ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Institute for Dynamic Systems and Control
IDSC
Institut für Dynamische Systeme und Regelungstechnik

Final Exam **January 28th, 2010**

Dynamic Programming & Optimal Control (151-0563-00) Prof. R. D'Andrea

---

Solutions

Exam Duration: **150 minutes**

Number of Problems: **4**

Permitted aids: One A4 sheet of paper.

Use only the provided sheets for your solutions.

Problem 1 25%

Consider the dynamic system

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

where the state is constrained to be in the range $-1 \leq x_k \leq 3$ and the input u_k is restricted to be 1 or -1.

Given the initial state $x_0 = 1$, the goal is to minimize the cost

$$\underset{w_0, w_1}{E} \left\{ \sum_{k=0}^{1} (x_k + u_k + w_k)^2 \right\}.$$

The disturbance w_k takes the values 0 and 1. If $x_k \neq 0$, both values have equal probability. If $x_k = 0$, the disturbance w_k is 0 with probability 1.

Apply the Dynamic Programming algorithm to find the optimal control policy and the optimal cost $J_0(1)$.

Solution 1

The optimal control problem is considered over a time horizon $N = 2$ and the cost, to be minimized, is defined by

$$g_2(x_2) = 0 \qquad \text{and} \qquad g_k(x_k, u_k, w_k) = (x_k + u_k + w_k)^2, \quad k = 0, 1.$$

Note that the state x_k takes on only integer values since $w_k \in \{0, 1\}$, $u_k \in \{-1, 1\}$, and $x_0 = 1$.

The DP algorithm proceeds as follows:

2nd stage:
The initial condition for the Dynamic Programming recursion is

$$J_2(x_2) = 0$$

for all feasible $x_2 \in \{-1, 0, 1, 2, 3\}$.

1st stage:
Proceeding backwards, we get:

$$\begin{aligned} J_1(x_1) &= \min_{u_1 \in \{-1, 1\}} \underset{w_1}{E}\left\{(x_1 + u_1 + w_1)^2 + J_2(x_2)\right\} \\ &= \min_{u_1 \in \{-1, 1\}} \underset{w_1}{E}\left\{(x_1 + u_1 + w_1)^2\right\}. \end{aligned}$$

Given the initial condition $x_0 = 1$ and the input constraint $u_1 \in \{-1, 1\}$, we know from the deterministic dynamic relationship, $x_1 = x_0 + u_0$, that $x_1 \in \{0, 2\}$. That is, only the two values, $J_1(0)$ and $J_1(2)$, need to be considered:

$$\begin{aligned} J_1(0) &= \min_{u_1 \in \{-1, 1\}} \left[0 \cdot (0 + u_1 + 1)^2 + 1 \cdot (0 + u_1 + 0)^2\right] \\ &= 1, && \text{with} \quad \mu_1^*(0) = 1, -1. \\ J_1(2) &= \min_{u_1 \in \{-1, 1\}} \left[0.5 \cdot (2 + u_1 + 1)^2 + 0.5 \cdot (2 + u_1 + 0)^2\right\} \\ &= \min\left[2 + 0.5,\ 8 + 4.5\right] \\ &= 2.5, && \text{with} \quad \mu_1^*(2) = -1. \end{aligned}$$

0th stage:
Finally, the optimal cost is calculated for the initial condition $x_0 = 1$:

$$\begin{aligned} J_0(1) &= \min_{u_0 \in \{-1, 1\}} \underset{w_0}{E}\left\{(1 + u_0 + w_0)^2 + J_1(x_1)\right\} \\ &= \min_{u_0 \in \{-1, 1\}} \left[\underset{w_0}{E}\left\{(1 + u_0 + w_0)^2\right\} + J_1(1 + u_0)\right]. \\ &= \min_{u_1 \in \{-1, 1\}} \left[0.5 \cdot (1 + u_0 + 1)^2 + 0.5 \cdot (1 + u_0 + 0)^2 + J_1(1 + u_0)\right] \\ &= \min\left[0.5 + 0 + 1,\ 4.5 + 2 + 2.5\right] \\ &= 1.5, && \text{with} \quad \mu_0^*(1) = -1. \end{aligned}$$

To sum up, the optimal (expected) cost is $J_0(1) = 1.5$ and the optimal policy is to choose -1 at stage 0, whereas both input values, -1 and 1, lead to the same cost in stage 1.

