Exercise 7.1 (student presents)

A man is fishing on his boat on a sea shore. We model his movement in discrete time. The location of the fisherman on period k is denoted by the state variable x_k. It defines how many sea miles to the right (positive value) or to the left (negative value), from the dock the fisherman is. On period $k = 0$, the fisherman starts from the dock. The fisherman knows that a school of trouts appears two sea miles to the right from the dock on period $k = 1$ and one sea mile to the right from the dock on period $k = 2$.

There is a heavy storm on the sea, which is affecting the fisherman's movement. The direction of the waves can change from period to period. Depending on the direction of the waves, they bring him either one sea mile to the right or left, with equal probability, on each period. This is modeled by a random variable w_k, which attains the values -1 and 1 with equal probabilities.

The fisherman's boat is hard to maneuver, thus on each period he can choose to stay on place or commit to moving one sea mile to the right or left. Thus, the control is also discrete: $u_k \in \{-1, 0, 1\}$ when $k = 0, 1$.

$$x_{k+1} = x_k + u_k + w_k \qquad k = 0, 1, \tag{1}$$

where the initial state is $x_0 = 0$.

The fisherman wants to be where the fish is on period $k = 1, 2$, and wants to minimize the effort it takes to steer the boat. Thus, his cost is

$$(x_2 - 1)^2 + (x_1 - 2)^2 + u_1^2 + u_0^2. \tag{2}$$

a) Which locations x_k can the fisherman reach with his boat on $k = 1, 2$?

b) Calculate the optimal cost-to-go $J_0(x_0)$ and optimal control policy $\{\mu_0^*(x_0), \mu_1^*(x_1)\}$.

c) If the fisherman knows that the school of trouts are 3 sea miles to the left of the dock on period $k = 0$, will it affect the optimal control policy? Hint: Add the term $(x_0 - 3)^2$ to the cost.

Solution

a) $k = 1$:

$$\begin{aligned}
\{u_0 = 1, w_0 = 1\} &\;\Rightarrow\; x_1 = 2\\
\{u_0 = 0, w_0 = 1\} &\;\Rightarrow\; x_1 = 1\\
\{u_0 = 1, w_0 = -1\} \vee \{u_0 = -1, w_0 = 1\} &\;\Rightarrow\; x_1 = 0\\
\{u_0 = 0, w_0 = -1\} &\;\Rightarrow\; x_1 = -1\\
\{u_0 = -1, w_0 = -1\} &\;\Rightarrow\; x_1 = -2
\end{aligned}$$

$k = 2$:

$$\begin{aligned}
x_1 = 2 &\;\Rightarrow\; x_2 = \{0, 1, 2, 3, 4\}\\
x_1 = 1 &\;\Rightarrow\; x_2 = \{-1, 0, 1, 2, 3\}\\
x_1 = 0 &\;\Rightarrow\; x_2 = \{-2, -1, 0, 1, 2\}\\
x_1 = -1 &\;\Rightarrow\; x_2 = \{-3, -2, -1, 0, 1\}\\
x_1 = -2 &\;\Rightarrow\; x_2 = \{-4, -3, -2, -1, 0\}
\end{aligned}$$

Thus $x_1 \in \{-2, -1, 0, 1, 2\}$ and $x_2 \in \{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$.

b) Lets calculate the cost of the final state ($k = 2$) $J_2(x_2) = (x_2 - 1)^2$ for all states x_2:

$$\begin{array}{rrr}
J_2(4) = 9 & J_2(3) = 4 & J_2(2) = 1\\
J_2(1) = 0 & J_2(0) = 1 & J_2(-1) = 4\\
J_2(-2) = 9 & J_2(-3) = 16 & J_2(-4) = 25.
\end{array}$$

Now the DP-algorithm can be used in period $k = 1$:

$$J_1(x_1) = \min_{u_1 \in \{-1,0,1\}} E_{w_1}\Big\{(x_1 - 2)^2 + u_1^2 + J_2(x_1 + u_1 + w_1)\Big\}$$

