B9120-001 Dynamic Programming Fall 2023
Prof. Daniel Russo

Homework Assignment 7: Due Thursday November 2

DP theory without contraction.
This problem looks at stochastic shortest path problems. Compared to the treatment provided in class, we remove the assumption that all policies are assured to terminate in finite time, but impose the assumptions (a) that *some policy terminates* and (b) that costs are non-negative (so termination is good).

In this case, the Bellman operator is not a contraction operator. Nevertheless, many of the most important conclusions of our theory continue to hold.

This problem exposes you to DP arguments that leverage monotonicity properties instead of contraction properties, and should make it easier to follow parts of class in the coming weeks.

Setup

- The state space is $\mathcal{X} \cup \{\emptyset\}$ where $\mathcal{X}$ is finite.
- The control space is finite.
- The 'terminal state' $\emptyset$ is costless [$g(\emptyset, u) = 0$] and absorbing [$\mathbb{P}(x_{k+1} = \emptyset | x_k = \emptyset, u_k = u) = 1$]. This implies that any policy incurs zero expected cost starting from $\emptyset$.

We place two major regularity conditions on the problem:

a) Single period costs are non-negative, i.e $g(x, u) \geq 0$ for each $x \in \mathcal{X}$ and $u \in U(x)$.

b) There exists at least one stationary policy μ that reaches the terminal state with probability 1, regardless of the initial state. You may use without proof that[1]

$$\mathbb{E}[\tau \mid x_0 = x] < \infty$$

for any $x \in \mathcal{X}$, where $\tau = \inf\{k \geq 0 : x_k = \emptyset\}$ is the hitting time of the terminal state.

[1]this follows from the equivalence between null and positive recurrence in finite Markov chains

Problem

Each part has a proof that is a few lines. You should not be writing a lot for any step. If miss the proof of part a, you can still proceed to complete other arguments assuming part a's conclusion holds; the same for other parts.

a) Consider the sequence

$$J_k = T^k J_0 \qquad k \in \{0, 1, \ldots\},$$

where J_0 is the vector of all zeros (*important*!). Conclude by inductively applying the monotonicity property of the Bellman operator that $J_k(x)$ is monotonically increasing for each x. Use the monotone convergence theorem to conclude that the limiting vector $J_\infty = \lim_{k\to\infty} J_k$ exists.

(It may take infinite values at some elements. We don't yet rule that out.)

b) Consider an arbitrary (possibly non-stationary) policy $\pi = (\mu_0, \mu_1, \mu_2, \ldots)$. Use the monotone convergence theorem to conclude that

$$J^\pi = \lim_{k\to\infty} T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{k-1}} J_0$$

exists. *Hint: write* $(T_{\mu_0} T_{\mu_1} \cdots T_{\mu_{k-1}} J_0)(x)$ *as the expected total cost over a* k *period horizon.*

c) Argue that $J_\infty \preceq J^\pi$ for every π and hence that J_∞ has finite elements.

d) Use the continuity of T to conclude that J_∞ is a fixed point of T, i.e. $J_\infty = TJ_\infty$.

e) Take μ^* to be a policy satisfying $T_{\mu^*} J_\infty = J_\infty$, (i.e. $\mu^*(x)$ attains the minimum in the Bellman equation). Show that $J^{\mu^*} \preceq J_\infty$, implying by part (c) that μ^* is optimal.

Hint: compare $T_{\mu^*}^k J_0$ *and* $T_{\mu^*}^k J_\infty$.

f) Bonus question *(we will not deduct points if you miss this)*:

Conclude that $J_\infty = J_{\mu^*}$ is the unique non-negative fixed point of the Bellman equation.