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

At each iteration the value function is updated as follows

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

And the corresponding policy is given by

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

Let $x_m^{(k)}$ $(m = 0, ..., k)$ be such that $x_0^{(k)} = x$ and $x_{m+1}^{(k)} = f(x_m^{(k)}, \mu_{k-m}, \omega(x_m^{(k)}))$. In this way, it holds that

$$\sum_{m=0}^{k-1} \alpha_\delta^m g(x_m^{(k)}, \mu_{k-m}(x_m^{(k)})) + \alpha_\delta^k V_0(x_k^{(k)}) = V_k(x).$$

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

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

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

$$\begin{aligned} &-\alpha^k M + V_\pi(x_0) \\ &\leq \mathbb{E}[\alpha_\delta^k V_0(x_k) + \sum_{m=0}^{k-1} \alpha_\delta^m g(x_m, \mu_m(x_m))] + \alpha_\delta^k \frac{c+a}{1-\alpha_\delta} \\ &\leq \alpha^k M + V_\pi(x_0). \end{aligned} \tag{1.5}$$

The expectation in the middle equals the value produced by the value iteration algorithm. If we choose the π that minimizes $V_\pi(x_0)$ will equal the optimal cost $V(x_0)$ and we have

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