IE5571/8571 Homework 3

You can discuss problems with other students but must write up your own solution.

Problem 1: Exercises 4.9, 5.6, and 5.8 from textbook.

Problem 2

John wants to sell his car. He receives one offer each month and must decide immediately whether to accept the offer. Once rejected the offer is lost. The possible offers are \$500, \$600, \$700, \$800, and \$1000, with respective probabilities 1/4, 1/4, 1/6, 1/6, and 1/6. Each month he has to pay \$60 to maintain the car.

Assume a discount factor of $\alpha = 0.97$.

(a) Starting with an arbitrary initial condition, compute two iterations of the value iteration algorithm.

(b) Use policy iteration to find a policy that minimizes his expected discounted cost.

Problem 3

A manufacturer relies on one key machine. Due to heavy use, the machine deteriorates rapidly. At the end of each week a thorough inspection is done that classifies the machine into one of four possible states

0–Good as new

1–Operable minor deterioration

2–Operable major deterioration

3–Inoperable

Without any repairs the state of the machine evolves as a Markov chain with a transition matrix

$$\mathbb{P} = \begin{bmatrix} 0 & 7/8 & 1/16 & 1/16 \\ 0 & 3/4 & 1/8 & 1/8 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Thus if the the state of the machine is 1 then it will be in state 1 next week with probability 3/4.

It's obviously not acceptable to have the machine inoperable, so the manufacturer would like to replace the machine if it's in state 3. This costs \$6000. In addition, the manufacturer accrues costs from producing defective items if the machine is in states 1 or 2. In particular, it costs \$1000 per week the machine is in state 1, and \$3000 per week the machine is in state 2. Finally, the manufacturer can repair the machine for \$2000 if it's in state 2. This repair returns the machine to state 1.

Using a discount factor of $\alpha = 0.95$ find the optimal policy via policy iteration.

Problem 4

Consider an infinite-period inventory system with a single product where, at the beginning of each period, a decision is be made about how many items to produce during that period. The setup cost is \$10, and the unit production cost is \$5. The holding cost for each item not sold during each period is \$4, and a maximum of 2 items can be stored. During each period, demand is 0, 1, or 2 items each with probability 1/3. If demand exceeds the supply available during that period, those sales are lost and a shortage cost is incurred, namely \$8 for a shortage of 1 units and \$32 for a shortage of 2 units.

Using a discount factor of $\alpha = 0.95$ find the optimal policy via policy iteration.

(IE 8571) Problem 5

Prove the Policy Improvement Theorem for randomized policies.

(IE 8571) Problem 6

Recall that for episodic MDP's there exist a terminal state which we denote here by Δ. In order to guarantee convergence of algorithms like VI and PI it is commonly assumed that there exists a finite integer m such that for all policies π

$$\rho_\pi = \max_{s \in \mathcal{S}} P_\pi(S_m \neq \Delta | X_0 = s) < 1. \tag{1}$$

Since $|\mathcal{S}| < \infty$ and $|\mathcal{A}(s)| < \infty$ the number of deterministic policies is finite. If we denote the space of deterministic policies by $\mathcal{P}_0$ then (1) implies that

$$\max_{\pi \in \mathcal{P}_0} \rho_\pi < 1.$$

If we denote the space of all possible policies by $\mathcal{P}$, show that in fact

$$\max_{\pi \in \mathcal{P}} \rho_\pi < 1.$$