CSE 150 Assignment 5 - Policy Improvement

Josh Anthony - A14281769

Consider the Markov decision process (MDP) with two states s ∈ {0, 1}, two actions a ∈ {0, 1}, discount factor $\gamma = \frac{2}{3}$, and rewards and transition matrices as shown below:

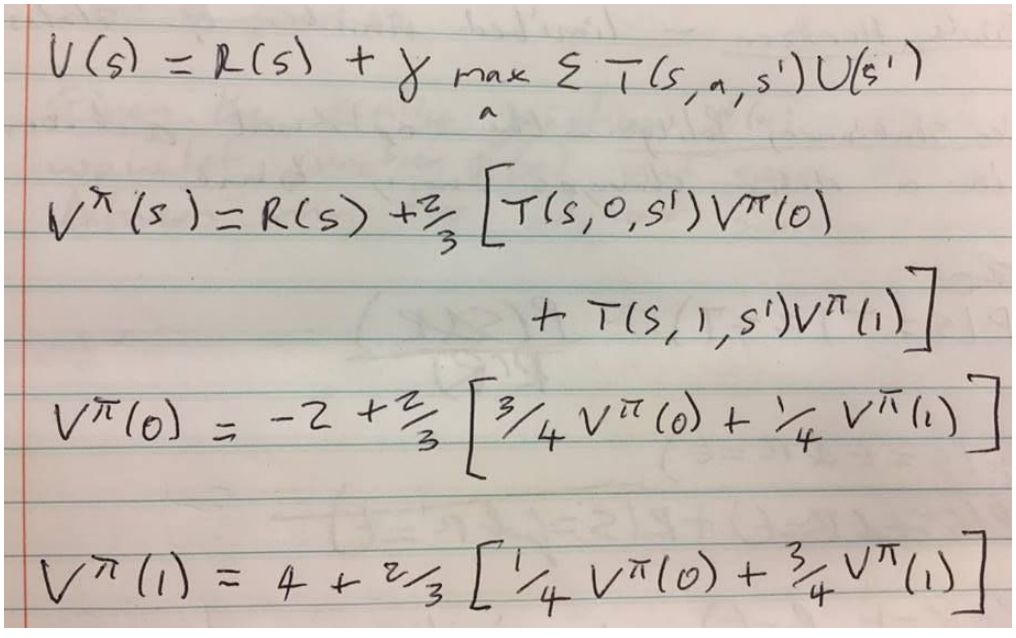
s	R(s)
0	-2
1	4

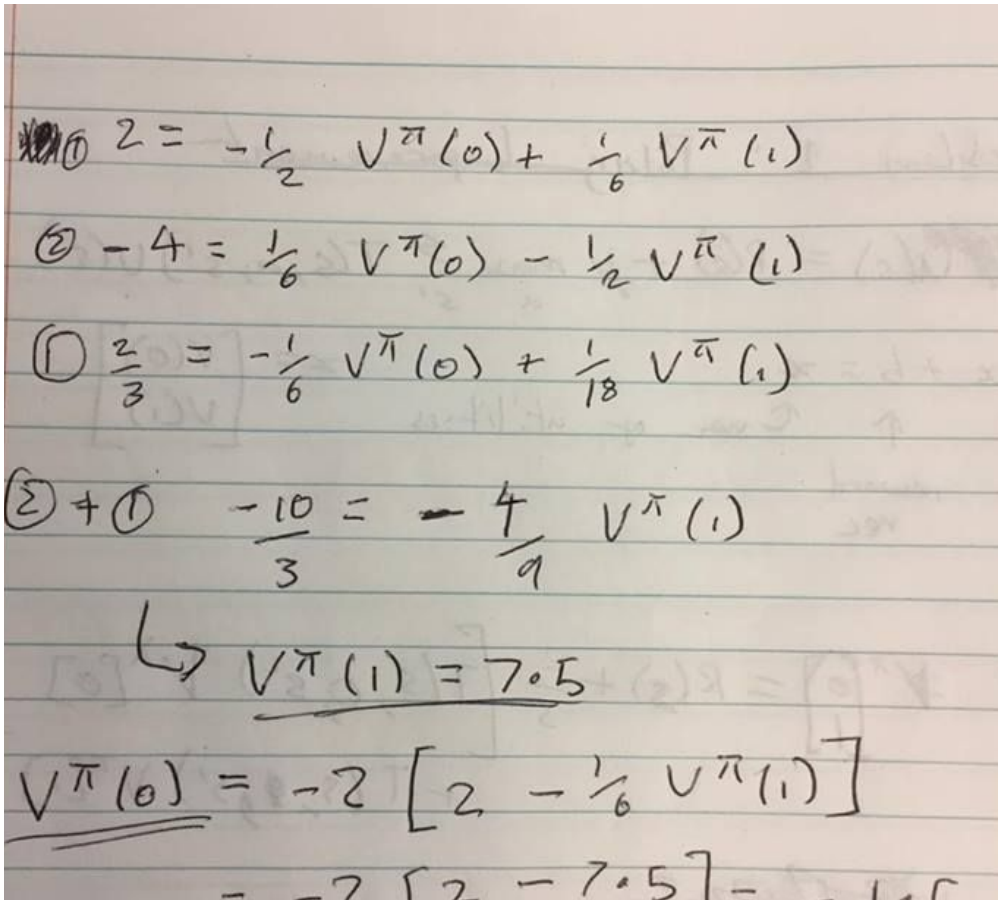
s	s'	P(s' | s, a=0)
0	0	3/4
0	1	1/4
1	0	1/4
1	1	3/4

