B9120-001 Dynamic Programming Winter 2020

Prof. Daniel Russo

Homework Assignment 6: Due Monday March 23

Approximate Fixed Point Operators.

Let $\mathcal{J} \subset \mathbb{R}^n$ and let $T : \mathcal{J} \to \mathcal{J}$ be a contraction operator with modulus α is some norm $\|\cdot\|$. That is, $\|TJ - TJ'\| \leq \alpha\|J - J'\|$ for all $J, J' \in \mathcal{J}$. Let J^* denote its unique fixed point. Let $\hat{T} : \mathcal{J} \to \mathcal{J}$ represent an approximation to T. Define the induced norm

$$\|\hat{T} - T\| = \sup_{J \in \mathcal{J}} \|\hat{T}J - TJ\|.$$

Part (a) Suppose $\hat{J} = \hat{T}\hat{J}$ is a fixed point of $\hat{T}$. Show

$$\|\hat{J} - J^*\| \leq \frac{\|\hat{T} - T\|}{1 - \alpha}$$

Part (b) No longer assume $\hat{T}$ has a fixed point. Consider the iteration $\hat{J}_{k+1} = \hat{T}\hat{J}_k$ for $k = 0, 1, 2, \cdots$. Show

$$\limsup_{k \to \infty} \|\hat{J}_k - J^*\| \leq \frac{\|\hat{T} - T\|}{1 - \alpha}$$

Remark: *For concreteness, one can have in mind for this problem that* $\mathcal{J} = \{J \in \mathbb{R}^n : \|J\|_\infty \leq \frac{M}{1-\alpha}\}$ *denotes a bounded subset of cost-to-go functions,* T *is the Bellman optimality operator and* $\hat{T}$ *is an approximation to* T *that is consistent with the apriori bounds on* J_μ*. For example,* $\hat{T}J(i) = \min_u \hat{g}(i,u) + \sum_j \hat{P}_{ij}(u)J(j)$ *could be a Bellman operator for an MDP whose transition probabilities were estimated from data.*

Solution

(a) We have

$$\begin{aligned}
\|\hat{J} - J^*\| &\leq \|\hat{J} - T\hat{J}\| + \|T\hat{J} - J^*\| \\
&= \|\hat{T}\hat{J} - T\hat{J}\| + \|T\hat{J} - TJ^*\| \\
&\leq \|\hat{T} - T\| + \alpha\|\hat{J} - J^*\|.
\end{aligned}$$

The result follows by rearranging terms.

(b) We have

$$\begin{aligned}\|\hat{J}_{k+1} - J^*\| &= \|\hat{T}\hat{J}_k - TJ^*\| \\ &\leq \|\hat{T}\hat{J}_k - T\hat{J}_k\| + \|T\hat{J}_k - TJ^*\| \\ &\leq \|\hat{T} - T\| + \alpha\|\hat{J}_k - J^*\| \\ &\leq \cdots \\ &\leq \alpha^k\|\hat{J}_0 - J^*\| + \sum_{i=0}^{k}\alpha^i\|\hat{T} - T\| \to \frac{\|\hat{T} - T\|}{1-\alpha}.\end{aligned}$$

Asynchronous value iteration

Consider the iteration

$$J_{k+1}(x) \leftarrow \begin{cases} TJ_k(x) & \text{if } x = x_k \\ J_k(x) & \text{if } x \neq x_k \end{cases}$$

for some sequence of states $(x_i : i = 0, 1, 2, \cdots)$. Suppose the state space is finite and each state is updated infinitely often. Show that $\|J_k - J^*\|_\infty \to 0$.

Solution

Define the lower and upper bounds L_0 and U_0 on the cost-to-go function

$$L_0 := J^* - \|J_0 - J^*\|_\infty e \preceq J_0 \preceq J^* + \|J_0 - J^*\|_\infty e := U_0$$

where e denotes a vector of ones. Consider applying asynchronous value iteration with initial cost-to-go functions L_0 and U_0, producing the estimates

$$L_{k+1}(x) \leftarrow \begin{cases} TL_k(x) & \text{if } x = x_k \\ L_k(x) & \text{if } x \neq x_k \end{cases} \qquad U_{k+1}(x) \leftarrow \begin{cases} TU_k(x) & \text{if } x = x_k \\ U_k(x) & \text{if } x \neq x_k \end{cases}$$

By the monotonicity of the Bellman operator, we can see that $L_k \preceq J_k \preceq U_k$ and $L_k \preceq J^* \preceq U_k$ hold for all k.

To show the result, we show that at the first time K_m at which every state has been updated m times we have

$$T^m L_0 \preceq L_{K_m} \preceq J_{K_m} \preceq U_{K_m} \preceq T^m U_0. \tag{1}$$

Taking $m \to \infty$ yields the result.

To show this, we focus on the time K_1 and the upper bounding sequence $(U_k : k = 0, 1 \cdots)$. We have $TU_1 = J^* + \alpha \|J_0 - J^*\|_\infty e \preceq U_0$. Therefore, $U_1(x) = U_0(x)$ for $x \neq x_0$ and, for the updated state x_0, we have

$$U_1(x_0) = TU_0(x_0) \leq U_0(x_0).$$

Now, since $U_1 \preceq U_0$, we have $U_2(x) = U_1(x)$ for $x \neq x_0$ and

$$U_2(x_1) = TU_1(x_1) \leq TU_0(x_1).$$

In words, when a state x is updated for the first time, its value decreases below $TU_0(x)$. Each subsequent period in which x is updated, its value only decreases. Proceeding in this fashion, we can see that at the first time K_1 at which every state has been updated at least once, we have $J_{K_1} \preceq U_{K_1} \preceq TU_0$. The result follows by inductively repeating the same argument.

Comment: *An alternative proof defines an operator T_i that updates only state x_i. We observe that this operator is non-expansive and that $T_{k_1} \cdots T_0$ is a contraction. The proof we gave highlights monotonicity properties instead.*