Chapter 1

Iteration Methods

1.1 Convergence of Value Iteration

In this section, the convergence of value iteration for the simple discounted problem is proven. The proof is done similarly as in Bertsekas volume 1 p.298-299.[Will be replaced by proper citation in report] The Bellman equations that we are going to use, are

$$V(x_k) = \begin{cases} \min\{c + \alpha_\delta V(1), \alpha_\delta \mathbb{E}[V(S(x_k))]\}, & \text{if } x_k > 0 \\ c + a + \alpha_\delta V(1), & \text{else.} \end{cases} \tag{1.1}$$

Where $\mathbb{E}[V(S(x_k))] = \mathbb{P}(\omega_k = 0)V(0) + \mathbb{P}(\omega_k = 1)V(x_k + 1)$.

We define an operator T. Such that at each iteration the value function is updated as follows

$$V_{k+1}(x) = TV_k = \begin{cases} \min\{c + \alpha_\delta V_k(1), \alpha_\delta \mathbb{P}(\omega(x) = 0)V_k(0) + \alpha_\delta \mathbb{P}(\omega(x) = 1)V_k(x+1)\}, & \text{if } x > 0 \\ c + a + \alpha_\delta V_k(1), & \text{else.} \end{cases} \tag{1.2}$$

And the corresponding policy is given by

$$\mu_{k+1}(x) = \begin{cases} 1, & \text{if } x = 0 \text{ or} \\ & c + \alpha_\delta V_k(1) < \alpha_\delta \mathbb{P}(\omega(x) = 0)V_k(0) + \alpha_\delta \mathbb{P}(\omega(x) = 1)V_k(x+1)\} \\ x + 1, & \text{else.} \end{cases} \tag{1.3}$$

For this T, it holds that for every V_a, V_b, the following property holds [Bertsekas]

$$\forall_x [V_a \leq V_b] \Rightarrow \forall_x [TV_a \leq TV_b].$$

Also, defining $e(x) = 1$, we have [Bertsekas]

$$T(V + Ce)(x_0) = TV(x_0) + \alpha_\delta C.$$

Furthermore, for every $\pi = \{\mu_0, \mu_1, ...\}$ with corresponding x_m, we have

$$\begin{aligned} V_\pi(x_0) &= \sum_{m=0}^{\infty} \alpha_\delta^m g(x_m, \mu_m(x_m)) \\ &\leq \sum_{k=0}^{\infty} \alpha_\delta^k (c+a) \\ &= \frac{c+a}{1-\alpha_\delta}. \end{aligned} \tag{1.4}$$

Moreover, for the optimal cost V^* we have that for all x_0 $V^*(x_0) \leq V_\pi(x_0) \leq \frac{c+a}{1-\alpha_\delta}$. Also, $TV^* = V^*$ because V^* satisfies the Bellman equations.

Now we choose an initial value V_0 such that there exists an M such that for each x, $|V_0(x)| \leq M$ holds. The following inequality now holds. We can now write

$$V^*(x_0) - \frac{c+a}{1-\alpha_\delta} \leq V_0(x_0) \leq M + V^*(x_0). \tag{1.5}$$

If we apply T k times to this equation and let $k \to \infty$, we get

$$\begin{aligned} &T^k(V^*(x_0) - \frac{c+a}{1-\alpha_\delta}) \\ &= V^*(x_0) - \alpha_\delta^k \frac{c+a}{1-\alpha_\delta} \\ &\leq T^k V_0(x_0) \\ &= V^k(x_0) \\ &\leq T^k(M + V^*(x_0)) \\ &\leq \alpha_\delta^k M + V^*(x_0). \end{aligned} \tag{1.6}$$

And in the limit, we get $V^k(x_0) \to V^*(x_0)$. Such that the convergence is proven for all bounded positive V_0.