From this follows

$$\begin{aligned}
J_1(2) = 2 &\;\;\&\;\; \mu_1^*(2) = 0 \vee -1\\
J_1(1) = 2 &\;\;\&\;\; \mu_1^*(1) = 0\\
J_1(0) = 6 &\;\;\&\;\; \mu_1^*(0) = 0 \vee 1\\
J_1(-1) = 12 &\;\;\&\;\; \mu_1^*(-1) = 1\\
J_1(-2) = 22 &\;\;\&\;\; \mu_1^*(-2) = 1.
\end{aligned}$$

Similarly the DP-algoritm can be used in period $k = 0$:

$$J_0(x_0) = \min_{u_0 \in \{-1,0,1\}} E_{w_0}\Big\{u_0^2 + J_1(x_0 + u_0 + w_0)\Big\}$$

From this follows

$$J_0(0) = 5 \;\;\&\;\; \mu_0^*(0) = 1.$$

Thus, the fisherman should steer the boat one sea mile to the right in period $k = 0$, which will get him either to $x_1 = 0$ or $x_1 = 2$ in period $k = 1$. In both of those states, the optimal control is to stay on place.

c) This does not matter since the term $(x_0 - 3)^2$ is constant, i.e., the fisherman can not change his location x_0 on period $k = 0$.

Exercise 7.2 (solved in class)

Lets assume that the fisherman in Exercise 7.1 manages to upgrade his boat. His newer boat has an improved steering wheel, and hence his new control is continuous: $u_k \in [-1, 1]$ for all $k = 0, 1$. Calculate the optimal cost-to-go $J_1(x_1)$ and the optimal control policy $\mu_1^*(x_1)$ when $x_1 = 2, 1, -2$.

Solution

The cost in period $k = 2$ is $J_2(x_2) = (x_2 - 1)^2$. The DP-algorithm in period $k = 1$ is

$$\begin{aligned} J_1(x_1) &= \min_{u_1 \in [-1,1]} E_{w_1} \Big\{ (x_1 - 2)^2 + u_1^2 + J_2(x_1 + u_1 + w_1) \Big\} \\ &= \min_{u_1 \in [-1,1]} \Big\{ (x_1 - 2)^2 + u_1^2 + 0.5(x_1 + u_1 + 1 - 1)^2 + 0.5(x_1 + u_1 - 1 - 1)^2 \Big\} \end{aligned}$$

The minimum is found by setting the derivative of the right-side of the equation with regards to u_1 to zero:

$$4u_1 + 2x_1 - 2 = 0,$$

from which we get

$$u_1 = \frac{1 - x_1}{2} = \mu_1(x_1). \tag{3}$$

When $x_1 = 2$ and $x_1 = 1$, (3) is a feasible control. We get

$$\begin{aligned} \mu_1^*(x_1 = 2) = -0.5 \qquad & J_1(x_1 = 2) = 1.5 \\ \mu_1^*(x_1 = 1) = 0 \qquad & J_1(x_1 = 1) = 2 \end{aligned}$$

When $x_1 = -2$ the control (3) is not feasible. Because the function which is minimized is monotonically decreasing with regards to u_1 in the range $[-1, 1]$, $\mu_1(x_1 = -2) = 1$ should be chosen. Then $J_1(x_1 = -2) = 22$.

Exercise 7.3 (student presents)

The system is

$$x_{k+1} = x_k + u_k + w_k \qquad k = 0, 1, 2,$$

the initial state $x_0 = 0$, state constraints $x_k \in [-2, 4]$ for all k, and the stochastic term w_k attains the value 1 and -1 with equal probabilities for all k. The cost is

$$g_k(x_k, u_k) = \begin{cases} x_k^2 + u_k^2 & x_k \in [0, 2] \\ x_k^2 + u_k^2 + 13 & x_k < 0 \vee x_k > 2 \end{cases}$$

The final stage cost is $g_3(x_3) = 0$. Minimize the cost

$$\sum_{k=0}^{2} g_k(x_k, u_k)$$

and calculate the optimal control policy $u_k = \mu_k^*(x_k)$ with the following control constraints:

$$u_k \in \begin{cases} \{1\} & x_k < 0 \\ \{-1\} & x_k > 2 \\ \{-1, 0, 1\} & x_k = 0, 1, 2. \end{cases}$$