s	s'	P(s' | s, a=1)
0	0	1/2
0	1	1/2
1	0	1/2
1	1	1/2

a) (2.5 points) Consider the policy π that chooses the action a = 0 in each state. For this policy, solve the linear system of Bellman equations to compute the state-value function $V^{\pi}(s)$ for s ∈ {0, 1}. Your answers should complete the following table:

s	π(s)	$V^{\pi}(s)$
0	0	
1	0	

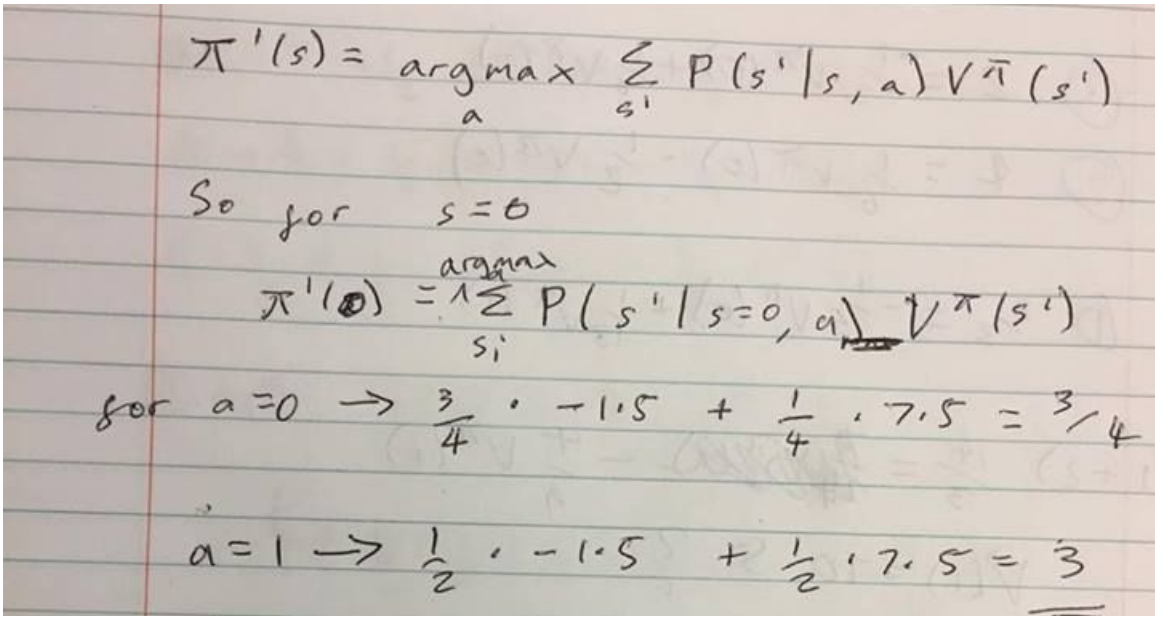


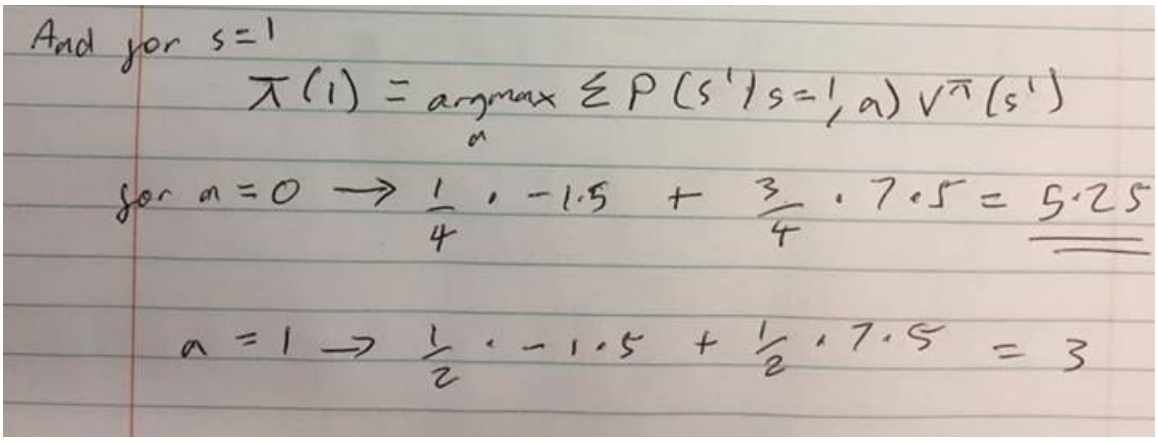


So $V^{\pi}(0) = -1.5$ and $V^{\pi}(1) = 7.5$

b) (2.5 points) Compute the greedy policy $\pi'(s)$ with respect to the state-value function $V^{\pi}(s)$ from part (a). Your answers should complete the following table:

s	π(s)	π'(s)
0	0	
1	0	





So $\pi'(0) = 1$ and $\pi'(1) = 0$