Example – Cake eating problem (discounted costs)

Discounted objective:

$$\alpha^N g_N(x_N) + \sum_{k=0}^{N-1} \alpha^k g_k(x_k, u_k, w_k),$$

with discount factor $\alpha \in (0,1)$. DP algorithm can be written in an alternative form:

$$V_N(x_N) = g_N(x_N)$$

$$V_k(x_k) = \min_{u_k \in U_k(x_k)} E\left\{g_k(x_k, u_k, w_k) + \alpha V_{k+1}(f_k(x_k, u_k, w_k))\right\}$$

Imagine you bought a large cake that you want to eat. Also, you do not want eat it all in one sitting, rather you would like to eat it in N portions/days. However, as the days go by, the cake is getting a bit stale (it is a bit worse each day). What is the best way to divide the cake, so that your enjoyment is maximized?

x_0: Initial cake size,

N: Number of time steps (portions/days),

x_k: Cake size at time $k = 1, \ldots, N$,

u_k: Amount of eaten cake at time $k = 0, \ldots, N-1$,

α : Discount factor – cake getting stale.

System model:

$$x_{k+1} = x_k - u_k,$$

utility (negative because of the minimization) just from eating cake, no uncertainty:

$$g(x_k, u_k, w_k) = -2\sqrt{u_k},$$

terminal state has no value (throw out rest of the cake):

$$g_N(x_N) = 0.$$

Relationship:

$$V_k(x_k) = \min_{0 \le u_k \le x_k} [-2\sqrt{u_k} + \alpha V_{k+1}(x_k - u_k)]$$

Assignment: Compute the optimal control for $x_0 = 1, \alpha = 0.8, N = 3$.

Time $k = N$:

$$V_N(x_N) = g_N(x_N) = 0.$$

Time $k = N-1$:

$$V_{N-1}(x_{N-1}) = \min_{0 \le u_{N-1} \le x_{N-1}} [-2\sqrt{u_{N-1}} + \alpha \cdot 0]$$

$$u_{N-1} = x_{N-1}, \quad V_{N-1}(x_{N-1}) = -2\sqrt{x_{N-1}} \quad (\to x_N = 0).$$

Time $k = N-2$:

$$V_{N-2}(x_{N-2}) = \min_{0 \le u_{N-2} \le x_{N-2}} [-2\sqrt{u_{N-2}} + \alpha(-2\sqrt{x_{N-1}})]$$

$$V_{N-2}(x_{N-2}) = \min_{0 \le u_{N-2} \le x_{N-2}} [-2\sqrt{u_{N-2}} + \alpha(-2\sqrt{x_{N-2} - u_{N-2}})]$$

Let's differentiate! (maybe we get lucky and the minimizer will belong to the interval):

$$0 = -\frac{1}{\sqrt{u_{N-2}}} + \alpha \frac{1}{\sqrt{x_{N-2} - u_{N-2}}}$$

Moving around and squaring:

$$\frac{1}{u_{N-2}} = \alpha^2 \frac{1}{x_{N-2} - u_{N-2}},$$

Reciprocal and finish:

$$u_{N-2} = \frac{x_{N-2} - u_{N-2}}{\alpha^2} \quad \rightarrow \quad u_{N-2} = \frac{x_{N-2}}{\alpha^2 + 1}.$$

Since $0 < \alpha < 1$, the control is feasible (how lucky!) and

$$V_{N-2}(x_{N-2}) = -2\sqrt{\frac{x_{N-2}}{\alpha^2+1}} - 2\alpha\sqrt{x_{N-2}(1 - \frac{1}{\alpha^2+1})} = -2\sqrt{\frac{x_{N-2}}{\alpha^2+1}} - 2\alpha^2\sqrt{\frac{x_{N-2}}{\alpha^2+1}}$$

$$V_{N-2}(x_{N-2}) = -2\sqrt{\alpha^2+1}\sqrt{x_{N-2}}.$$

Time $k = N-3$:

$$V_{N-3}(x_{N-3}) = \min_{0 \le u_{N-3} \le x_{N-3}} [-2\sqrt{u_{N-3}} + \alpha(-2\sqrt{\alpha^2+1}\sqrt{x_{N-2}})]$$

