Lecture 13 – Deteriorating Tool Example

A tool **deteriorates** stochastically with states $\mathrm{S} = \{0,1,2,\ldots\}$ with a theoretically unbounded state space.The decision maker chooses from actions $\mathrm{A} = \{\mathrm{D}, \mathrm{R}\}$, to replace the tool (R) or defer replacement (D). The tool deteriorates by i states with probability $\mathrm{p(i)}$ at each stage. Action R returns the tool to the ideal state 0.

Transition probabilities:

$$\mathrm{p_t}(j|s,D) = \begin{cases} 0, & j < s \\ p(j-s), & j \geq s \end{cases}$$

Where 0 for the case $j < s$ implies the tool can only stay in the same state or decay to a worse state and $p(j-s)$ implies the transition probability depends only on the number of states by which the tool deteriorates and

$\mathrm{p_t}(j|s,R) = p(j)$, $j \geq 0$.

Rewards:

$$\mathrm{r_t}(s,a) = \begin{cases} W - h(s), & a = D \\ W - K, & a = R \end{cases}$$

Where W is a fixed reward for each epoch, K is a fixed cost of replacement, and h(s) is a state dependent nonnegative maintenance cost. The salvage value at the end of period N is $\mathrm{R_N(s)}$.

Exercise: Using Theorem 4.7.5 provide conditions under which there exists an optimal policy that is monotone.

To complete this exercise you must experiment with alternative choices of conditions that could lead to satisfaction of the conditions in Theorem 4.7.5. The following proposition identifies specific conditions for which the conditions of Theorem 4.7.5 can be shown to hold.

Proposition: There exists an optimal monotone policy if $h(s)$ is nondecreasing and $R_N(s)$ is nonincreasing.

Proof: We first note that the conditions in the proposition guarantee that conditions 1 and 5 of Theorem 4.7.5 hold (in fact this is what motivates proposing these conditions). Next, we consider condition 2, that $q_t(k|s,a)$ is nondecreasing in $s, \forall k, a$. The case of action R holds trivially. For action D:

$$\Delta q_t = q_t(k|s+1,D) - q_t(k|s,D) = \begin{cases} \sum_{j=k}^{\infty}\big(p(j-s-1) - p(j-s)\big) = p(k-s-1) & if\ k > s \\ 0 & if\ k \leq s \end{cases}$$

Therefore $\Delta q_t \geq 0$.

Next, consider condition 3 that $r_t(s,a)$ is superadditive. This follows because:

$$r_t(s+1,R) + r_t(s,D) = W - K + W - h(s) \geq r_s(s+1,D) + r_t(s,R) = W - h(s+1) + W - K$$

Which follows by the assumption in the proposition that $h(s)$ is nondecreasing and by Lemma 4.7.6 in Puterman for the case of a reward function with only two actions.

Finally we need to show that condition 4 is satisfied, i.e., that $\sum_{j=0}^{\infty} p_t(j|s,a)v_t(j)$ is superadditive. By proposition 4.7.3 $v_t(s)$ is nonincreasing in s. Condition 4 implies the following:

$$\sum_{j=0}^{\infty} p_t(j)v_t(j) + \sum_{j=s}^{\infty} p_t(j-s)v_t(j) \geq \sum_{j=0}^{\infty} p_t(j)v_t(j) + \sum_{j=s+1}^{\infty} p_t(j-s-1)v_t(j)$$

This is true if

$$\sum_{j=s}^{\infty} p_t(j-s)v_t(j) \geq \sum_{j=s+1}^{\infty} p_t(j-s-1)v_t(j) \geq 0$$

Reorganizing the sums yields:

$$\sum_{j=0}^{\infty} p_t(j)(v_t(j+s) - v_t(j+s+1)) \geq 0$$

which follows from the fact that $v_t(s)$ is nonincreasing.