Solution

$k = 3$:

$$J_3(x_3) = g_3(x_3) = 0.$$

$k = 2$:

$$J_2(x_2) = \min_{u_2} E_{w_2}\Big[g_2(x_2, u_2) + J_3(x_3)\Big].$$

Tabulate values of J_2, with the smallest values **bolded**:

x_2/u_2	-1	0	1	$\mu_2^*(x_2)$
-2	-	-	**18**	1
-1	-	-	**15**	1
0	1	**0**	1	0
1	2	**1**	2	0
2	5	**4**	5	0
3	**23**	-	-	-1
4	**30**	-	-	-1

$k = 1$:

$$\begin{aligned} J_1(x_1) &= \min_{u_1} E_{w_1}\Big[g_1(x_1, u_1) + J_2(x_2)\Big] \\ &= \min_{u_1}\Big[g_1(x_1, u_1) + 0.5J_2(x_1 + u_1 + 1) + 0.5J_2(x_1 + u_1 - 1)\Big]. \end{aligned}$$

Tabulate values of J_1, with the smallest values **bolded**:

x_1/u_1	-1	0	1	$\mu_1^*(x_1)$
-2	-	-	**27**	1
-1	-	-	**23**	1
0	10	8	**3**	1
1	10	**3**	14	0
2	**7**	16	23	-1
(3	**35**	-	-	-1)
(4	**47**	-	-	-1)

(The states $x_1 = 3$ and $x_1 = 4$ can't be reached by starting from $x_0 = 0$.)

$k = 0$:

$$\begin{aligned} J_0(x_0) &= \min_{u_0} E_{w_0}\Big[g_0(x_0, u_0) + J_1(x_1)\Big] \\ &= \min_{u_0}\Big[g_0(x_0, u_0) + 0.5J_1(x_0 + u_0 + 1) + 0.5J_1(x_0 + u_0 - 1)\Big]. \end{aligned}$$

Now the state x_0 is set $x_0 = 0$, thus the minimal value $J_0 = 6$ is attained when $u_0 = \mu_0 = 1$.

Exercise 7.4 (solved in class)

The production process of a silicon wafer goes through two ovens U_1 and U_2. Let x_0 be the initial temperature of the silicon wafer, and x_k its temperature when it is moved out of oven U_k, $k = 1, 2$. The temperature of oven k is u_{k-1}. Lets assume that the temperature of the silicon wafer is modeled by the state equation

$$x_{k+1} = (1-a)x_k + au_k + w_k, \qquad k = 0, 1$$

where $a \in (0, 1)$ is some parameter of the process and w_k, $k = 0, 1$, are independent random variables with zero mean and finite variance. The aim is to get the temperature of the silicon wafer in the end of the process as close as possible to the target temperature T so that the energy of the process is minimized

$$r(x_2 - T)^2 + u_0^2 + u_1^2$$

where $r > 0$ is some constant. Solve how the oven U_2 should be warmed.

Solution

$k = 2$:

$$J_2(x_2) = r(x_2 - T)^2.$$

$k = 1$:

$$\begin{aligned} J_1(x_1) &= \min_{u_1} E_{w_1}\Big\{u_1^2 + J_2(x_2)\Big\} \\ &= \min_{u_1} E_{w_1}\Big\{u_1^2 + J_2\big((1-a)x_1 + au_1 + w_1\big)\Big\} \\ &= \min_{u_1} E_{w_1}\Big\{u_1^2 + r\big((1-a)x_1 + au_1 + w_1 - T\big)^2\Big\} \\ &= \min_{u_1} \Big[u_1^2 + r\big((1-a)x_1 + au_1 - T\big)^2 \\ &\quad +2rE\{w_1\}\big((1-a)x_1 + au_1 - T\big) + rE\{w_1^2\}\Big]. \end{aligned}$$

Because w_1 has zero mean, the third term inside the function to be minimized vanishes:

$$J_1(x_1) = \min_{u_1} \Big[u_1^2 + r\big((1-a)x_1 + au_1 - T\big)^2 + rE\{w_1^2\}\Big]. \tag{4}$$

The minimization problem is solved by setting the derivative to zero:

$$0 = 2u_1 + 2ra((1-a)x_1 + au_1 - T). \tag{5}$$

The optimal control $\mu_1^*(x_1) = u_1^*$ is

$$\mu_1^*(x_1) = \frac{ra(T-(1-a)x_1)}{1+ra^2}$$

Notice that the random variable term w_1 vanished when writing the equation (5).

