MATH4406 (Control Theory), HW5 (Unit 5)
Infinite Horizon Discounted MDP – Part 2.

Prepared by Yoni Nazarathy, Last Updated: September 21, 2014.

This homework project is the second homework about infinite horizon discounted MDP. The previous homework was computational. This one is of a more theoretical nature. Quiz 3, will be based on this homework assignment.

Part 1: The Machine Replacement Model.

Suppose that at each epoch, a production machine is inspected and its condition and state is noted. States are $\{0, 1, 2, \ldots\}$ with state 0 being "perfectly new". With each state i, an operating cost of $C(i)$ is incurred at each epoch where $C(\cdot)$ is assumed to be an increasing function in i. After inspecting the state, at each epoch a decision is made: $a = 0$ (not replace the machine) or $a = 1$ (replace the machine). If the decision is to replace, a cost $R > 0$ is immediately incurred. This then moves the state of the machine to 0 for the next epoch. If the decision is not to replace then the condition of the machine evolves randomly according to the probabilities $P_{i,j}$. I.e. this is the probability of the machine changing from state i to state j, if not replaced.

We assume the following about $P_{i,j}$: For each k, $\sum_{j=k}^{\infty} P_{i,j}$ is an increasing function of i. This means that if T_i is a random variable representing the next state visited after i (assuming no replacement) then,

$$\mathbb{P}(T_{i+1} > k) \;\geq\; \mathbb{P}(T_i > k).$$

This is known as a *stochastic order* and is equivalent to having

$$\mathbb{E}\left[f(T_{i+1})\right] \geq \mathbb{E}\left[f(T_i)\right],$$

for all increasing functions f.

Some of the items below are easy. Some are harder. The key items are **5** and **6** where you will prove structural properties of the problem.

1: Describe (briefly) a real life situation where this machine replacement model may be applicable. In that situation, explicitly state some example R, $C(\cdot)$ and $P_{i,j}$.

2: Argue (briefly) why the stochastic ordering assumption above is sensible.

3: Pose the problem as an MDP with infinite horizon and discounted objective with discount factor $\lambda \in (0, 1)$. Write the state-space, action-set, transition probabilities, rewards etc.

4: Use the "standard" form of the optimality equation (e.g. Eq (6.2.2) on page 146 of [Put94]) to show that the optimality equation for this MDP can be written as,

$$v(i) = C(i) + \min\{R + \lambda v(0),\ \lambda \sum_{j=0}^{\infty} P_{i,j} v(j)\}.$$

Here $v(\cdot)$ is the value-function but with respect to the minimisation criterion.

5: Use the value-iteration, $v^0(i) = C(i)$ and for $n \geq 1$,

$$v^n(i) = C(i) + \min\{R + \lambda\, v^{n-1}(0), \alpha \sum_{j=0}^{\infty} P_{i,j} v^{n-1}(j)\},$$

to prove that $v(i)$ is an increasing function.

6: Use the above to prove that there exists an $\bar{i} \leq \infty$ such that an optimal policy is to replace when $i \geq \bar{i}$ and not replace if $i < \bar{i}$.

7: Provide an example where $\bar{i} = \infty$.

8: Devise an algorithm for finding $\bar{i}$. Specify the algorithm precisely. You do not need to implement it.

Part 2: More on using Contraction Mappings and Rates of Convergence.

Look at Theorem 6.3.3, on page 163 of [Put94].

9: The theorem describes the rates of convergence of value-iteration, but you first need to understand what the statement means. Write out the theorem statement and briefly indicate both the exact mathematical meaning of (a)–(e) (defining "linear rate" etc...) and then describe the applicative meaning in each of those cases. For this read the "Rates of Convergence" sub-section 6.3.1 starting on page 159.

10: Write out the proof of the theorem, filling in any steps that are missing in the book.

Look at Theorem 6.4.8. on page 181 of [Put94].

11: This theorem states conditions for policy iteration to converge quadratically (better than value-iteration which is linear). Study these conditions and describe in which situations you believe policy-iteration is a better algorithmic choice and in which situations it is not. Use Corollary 6.4.9 if needed.

12 (optional): Write out the proof of the Theorem. In the process, fill out any missing details - including results used on the way.