1.2 Convergence of custom iteration

The convergence of the iteration method for the simple discounted problem will now be proven. The following equation will be solved

$$V(\mu^*) = \inf_{\mu>0} \mathbb{P}(Q > \mu)e^{-\beta\mu}(c+V(\mu^*)) + \mathbb{P}(Q \leq \mu)\mathbb{E}[e^{-\beta Q}|Q \leq \mu](c+a+V(\mu^*))]. \tag{1.7}$$

Note that these are not the Bellman equations since the discount depends on the chosen action. Let $\alpha_\mu = \mathbb{P}(Q > \mu)e^{-\beta\mu}\mathbb{P}(Q \leq \mu)\mathbb{E}[e^{-\beta Q}|Q \leq \mu]$ denote the discount when choosing control limit μ. Note that this is decreasing in μ since $\frac{d}{d\mu}\alpha_\mu = -\beta\mathbb{P}(Q > \mu)e^{-\beta\mu} < 0$. Since $\lim_{\mu\to 0} V(\mu) = \infty$, we know that for every $B > 0$ for sufficiently small ε, we have $\mu < \varepsilon \Rightarrow V(\mu) > B$.

The cost that is incurred when a control μ is chosen equals

$$g(\mu) := \mathbb{P}(Q > \mu)e^{-\beta\mu}c + \mathbb{P}(Q \leq \mu)\mathbb{E}[e^{-\beta Q}|Q \leq \mu](c+a).$$

Note that $g(\mu) < c + a$ for all μ so that

$$V(\mu^*) = \sum_{n=0}^{\infty} \alpha_{\mu^*}^k g(\mu^*) \leq \sum_{n=0}^{\infty} \alpha_\varepsilon^k (c + a) = \frac{c + a}{1 - \alpha_\varepsilon}$$

The iteration is given by

$$V_{n+1} = TV_n = \inf_{\mu_{n+1} > 0} g(\mu_{n+1}) + \alpha_{\mu_{n+1}} V_n \tag{1.8}$$

By $\mu(V)$ we will denote the μ at which TV is attained. We will prove that $T(A_1 + A_2) \leq TA_1 + \alpha_\varepsilon A_2$,$T(A_1) \leq T(A_2)$ and $T(A_1 - A_2) \geq TA_1 - \alpha_\varepsilon A_2$ for $A_1 \geq A_2$:

- $T(A_1+A_2) = g(\mu(A_1+A_2))+\alpha_{\mu(A_1+A_2)}(A_1+A_2) \leq g(\mu(A_1))+\alpha_{\mu(A_1)}(A_1+A_2) \leq g(\mu(A_1)) + \alpha_{\mu(A_1)} A_1 + \alpha_\varepsilon A_2 = TA_1 + \alpha_\varepsilon A_1$
- $T(A_1) = g(\mu(A_1)) + \alpha_{\mu(A_1)} A_1 \leq g(\mu(A_2)) + \alpha_{\mu(A_2)} A_1 \leq g(\mu(A_2)) + \alpha_{\mu(A_2)} A_2 = T(A_2)$
- $T(A_1 - A_2) = g(\mu(A_1 - A_2)) + \alpha_{\mu(A_1 - A_2)}(A_1 - A_2) \geq g(\mu(A_1 - A_2)) + \alpha_{\mu(A_1 - A_2)} A_1 - \alpha_\varepsilon A_2 \geq g(\mu(A_1)) + \alpha_{\mu(A_1)} A_1 - \alpha_\varepsilon A_2 = TA_1 - \alpha_\varepsilon A_2$

If our initial $V_0 < B$, then the following inequality now holds

$$V(\mu^*) - \frac{c + a}{1 - \alpha_\varepsilon} \leq V_0 \leq V_{\mu^*} + B. \tag{1.9}$$

If we now apply T k times on this inequality, we get

$$V_{\mu^*} - \alpha_\varepsilon^k \frac{c + a}{1 - \alpha_\varepsilon} \leq T^k (V_{\mu^*} - \frac{c + a}{1 - \alpha_\varepsilon}) \leq T^k V_0 = V_k \leq T^k (V_{\mu^*} + B) \leq V_{\mu^*} + \alpha_\varepsilon^k B. \tag{1.10}$$

Concluding $\lim_{k \to \infty} V_k = V_{\mu^*}$. So that the convergence for value iteration is proven. Note that the difficulty of this iteration still lies in finding the μ_{n+1} that minimizes (1.8). For increasing hazard rates, there is at most one μ such that

$$h(\mu) = \beta \frac{c + V_n}{a}.$$

And μ_{n+1} should be chosen as either this μ or ∞.