Lecture 06

Saturday, 14 January 2023 11:09 AM

* Policy Evaluation

Given a policy $\pi = \{\mu_0, \dots, \mu_{N-1}\}$, the value function of π is defined as

$$J^{\pi}(x_0) = \mathbb{E}_{w_k;\, k=0,1,\dots,N-1}\left[g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k) \right]$$

So $J^{\pi} : S_0 \longrightarrow \mathbb{R}$ takes the total cost if policy π is applied at initial state x_0.

Corollary 1 : For every initial state x_0, the value of policy π, $J^{\pi}(x_0)$ of the basic problem is equal to $J_0(x_0)$ when the function is given by the last step of the following algorithm, which proceeds backward in time from period $N-1$ to period 0.

$$J_N(x_N) = g_N(x_N)$$

$$J_k(x_k) = \mathbb{E}_{w_k}\left[g_k(x_k, \mu_k(x_k), w_k) + J_{k+1}\big(f_k(x_k, \mu_k(x_k), w_k)\big) \right]$$

$$k = 0, 1, \dots, N-1$$

Proof: Just expand.

- This is policy evaluation algorithm.