Markov Decision Processes

1) Invent a simple Markov decision process (MDP) with the following properties: a) it has a goal state, b) its immediate action costs are all positive, c) all of its actions can result with some probability in the start state, and d) the optimal policy without discounting differs from the optimal policy with discounting and a discount factor of 0.9. Prove d) using value iteration.

2) Consider the following problem (with thanks to V. Conitzer): Consider a rover that operates on a slope and uses solar panels to recharge. It can be in one of three states: high, medium and low on the slope. If it spins its wheels, it climbs the slope in each time step (from low to medium or from medium to high) or stays high. If it does not spin its wheels, it slides down the slope in each time step (from high to medium or from medium to low) or stays low. Spinning its wheels uses one unit of energy per time step. Being high or medium on the slope gains three units of energy per time step via the solar panels, while being low on the slope does not gain any energy per time step. The robot wants to gain as much energy as possible.

a) Draw the MDP graphically. b) Solve the MDP using value iteration with a discount factor of 0.8. c) Describe the optimal policy.

Now answer the three questions above for the following variant of the robot problem: If it spins its wheels, it climbs the slope in each time step (from low to medium or from medium to high) or stays high, all with probability 0.3. It stays where it is with probability 0.7. If it does not spin its wheels, it slides down the slope to low with probability 0.4 and stays where it is with probability 0.6. Everything else remains unchanged from the previous problem.

3) Consider the following problem (with thanks to V. Conitzer): Consider a rover that operates on a slope. It can be in one of four states: top, high, medium and low on the slope. If it spins its wheels slowly, it climbs the slope in each time step (from low to medium or from medium to high or from high to top) with probability 0.3. It slides down the slope to low with probability 0.7. If it spins its wheels rapidly, it climbs the slope in each time step (from low to medium or from medium to high or from high to top) with probability 0.5. It slides down the slope to low with probability 0.5.

Spinning its wheels slowly uses one unit of energy per time step. Spinning its wheels rapidly uses two units of energy per time step. The rover is low on the slope and aims to reach the top with minimum expected energy consumptions.

a) Draw the MDP graphically. b) Solve the MDP using undiscounted value iteration (that is, value iteration with a discount factor of 1). c) Describe the optimal policy.

4) You won the lottery and they will pay you one million dollars each year for 20 years (starting this year). If the interest rate is 5 percent, how much money do you need to get right away to be indifferent between this amount of money and the annuity?

5) Assume that you are trying to pick up a block from the table. You drop it accidentally with probability 0.7 while trying to pick it up. If this happens, you try again to pick it up. How many attempts does it take on average before you pick up the block successfully?

6) Assume that you use undiscounted value iteration (that is, value iteration with a discount factor of 1) for a Markov decision process with goal states, where the action costs are greater than or equal to zero. Give a simple example that shows that the values that value iteration converges to can depend on the initial values of the states, in other words, the values that value iteration converges to are not necessarily equal to the expected goal distances of the states.

7) An MDP with a single goal state ($S3$) is given below. a) Given the expected goal distances $c(S1) = 7$, $c(S2) = 4.2$, and $c(S3) = 0$, calculate the optimal policy. b) Suppose that we want to follow a policy where we pick action $a2$ in state $S1$ and action $a3$ in state $S2$. Calculate the expected goal distances of $S1$ and $S2$ for this policy.