MS-E2191 Graduate Seminar on Operations Research
Fall 2020
Homework 11

MODEL SOLUTION

Please fill in the missing parts of the proof below.

Proposition: Convergence of the DP Algorithm. For any bounded function $J: X \to \mathbb{R}$, we have

$$J^*(x) = \lim_{N\to\infty} (T^N J)(x), \quad \forall x \in X.$$

Proof: For every positive integer N, initial state $x_0 \in X$, and optimal policy $\pi = \{\mu_0, \mu_1, \dots\}$, we break the cost into portions acquired from the first N stages and from the remaining stages

$$J_\pi(x_0) = \lim_{K\to\infty} E\{\sum_{k=0}^{K} \alpha^k g(x_k, \mu_k(x_k), w_k)\}$$

$$= E\{\sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k)\} + \lim_{K\to\infty} E\{\sum_{k=N}^{K} \alpha^k g(x_k, \mu_k(x_k), w_k)\}.$$

By assuming that the cost is bounded, $|g(x_k, \mu_k(x_k), w_k)| \le M$, we obtain

$$|\lim_{K\to\infty} E\{\sum_{k=N}^{K} \alpha^k g(x_k, \mu_k(x_k), w_k)| \le M \sum_{k=N}^{\infty} \alpha^k = \frac{\alpha^N M}{1-\alpha}.$$

Using the relations above it follows that

$$J_\pi(x_0) - \frac{\alpha^N M}{1-\alpha} - \alpha^N \max_{x\in X} |J(x)|$$

$$\le E\{\alpha^N J(x_N) + \sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k)\}$$

$$\le J_\pi(x_0) + \frac{\alpha^N M}{1-\alpha} + \alpha^N \max_{x\in X} |J(x)|.$$

By taking the minimum over π, we obtain for all x_0 and N

$$J^*(x_0) - \frac{\alpha^N M}{1-\alpha} - \alpha^N \max_{x\in X} |J(x)|$$

$$\le (T^N J)(x_0)$$

$$\le J^*(x_0) + \frac{\alpha^N M}{1-\alpha} + \alpha^N \max_{x\in X} |J(x)|.$$

By taking the limit $N \to \infty$, the result follows. **Q.E.D.**