Looks worse, but is exactly the same structure as V_{N-2}. To see this, set $\beta = \alpha\sqrt{\alpha^2+1}$, then

$$V_{N-3}(x_{N-3}) = \min_{0 \le u_{N-3} \le x_{N-3}} [-2\sqrt{u_{N-3}} + \beta(-2\sqrt{x_{N-3} - u_{N-3}})]$$

Using the results from $V_{N-2}(x_{N-2})$ we get:

$$u_{N-3} = \frac{x_{N-3}}{\beta^2+1}, \quad V_{N-3}(x_{N-3}) = -2\sqrt{\beta^2+1}\sqrt{x_{N-3}}.$$

From this, we can get the optimal policies and costs in the following way:

$$\begin{aligned}
&\gamma_0 = --, \quad u_N = -- \quad V_N = 0,\\
&\gamma_1 = 0, \quad u_{N-1} = \frac{x_{N-1}}{\gamma_1^2+1}, \quad V_{N-1} = -2\sqrt{\gamma_1^2+1}\sqrt{x_{N-1}},\\
&\gamma_2 = \alpha, \quad u_{N-2} = \frac{x_{N-2}}{\gamma_2^2+1} \quad V_{N-2} = -2\sqrt{\gamma_2^2+1}\sqrt{x_{N-2}},\\
&\gamma_3 = \alpha\sqrt{\gamma_2^2+1}, \quad u_{N-3} = \frac{x_{N-3}}{\gamma_3^2+1} \quad V_{N-3} = -2\sqrt{\gamma_3^2+1}\sqrt{x_{N-3}},\\
&\vdots\\
&\gamma_k = \alpha\sqrt{\gamma_{k-1}^2+1}, \quad u_{N-k} = \frac{x_{N-k}}{\gamma_k^2+1} \quad V_{N-k} = -2\sqrt{\gamma_k^2+1}\sqrt{x_{N-k}},\\
&\vdots\\
&\gamma_N = \alpha\sqrt{\gamma_{N-1}^2+1}, \quad u_0 = \frac{x_0}{\gamma_N^2+1} \quad V_0 = -2\sqrt{\gamma_N^2+1}\sqrt{x_0},
\end{aligned}$$

Stochastic element: a drunk roommate comes in the evening and eats a portion λ of the cake with probability p, or does not eat anything with probability $1-p$. The modified problem:

$$V_N(x_N) = 0, \quad x_{k+1} = x_k - u_k - w_k(x_k - u_k), \quad w_k = \begin{cases} \lambda, & \text{with probability } p,\\ 0, & \text{with probability } 1-p. \end{cases}$$

Then V_{N-1} and u_{N-1} remain the same

$$V_{N-1}(x_{N-1}) = \min_{0 \le u_{N-1} \le x_{N-1}} E\{-2\sqrt{u_{N-1}} + \alpha \cdot 0\} = -2\sqrt{x_{N-1}}.$$

But the rest differs:

$$
\begin{aligned}
V_{N-2}(x_{N-2}) &= \min_{0\le u_{N-2}\le x_{N-2}} E\{-2\sqrt{u_{N-2}} + \alpha(-2)\sqrt{(1-w_{N-2})(x_{N-2}-u_{N-2})}\} \\
&= \min_{0\le u_{N-2}\le x_{N-2}} \Big[p(-2\sqrt{u_{N-2}} - 2\alpha\sqrt{(1-\lambda)(x_{N-2}-u_{N-2})}) + (1-p)(-2\sqrt{u_{N-2}} - 2\alpha\sqrt{(x_{N-2}-u_{N-2})})\Big] \\
&= \min_{0\le u_{N-2}\le x_{N-2}} \Big[-2\sqrt{u_{N-2}} - 2\alpha(p\sqrt{(1-\lambda)(x_{N-2}-u_{N-2})} + (1-p)\sqrt{x_{N-2}-u_{N-2}})\Big] \\
&= \min_{0\le u_{N-2}\le x_{N-2}} \Big[-2\sqrt{u_{N-2}} - 2\alpha(p\sqrt{1-\lambda} + (1-p))\sqrt{x_{N-2}-u_{N-2}}\Big] \\
&\to u_{N-2}(x_{N-2}) = \frac{x_{N-2}}{\bar{\alpha}^2+1}, \quad V_{N-2}(x_{N-2}) = -2\sqrt{\bar{\alpha}^2+1}\sqrt{x_{N-2}}, \quad \bar{\alpha} = \alpha(p\sqrt{1-\lambda}+(1-p))
\end{aligned}
$$