Thomas Kesselheim
Alexander Braun

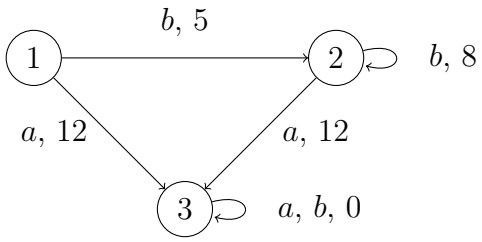
June 17, 2021

# Algorithms and Uncertainty

Summer Term 2021

Tutorial Session - Live Tasks 7

**Exercise 1:**
We consider a Markov decision process with $\mathcal{S} = \{1, 2, 3\}$, $\mathcal{A} = \{a, b\}$. The state transitions are deterministic as displayed in this diagram; the numbers in the edge labels are the respective rewards.




We consider an infinite time horizon with discount factor $\gamma = \frac{1}{2}$.

(a) Give an optimal policy and the function $s \mapsto V^*(s)$.

(b) Perform the first six steps of value iteration starting from $W^{(0)} = (0, 0, 0)$.

(c) Perform policy iteration until convergence starting from the policy that always uses action $a$.