Lecture 12

* Discounted problems with bounded cost per stage.

We assume bounded cost per stage.

Assumption 1: The cost per stage g satisfies

$$|g(x,u,w)| \leq M, \quad \text{for all } (x,u,w) \in S \times U \times D$$

where M is a scalar. Furthermore, $0 < \alpha < 1$.

Proposition 2 (Convergence of DP algorithm). For any bounded function $J: S \to \mathbb{R}$, the optimal cost function satisfies

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

Proof: For every positive integer K, initial state $x_0 \in S$, and policy $\pi = \{\mu_0, \mu_1, \dots\}$, we break down the cost $J_\pi(x_0)$ into the portion incurred over the first K stages and over the remaining stages.

$$J_\pi(x_0) = \lim_{N\to\infty} \mathbb{E}\left[\sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k)\right]$$

$$= \mathbb{E}\left[\sum_{k=0}^{K-1} \alpha^k g(x_k, \mu_k(x_k), w_k)\right]$$

$$+ \lim_{N\to\infty} \mathbb{E}\left[\sum_{k=K}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k)\right]$$

By assumption 1 we have

$$\left|\lim_{N\to\infty} \mathbb{E}\left[\sum_{k=K}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k)\right]\right| \leq M \sum_{k=K}^{\infty} \alpha^k = \frac{\alpha^K M}{1-\alpha}$$

$$\left|\lim_{N\to\infty} \mathbb{E}\left[\sum_{k=K} \alpha^k g(x_k, \mu_k(x_k), w_k)\right]\right| \leq M \sum_{k=K} \alpha^k = \frac{\alpha^K M}{1-\alpha}$$

Using the relations, it follows that

$$J_\pi(x_0) - \frac{\alpha^K M}{1-\alpha} - \alpha^K \max_{x\in S} |J(x)|$$
$$\leq \mathbb{E}\left[\alpha^K J(x_K) + \sum_{k=0}^{K-1} \alpha^k g(x_k, \mu_k(x_k), w_k)\right]$$
$$\leq J_\pi(x_0) + \frac{\alpha^K M}{1-\alpha} + \alpha^K \max_{x\in S} |J(x)|$$

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

$$J^*(x_0) - \alpha^K\left(\frac{M}{1-\alpha} + \max_{x\in S} |J(x)|\right)$$
$$\leq (T^K J)(x_0) \leq J^*(x_0) + \alpha^K\left(\frac{M}{1-\alpha} + \max_{x\in S} |J(x)|\right)$$

and by taking limit as $K\to\infty$, result follows. ▨

<u>Corollary 1</u>: For every stationary policy μ, the associated cost function satisfies

$$J_\mu(x) = \lim_{N\to\infty} (T_\mu^N J_\mu)(x) \quad \forall x \in S$$

The next proposition shows that J^* is the unique sol^n of the Bellman's equation.

<u>Proposition 3</u> (Bellman's Equation). The optimal cost function J^* satisfies

$$J^*(x) = \min_{u\in U(x)} \mathbb{E}\left[g(x,u,w) + \alpha J^*(f(x,u,w))\right] \quad \forall x\in S,$$

or equivalently

$$J^* = TJ^*$$

Furthermore, J^* is the unique solution of this equation within

Furthermore, J^* is the unique solution of this equation within the class of bounded functions.

Proof: From proof of prop. 1 with $J_0 = 0$ we get

$$J^*(x) - \frac{\alpha^N M}{1-\alpha} \leq (T^N J_0)(x) \leq J^*(x) + \frac{\alpha^N M}{1-\alpha}$$

Appling the map T and using Lemma 1 and Lemma 2,

$$(TJ^*)(x) - \frac{\alpha^{N+1} M}{1-\alpha} \leq (T^{N+1} J_0)(x) \leq (TJ^*)(x) + \frac{\alpha^{N+1} M}{1-\alpha}$$

By taking limit $N \to \infty$ and using prop 1, we get

$$(TJ^*)(x) = J^*(x) \quad \forall x \in S$$

$$\Rightarrow TJ^* = J^*.$$

To show uniqueness, observe that if J is bounded and satisfies $J = TJ$ then

$$J = \lim_{N\to\infty} T^N J.$$

So by prop 1, $J = J^*$. ▨

Corollary 2: For every stationary policy μ the associated cost function satisfies

$$J_\mu(x) = \mathbb{E}_w\left[g(x,\mu(x),w) + \alpha J_\mu\big(f(x,\mu(x),w)\big)\right], \ \forall x \in S$$

or, equivalently,

$$J_\mu = T_\mu J_\mu$$

Furthermore, J_μ is the unique solution of this equation within the class of bounded functions.

Proposition 4 (Necessary and Sufficient condition for optimality).

Proposition 4 (Necessary and Sufficient condition for optimality).
A stationary policy μ is optimal if and only if $\mu(x)$ attains the minimum in Bellman's equation for each $x \in S$, i.e.,

$$\min_{u \in U(x)} E\left[g(x,u,w) + \alpha J^*(f(x,u,w))\right]$$
$$= E\left[g(x,\mu(x),w) + \alpha J^*(f(x,\mu(x),w))\right]$$

or equivalently

$$TJ^* = T_\mu J^*.$$

Proof: If $TJ^* = T_\mu J^*$, then using Bellmans equation, we have

$$J^* = TJ^* = T_\mu J^*$$

So by uniqueness in cor. 2. we obtain $J^* = J_\mu$.

Conversely if $J^* = J_\mu$ then

$$J^* = T_\mu J^* = TJ^* \qquad \blacksquare$$

We will finally show convergence rate estimate in the next proposition.

Proposition 5. For any two bounded function $J: S \to \mathbb{R}$, $J': S \to \mathbb{R}$, and for all $k = 0, 1, \dots,$ there holds

$$\max_{x \in S} \left|(T^k J)(x) - (T^k J')(x)\right| \leq \alpha^k \max_{x \in S} |J(x) - J'(x)|.$$

Proof: Let $c = \max_{x \in X} |J(x) - J'(x)|$

Then we have $J(x) - c \leq J'(x) \leq J(x) + c$

Applying T^k in the relation and using lemma 1 & lemma 2, we obtain

$$(T^kJ)(x) - \alpha^k c \leq (T^kJ')(x) \leq (T^kJ)(x) + \alpha^k c$$

$$\Rightarrow \quad \left|(T^kJ)(x) - (T^kJ')(x)\right| \leq \alpha^k c \qquad \blacksquare$$

Remark: Take $J' = J^*$. We get

$$\max_{x\in S} \left|(T^kJ)(x) - J^*(x)\right| \leq \alpha^k \max_{x\in S} \left|J(x) - J^*(x)\right|$$

* The same result holds for policy evaluation too.

Corollary 3: For any two bounded function $J: S \to \mathbb{R}$, $J': S \to \mathbb{R}$, and any stationary policy μ, we have

$$\max_{x\in S} \left|(T_\mu^kJ)(x) - (T_\mu^kJ')(x)\right| \leq \alpha^k \max_{x\in S} \left|J(x) - J'(x)\right|, \; k \in \mathbb{Z}_+$$