# 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

$$V_{k+1}(x) = TV_k$$

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

And the corresponding policy is given by

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

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

$$(\forall_x : V_a \leq V_b) \Rightarrow (\forall_x : TV_a \leq TV_b). \tag{1.4}$$

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

$$T(V + Ce)(x_0) = TV(x_0) + \alpha_\delta C. \tag{1.5}$$

For every policy $\pi = \{\mu_0, \mu_1, ...\}$ with corresponding random variables $X_m$ ($m \geq 0$) such that $X_0 = x_0$ and $X_{m+1} = f(X_m, \mu_m(X_m), \omega_m)$ for $m \geq 0$, we define the random variable

$$V^\pi(x_0) = \sum_{m=0}^{\infty} \alpha_\delta^m g(X_m, \mu_m(X_m))$$

as the discounted cost starting from state $x_0$, using policy $\pi$. Since $g(x_m, u_m) \leq c + a$ for all $x_m, u_m$, we can write

$$\begin{aligned}
\mathbb{E}[V^\pi(x_0)] &= \mathbb{E}\left[\sum_{m=0}^{\infty} \alpha_\delta^m g(X_m, \mu_m(X_m))\right] \\
&= \sum_{m=0}^{\infty} \alpha_\delta^m \mathbb{E}[g(X_m, \mu_m(X_m))] \\
&\leq \sum_{m=0}^{\infty} \alpha_\delta^m (c+a) \\
&= \frac{c+a}{1-\alpha_\delta}.
\end{aligned} \tag{1.6}$$

Moreover, for the optimal expected discounted cost $V^*$ we have that for all $x_0$ $V^*(x_0) \leq \mathbb{E}[V^\pi(x_0)] \leq \frac{c+a}{1-\alpha_\delta}$. 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 $x$, $0 \leq V_0(x) \leq M$ holds. The following inequality now holds. We can now write for all $x_0$

$$V^*(x_0) - \frac{c+a}{1-\alpha_\delta} \leq 0 \leq V_0(x_0) \leq M \leq M + V^*(x_0). \tag{1.7}$$

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

$$\begin{aligned}
&T^k\left(V^*(x_0) - \frac{c+a}{1-\alpha_\delta}\right) \\
&= V^*(x_0) - \alpha_\delta^k \frac{c+a}{1-\alpha_\delta} \\
&\leq T^k V_0(x_0) \\
&= V_k(x_0) \\
&\leq T^k(M + V^*(x_0)) \\
&= \alpha_\delta^k M + V^*(x_0).
\end{aligned} \tag{1.8}$$

Where the first and last equalities follows from (1.5), the second equality from the definition of $T$ and the two inequalities follow from (1.4). Letting $k \to \infty$, 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. Let $V^\mu$ be the total discounted cost of the policy corresponding to repairing the machine when it has lived a time equal to control limit $\mu$.

Since this value is finite for every control limit $\mu > 0$, some $\mu^*$ must exist that minimizes this cost. For this $\mu^*$, the following equation must hold

$$V^{\mu^*} = \inf_{\mu>0} \mathbb{P}(Q > \mu)e^{-\beta\mu}(c + V^{\mu^*}) + \mathbb{P}(Q \leq \mu)\mathbb{E}[e^{-\beta Q}|Q \leq \mu](c + a + V^{\mu^*})]. \tag{1.9}$$

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 factor at which the costs for the next stage are discounted when choosing control limit $\mu$. The cost that is incurred when a control $\mu$ is chosen equals

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

We can now write

$$V^{\mu^*} = \sum_{n=0}^{\infty} \alpha_{\mu^*}^k g(\mu^*).$$

Note that $\alpha_\mu$ is decreasing in $\mu$ 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 $\varepsilon$, we have $\mu < \varepsilon \Rightarrow V^\mu > B$.

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

$$V_{n+1} = TV_n = \inf_{\mu_{n+1}>0} \left\{g(\mu_{n+1}) + \alpha_{\mu_{n+1}} V_n\right\} \tag{1.10}$$

By $\mu(V)$ we will denote the $\mu$ at which $TV$ is attained. For this $T$ we will prove the following properties:

**Lemma 1.** For $A_1, A_2$ such that $\frac{1}{2}B > A_1 \geq A_2 \geq 0$:

1. $T(A_1 + A_2) \leq TA_1 + \alpha_\varepsilon A_2$,
2. $T(A_1) \geq T(A_2)$,
3. $T(A_1 - A_2) \geq TA_1 - \alpha_\varepsilon A_2$.

**Proofs:**

1.

$$\begin{aligned} T(A_1 + A_2) &= g(\mu(A_1 + A_2)) + \alpha_{\mu(A_1+A_2)}(A_1 + A_2) \\ &\leq g(\mu(A_1)) + \alpha_{\mu(A_1)}(A_1 + A_2) \\ &\leq g(\mu(A_1)) + \alpha_{\mu(A_1)}A_1 + \alpha_\varepsilon A_2 \\ &= TA_1 + \alpha_\varepsilon A_1 \end{aligned} \tag{1.11}$$

where the first inequality follows from the fact that $\mu(A_1 + A_2)$ minimizes $g(\mu) + \alpha_\mu(A_1 + A_2)$ and the second from the fact that $a_\varepsilon > a_{\mu(A_1+A_2)}$.

2.

$$
\begin{aligned}
T(A_2) &= g(\mu(A_2)) + \alpha_{\mu(A_2)} A_2 \\
&\leq g(\mu(A_1)) + \alpha_{\mu(A_1)} A_2 \\
&\leq g(\mu(A_1)) + \alpha_{\mu(A_1)} A_1 \\
&= T(A_1)
\end{aligned}
\tag{1.12}
$$

where the first inequality follows from the fact that $\mu(A_2)$ minimizes $g(\mu)+\alpha_\mu A_2$ and the second from $A_1 \geq A_2$.

3.

$$
\begin{aligned}
T(A_1 - A_2) &= g(\mu(A_1 - A_2)) + \alpha_{\mu(A_1-A_2)}(A_1 - A_2) \\
&\geq g(\mu(A_1 - A_2)) + \alpha_{\mu(A_1-A_2)} A_1 - \alpha_\varepsilon A_2 \\
&\geq g(\mu(A_1)) + \alpha_{\mu(A_1)} A_1 - \alpha_\varepsilon A_2 \\
&= TA_1 - \alpha_\varepsilon A_2
\end{aligned}
\tag{1.13}
$$

where the first inequality follows from $a_\varepsilon > a_{\mu(A_1-A_2)}$ and the second from the fact that $\mu(A_1)$ minimizes $g(\mu) + \alpha_\mu A_1$.

If our initial $0 \leq V_0 < B$, then the following inequality now holds

$$
V^{\mu^*} - \frac{c+a}{1-\alpha_\varepsilon} \leq 0 \leq V_0 \leq B \leq V^{\mu^*} + B.
$$

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

$$
V^{\mu^*} - \alpha_\varepsilon^k \frac{c+a}{1-\alpha_\varepsilon} \leq T^k\left(V^{\mu^*} - \frac{c+a}{1-\alpha_\varepsilon}\right) \leq T^k V_0 = V_k \leq T^k(V^{\mu^*} + B) \leq V^{\mu^*} + \alpha_\varepsilon^k B.
\tag{1.14}
$$

Where the first and last inequalities follow from Lemma 1. 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 $\mu_{n+1}$ that minimizes (1.10). For increasing hazard rates, there is at most one $\mu$ such that

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

And $\mu_{n+1}$ should be chosen as either this $\mu$ or $\infty$.