Reading Assignment Two

Amol Kapoor, Adam Cobb
Worcester College, Oxford University

May 17, 2017

What are the equations analogous to (4.3), (4.4), and (4.5) for the action-value function q and its successive approximation by a sequence of functions q0 , q1 , q2 , . . . ?

For (4.3) and (4.4):

$$\begin{aligned} q_\pi(s,a) &= \mathbf{E}_\pi[R_{t+1} + \gamma q_\pi(S_{t+1}, A_{t+1}) | S_t = s, A_t = a] \\ &= \sum_{s'} p(s'|s,a)[r(s,a,s') + \sum_{a'} \pi(a'|s') \gamma q_\pi(s',a')] \end{aligned} \tag{1}$$

For (4.5):

$$\begin{aligned} q_{k+1}(s,a) &= \mathbf{E}_\pi[R_{t+1} + \gamma q_k(S_{t+1}, A_{t+1}) | S_t = s, A_t = a] \\ &= \sum_{s'} p(s'|s,a)[r(s,a,s') + \sum_{a'} \pi(a'|s') \gamma q_k(s',a')] \end{aligned} \tag{2}$$

Write a program for policy iteration and re-solve Jacks car rental problem with the following changes. One of Jacks employees at the first location rides a bus home each night and lives near the second location. She is happy to shuttle one car to the second location for free. Each additional car still costs 2 dollars, as do all cars moved in the other direction. In addition, Jack has limited parking space at each location. If more than 10 cars are kept overnight at a location (after any moving of cars), then an additional cost of 4 dollars must be incurred to use a second parking lot (independent of how many cars are kept there). These sorts of nonlinearities and arbitrary dynamics often occur in real problems and

cannot easily be handled by optimization methods other than dynamic programming. To check your program, first replicate the results given for the original problem. If your computer is too slow for the full problem, cut all the numbers of cars in half.

See attached code and images.