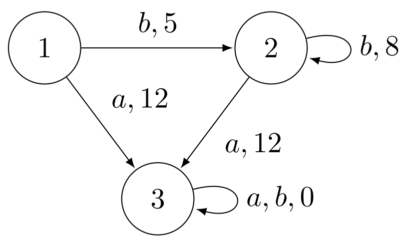
Exercise 8.1 *Online* DELEGATION (4 points)

Consider an online variant of DELEGATION where n independent boxes arrive sequentially in a fixed order. Both the sender $\mathcal{S}$ and the receiver $\mathcal{R}$ know the order of the boxes and their respective distributions in advance. At arrival of box i, $\mathcal{S}$ looks in the box and decides immediately whether to recommend it to $\mathcal{R}$ or not. If $\mathcal{S}$ lets it pass, the next box $i+1$ arrives. The process ends when $\mathcal{S}$ recommends a box (upon which $\mathcal{R}$ makes the accept/reject decision according to decision scheme ψ) or if $\mathcal{S}$ has let all boxes pass.

Show that there exists an instance of online DELEGATION where any decision scheme ψ is $\Omega(n)$-competitive (compared to the expected optimal value for $\mathcal{R}$).
Hint: Choose distributions for the boxes that incentivize $\mathcal{S}$ to recommend the last possible box.

Exercise 8.2 *Value Iteration versus Policy Iteration* (2 + 2 + 2 points)

Consider a Markov decision process with states $\mathcal{S} = \{1, 2, 3\}$ and actions $\mathcal{A} = \{a, b\}$ which is depicted below. The state transitions are deterministic. The numbers in the edge labels are the respective rewards. Assume an infinite time horizon with discount factor $\gamma = \frac{1}{2}$.



a) Derive an optimal Markovian policy π^* and $V^*(s)$ for all $s \in \mathcal{S}$.

b) Perform the first six steps of value iteration starting with initial vector $v^{(0)} = (0, 0, 0)$.

c) Starting from the policy that always performs action a, apply policy iteration until convergence.

Exercise 8.3 *Value Iteration with Caution* (4 points)

Consider a more cautious version of value iteration for MDPs with infinite time horizon with state set $\mathcal{S}$ and action set $\mathcal{A}$. It uses the operator t' which is defined by $t'(v)_s = \eta \cdot t(v)_s + (1-\eta) \cdot v_s$, for all states $s \in \mathcal{S}$, where t is the value iteration defined in the lecture and $\eta \in (0, 1)$ is an arbitrary parameter.

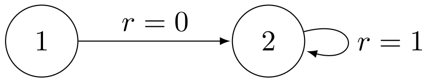
Show that t' converges to the unique fixed point of t.

Exercise 8.4 *Gittins Index* (2 + 2 points)

Consider the following instances for the Markovian Single-Armed Bandit problem with charges $\lambda \geq 0$. Let r denote the reward of a transition when action `play` is chosen, and p denotes the probability that the respective transition occurs ($p = 1$ unless stated otherwise). If `pause` is chosen, no transition occurs and the reward is zero in this round. At each iteration step, the probability that the process terminates is $\gamma \in (0, 1)$.

For each of the single-armed bandits, derive the Gittins indices of all states.

a)



b)