Lecture Note 4

1 Average-cost Problems

In the average cost problems, we aim at finding a policy u which minimizes

$$J_u(x) = \limsup_{T\to\infty} \frac{1}{T} \mathrm{E}\left[\sum_{t=0}^{T-1} g_u(x_t) \,\Big|\, x_0 = 0\right]. \tag{1}$$

Since the state space is finite, it can be shown that the lim sup can actually be replaced with lim for any stationary policy. In the previous lectures, we first find the cost-to-go functions $J^*(x)$ (for discounted problems) or $J^*(x,t)$ (for finite horizon problems) and then find the optimal policy through the cost-to-go functions. However, in the average-cost problem, $J_u(x)$ does not offer enough information for an optimal policy to be found; in particular, in most cases of interest we will have $J_u(x) = \lambda_u$ for some scalar λ_u, for all x, so that it does not allow us to distinguish the value of being in each state.

We will start by deriving some intuition based on finite-horizon problems. Consider a set of states $\mathcal{S} = \{x_1, x_2, \ldots, x^*, \ldots, x_n\}$. The states are visited in a sequence with some initial state x, say

$$\underbrace{x, \ldots\ldots,}_{h(x)} \underbrace{x^*, \ldots\ldots,}_{\lambda_u^1} \underbrace{x^*, \ldots\ldots,}_{\lambda_u^2} x^*, \ldots\ldots,$$

Let $T_i(x), i = 1, 2, \ldots$ be the stages corresponding to the ith visit to state x^*, starting at state x. Let

$$\lambda_u^i(x) = \mathrm{E}\left[\frac{\sum_{t=T_i(x)}^{T_{i+1}(x)-1} g_u(x_t)}{T_{i+1}(x) - T_i(x)}\right]$$

Intuitively, we must have $\lambda_u^i(x)$ is independent of initial state x and $\lambda_u^i(x) = \lambda_u^j(x)$, since we have the same transition probabilities whenever we start a new trajectory in state x^*. Going back to observe the definition of the function

$$J^*(x,T) = \min_u \mathrm{E}\left[\sum_{t=0}^{T} g_u(x_t) \Big| x_o = x\right],$$

we conjecture that the function can be approximated as follows.

$$J^*(x,T) \approx \lambda^*(x)T + h^*(x) + o(T), \quad \text{as } T \to \infty, \tag{2}$$

Note that, since $\lambda^*(x)$ is independent of the initial state, we can rewrite the approximation as

$$J^*(x,T) \approx \lambda^* T + h^*(x) + o(T), \quad \text{as } T \to \infty. \tag{3}$$

where term $h^*(x)$ can be interpreted as a residual cost that depends on the initial state x and will be referred to as the *differential cost function*. It can be shown that

$$h^*(x) = \mathrm{E}\left[\sum_{t=0}^{T_1(x)-1} (g_{u^*}(x) - \lambda^*)\right].$$

We can now speculate about a version of Bellman's equation for computing λ^* and h^*. Approximating $J^*(x,T)$ as in (3, we have

$$J^*(x,T+1) = \min_a \left\{ g_a(x) + \sum_y P_a(x,y) J^*(y,T) \right\}$$

$$\lambda^*(T+1) + h^*(x) + o(T) = \min_a \left\{ g_a(x) + \sum_y P_a(x,y) \left[\lambda^* T + h^*(y) + o(T) \right] \right\}$$

Therefore, we have

$$\boxed{\lambda^* + h^*(x) = \min_a \left\{ g_a(x) + \textstyle\sum_y P_a(x,y) h^*(y) \right\}} \tag{4}$$

As we did in the cost-to-go context, we set

$$T_u h = g_u + P_u h$$

and

$$Th = \min_u T_u h.$$

Then,we have

Lemma 1 (Monotonicity) *Let $h \leq \bar{h}$ be arbitrary. Then $Th \leq T\bar{h}$. ($T_u h \leq T_u \bar{h}$)*

Lemma 2 (Offset) *For all h and $k \in \Re$, we have $T(h + ke) = Th + ke$.*

Notice that the contraction principle does not hold for $Th = \min_u T_u h$.

2 Bellman's Equation

From the discussion above, we can write the Bellman's equation

$$\lambda e + h = Th. \tag{5}$$

Before examining the existence of solutions to Bellman's equation, we show the fact that the solution of the Bellman's equation renders the optimal policy by the following theorem.

Theorem 1 *Suppose that λ^* and h^* satisfy the Bellman's equation. Let u^* be greedy with respect to h^*, i.e., $Th^* \equiv T_{u^*} h^*$. Then,*

$$J_{u^*}(x) = \lambda^*, \forall x,$$

and

$$J^*_{u^*}(x) \leq J_u(x), \forall u.$$

Proof: Let $u = (u_1, u_2, \dots)$. Let N be arbitrary. Then

$$\begin{aligned}
T_{u_{N-1}} h^* &\geq Th^* = \lambda^* e + h^* \\
T_{u_{N-2}} T_{u_{N-1}} h^* &\geq T_{u_{N-2}} (h^* + \lambda^* e) \\
&= T_{u_{N-2}} h^* + \lambda^* e \\
&\geq Th^* + \lambda^* e \\
&= h^* + 2\lambda^* e
\end{aligned}$$

Then

$$T_1 T_2 \cdots T_{N-1} h^* \geq N\lambda^* e + h^*$$

Thus,we have

$$\mathrm{E}\left[\sum_{t=0}^{N-1} g_u(x_t) + h^*(x_N)\right] \geq (N-1)\lambda^* e + h^*$$

By dividing both sides by N and take the limit as N approaches to infinity, we have[1]

$$J_u \geq \lambda^* e$$

Take $u = (u^*, u^*, u^*, \dots)$, then all the inequalities above become the equality. Thus

$$\lambda^* e = J_{u^*}.$$

□

This theorem says that, if the Bellman's equation has a solution, then we can get a optimal policy from it.

Note that, if (λ^*, h^*) is a solution to the Bellman's equation, then $(\lambda^*, h^* + ke)$ is also a solution, for all scalar k. Hence, if Bellman's equation in (5) has a solution, then it has infinitely many solutions. However, unlike the case of discounted-cost and finite-horizon problems, the average-cost Bellman's equation does not necessarily have a solution. In particular, the previous theorem implies that, if a solution exists, then the average cost $J_{u^*}(x)$ is the same for all initial states. It is easy to come up with examples where this is not the case. For instance, consider the case when the transition probability is an identity matrix, i.e., the state visits itself every time, and each state incurs different transition costs $g(\cdot)$. Then the average cost λ^* depends on the initial state, which is not the property of the average cost. Hence, the Bellman's equation does not always hold.

[1] Recall that $J_u(x) = \limsup_{N\to\infty} \mathrm{E}\left[\frac{1}{N}\sum_{t=0}^{N-1} g_u(x_t) \,\middle|\, x_0 = x\right]$.