Total Marks: 25, Total Time: 4 hrs

Instructions

1. Work on your own. You can discuss with your classmates on the problems, use books or web. However, the solutions that are submitted must be your own and you must acknowledge all the sources (names of people you worked with, books, webpages etc., including class notes.) Failure to do so will be considered cheating. Identical or similar write-ups will be considered cheating as well.
2. In your submission, add the following declaration at the outset:
   "I pledge that I have not copied or given any unauthorized assistance on this exam."
3. The exam is divided into two sections. For the first section, either provide a proof or disprove using a counterexample. For the second section, provide a detailed answer, showing all the necessary steps.

I Prove or disprove

2 1. In an SSP problem with at least one proper policy, suppose that each improper policy π has $J_\pi(i) = \infty$ for at least one state i. Then, there is no improper policy π' such that $J_{\pi'}(j) = -\infty$ for at least one state j.

2 2. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a contraction mapping with modulus α under the Euclidean norm. Let x^* denote the fixed point of f. Then, $f(x) \leq x$ implies $x^* \leq x$. Here $\leq$ is element-wise.

2 3. In a finite horizon MDP setting, suppose we have time-invariant state and action spaces. Consider a modified problem, where the terminal cost $g_N(x)$ is replaced by $g'_N(x) = g_N(x) + 10$. Let J_k and J'_k denote the kth stage functions in the DP algorithm for the original and modified problems. Then, $J_k(x) \leq J'_k(x)$, for all x and k.

II Long answer problems

4 4. **Policy iteration**
Consider a machine replacement problem over N stages. In each stage, the machine can be in one of the n states, denoted $\{1, \ldots, n\}$. Let $g(i)$ denote the cost of operating the machine in state i in any stage, with

$$g(1) \leq g(2) \ldots \leq g(n).$$

The system is stochastic, with p_{ij} denoting the probability that the machine goes from state i to j. Note that the machine can either go worse or stay in the same conditions, which implies $p_{ij} = 0$ for $j < i$. The state of the machine at the beginning of each stage is known and the possible actions are (i) perform no maintenance, i.e., let the machine run in the current state; and (ii) repair the machine at a cost R. On repair, the machine transitions to state 1 and remains there for one stage.

Consider this problem in the discounted MDP framework, with a discount factor $\beta \in (0, 1)$. Assume $p_{ij} = 0$ if $j < i$, and $p_{ij} \leq p_{(i+1)j}$ if $i < j$. A threshold policy is defined to be a stationary

policy that repairs if and only the state is equal to or greater than some state, say k. Suppose we run policy iteration on this problem. If the initial policy is a threshold policy, show that all subsequent policies are threshold policies.

[4] 5. A finite horizon MDP setting is as follows:

Horizon $N = 3$, states $\{1, 2\}$, actions $\{a, b\}$ (available in each state), and transition probabilities defined by

$$p_{11}(a) = \frac{3}{4}, p_{12}(a) = \frac{1}{4}; \quad p_{11}(b) = \frac{1}{4}, p_{12}(b) = \frac{3}{4};$$
$$p_{21}(a) = \frac{1}{5}, p_{22}(a) = \frac{4}{5}; \quad p_{21}(b) = \frac{4}{5}, p_{22}(b) = \frac{1}{5};$$

The time-invariant single-stage costs are as follows:

$$g(1, a) = 1, g(1, b) = 0, g(2, a) = 2, g(2, b) = 5.$$

There is no terminal cost. Calculate the optimal expected cost using the DP algorithm, and specify an optimal policy.

6. Suppose $f : \mathbb{R}^n \to \mathbb{R}^n$ satisfies $\|f(x) - f(y)\| \leq \|x - y\|$, where $\|\cdot\|$ is the Euclidean norm. For some m, f^m is a contraction mapping with modulus α, and fixed point x^*. Here $f^m(x)$ is f applied m times on x.

Show that

[1.5] (a) x^* is the unique fixed point of f.

[1.5] (b) $\|x^* - x\| \leq \frac{m}{(1-\alpha)} \|f(x) - x\|, \forall x \in \mathbb{R}^n$.

7. The CS9890 mid-term is an open book exam with two questions. For each question, a student taking this exam can either (i) think and solve the problem; or (ii) search the internet for the solution. If he/she chooses to think, then he/she has a probability p_1 of finding the right solution approach. In the case when the right approach is found, the student has a probability p_2 of writing an answer that secures full marks. The corresponding probabilities for option (ii) are q_1 and q_2. The draconian instructor would let the student pass only if both problems are solved perfectly.

Assume $p_i, q_i > 0$, for $i = 1, 2$, and answer the following:

[3] (a) Formulate this problem as an SSP with the goal of passing the exam, and characterize the optimal policy.

[1] (b) Find a suitable condition on p_i, q_i under which it is optimal to think for each question.

[1] (c) When is it optimal to search the internet for each question?

8. A discounted MDP is specified below.

States $\{1, 2\}$, actions $\{a, b\}$ in state 1, and $\{c, d\}$ in state 2. The transition probabilities are

$$p_{11}(a) = p_{12}(a) = 0.5; \quad p_{11}(b) = 0.8, p_{12}(b) = 0.2;$$
$$p_{21}(c) = 0.4, p_{22}(c) = 0.6; \quad p_{21}(d) = 0.7, p_{22}(d) = 0.3;$$

The discount factor $\alpha = 0.9$.

The time-invariant single-stage costs are as follows:

$$g(1, a, 1) = -9, g(1, a, 2) = -3, g(1, b, 1) = -4, g(1, b, 2) = -4,$$
$$g(2, c, 1) = -3, g(2, c, 2) = 7, g(2, d, 1) = -1, g(2, d, 2) = 10.$$

For each of the policies given below, find the expected discounted cumulative cost.

1.5 (a) Policy π: $\pi(1) = a, \pi(2) = c$.

1.5 (b) Policy $\widetilde{\pi}$: $\widetilde{\pi}(1) = b, \widetilde{\pi}(2) = d$.