

Andrew Hwang  
AUG 20 128

Problem 1. Policy Iteration

$$V_{\pi}(s) = \sum_a \pi(a) \left[ \sum_{s'} P(s'|s,a) (R(s) + \gamma V_{\pi}(s')) \right]$$

$$= \sum_a \left[ R(s) \pi(a) + \sum_{s'} P(s'|s,a) \gamma V_{\pi}(s') \right]$$

$$= R(s) + \gamma \sum_{s'} P(s'|s,a) V_{\pi}(s')$$

a)  $V(0) = -2 + \frac{3}{4} V(0) + \frac{1}{4} V(1)$   
 $V(1) = 4 + \frac{1}{2} V(0) + \frac{1}{2} V(1)$   
 $V(0) = -2 + \frac{1}{4} V(0) + \frac{3}{4} V(1)$   
 $V(1) = 4 + \frac{1}{2} V(0) + \frac{1}{2} V(1)$   
 $\frac{3}{4} V(0) = -2 + \frac{3}{4} V(1)$   
 $V(0) = -2 + \frac{3}{4} V(1)$   
 $V(0) = -2 + \frac{1}{4} V(0) + \frac{3}{4} (-2 + \frac{3}{4} V(1))$   
 $= -2 + \frac{1}{4} V(0) + \frac{3}{4} (-2 + \frac{3}{4} V(1))$   
 $\frac{3}{4} V(0) = -2 + \frac{9}{16} V(1)$   
 $\frac{9}{16} V(0) = -\frac{8}{9}$   
 $V(0) = -\frac{16}{9}$   
 $= -1.78$   
 $V(1) = 4 + \frac{1}{2} (-1.78) + \frac{1}{2} V(1)$   
 $\frac{1}{2} V(1) = 4 - \frac{1}{4} = 3.75$   
 $V(1) = 7.5$   
 $V(0) = -1.5 \quad V(1) = 7.5$

b)  $\pi'(s) = \arg \max_a \sum_{s'} P(s'|s,a) V^{\pi'}(s')$

$s=0, a=0: V(0) \frac{3}{4} + V(1) \frac{1}{4} = -1.5 \cdot \frac{3}{4} + 7.5 \cdot \frac{1}{4} = -1.125 + 1.875 = 0.75$   
 $s=0, a=1: V(0) \frac{1}{2} + V(1) \frac{1}{2} = -1.5 \cdot \frac{1}{2} + 7.5 \cdot \frac{1}{2} = -0.75 + 3.75 = 3$   
 $s=1, a=0: V(0) \frac{1}{4} + V(1) \frac{3}{4} = -1.5 \cdot \frac{1}{4} + 7.5 \cdot \frac{3}{4} = -0.375 + 5.625 = 5.25$   
 $s=1, a=1: V(0) \frac{1}{2} + V(1) \frac{1}{2} = -1.5 \cdot \frac{1}{2} + 7.5 \cdot \frac{1}{2} = -0.75 + 3.75 = 3$

$s=0, a=1$   
 $s=1, a=0$