Exercise 7.5 (teacher demo)

Solve how the oven U_1 in Exercise 7.4 should be warmed.

Solution

Insert $u_1^* = \mu_1^*(x_1)$ into the equation of $J_1(x_1)$ (which you have solved in Exercise 7.4), and you get

$$J_1^*(x_1) = E\{w_1^2\}r + \frac{a^2r^2(T-(1-a)x_1)^2}{(1+a^2r)^2} + r\Big((1-a)x_1 + \frac{a^2r(T-(1-a)x_1)}{1+a^2r} - T\Big)^2 \tag{6}$$

Mathematica can be handy here! Lets insert this into the cost-to-go function of period 0 $J_0(x_0) = \min_{u_0} E_{w_0}\{u_0^2 + J_1(x_1)\}$. Simplifying the function gives us the following:

$$\begin{aligned}J_0(x_0) = \min_{u_0} E_{w_0}\Big\{ & \tfrac{1}{1+a^2r}\Big(E\{w_1^2\}r(1+a^2r) + u_0^2 + r\big(T^2 + 2a^2u_0^2 - 2a^3u_0^2 + a^4u_0^2 + \\ & 2au_0w_0 - 4a^2u_0w_0 + 2a^3u_0w_0 + w_0^2 - 2aw_0^2 + a^2w_0^2 - \\ & 2(a-1)^3(au_0+w_0)x_0 + (a-1)^4x_0^2 + 2(a-1)T(au_0 + w_0 + x_0 - ax_0)\big)\Big)\Big\}\end{aligned} \tag{7}$$

How is the mean on the right-side of the equation (7) calculated? Lets make a few remarks:

1. the random variables w_1 and w_0 are independent of each other
2. the random variable w_0 exists in the terms of the function to be minimized in forms $(...)\cdot w_0$, $[(...)\cdot w_0 + ...]$, and $(...)\cdot w_0^2$, but not in the form $[(...)\cdot w_0 + ...]^n$, $n = 2, 3, ...$
3. all terms where u_0 and w_0 exist together are of 1. degree

Thus the mean operator $E_{w_0}\{\cdot\}$ can be neatly used on the right side of the equation (7). The terms $E_{w_0}\{w_0\}$ vanish because of their zero mean. Next we form the first order condition, where we derivate with respect to u_0; there the rest of the terms with w_0 (remark 3) and w_1 (remark 1 and 2) vanishes. The first order condition is then:

$$\begin{aligned} 0 &= \frac{1}{1+a^2r} 2\big(u_0 + a^2r(T + 2u_0 - 3x_0) + \\ &\quad a^4r(u_0 - x_0) + ar(x_0 - T) + a^3r(3x_0 - 2u_0)\big). \end{aligned}$$

This results into

$$u_0^* = \mu_0^*(x_0) = \frac{(a-1)ar\big((a-1)^2x_0 - T\big)}{1 + 2a^2r - 2a^3r + a^4r}.$$

Exercise 7.6 (homework)

Solve with dynamic programming and motivate your solution. You decide to start selling starter packs for freshmen (fuksi). You assume that the packages have demand on the first three days. The demand is 0, 1, 2 or 3 on the first day with probabilities 20%, 40%, 30% and 10%, respectively, and 10%, 20%, 30% and 40% for the next days. The probabilities of the different days are independent of each other.

You sell the packages for 5 euro each, and you can manufacture them on the morning of each day inside your van, which fits a maximum number of 3 packages (in reality one package equals tens of units). The production costs of the packages are 4, 7 and 10 euros (1, 2 and 3 packages).

a) How do you operate, and what is the expected maximum profit, when the remaining packages are practically worthless after the opening days?

b) What if you start selling one day earlier? Then there are four sales days. On this 'zeroth' day, the demand is the same as on the first day?