Problem 2 25%

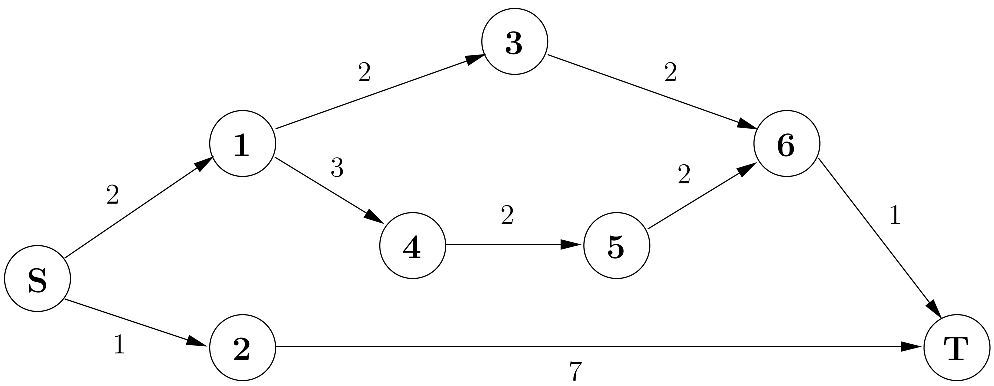


Figure 1

a) Find the shortest path from node S to node T for the graph given in Figure 1. Apply the *Label Correcting Method*. Use *Best-First Search* to determine at each iteration which node to remove from OPEN; that is, remove node i with

$$d_i = \min_{j \text{ in } \texttt{OPEN}} d_j,$$

where the variable d_i denotes the length of the shortest path from node S to node i that has been found so far.

Solve the problem by populating a table of the form given in Table 1. State the resulting shortest path and its length.

Iteration	Node exiting OPEN	OPEN	d_S	d_1	d_2	d_3	d_4	d_5	d_6	d_T
0	-	...								
...										

Table 1

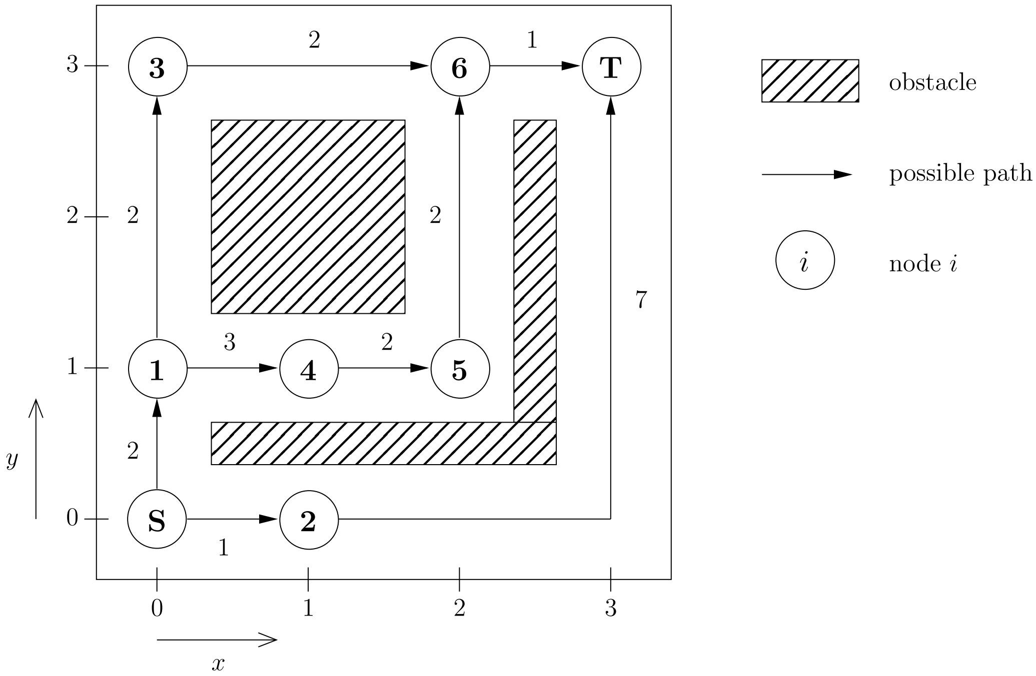
b) Assume the graph in Figure 1 originates from the map shown in Figure 2, where again the problem is to find the shortest path from node S to node T.

Traveling is only possible along the x- and y-axis, and the travel cost per unit distance is at least 1. Therefore, a *lower bound* on the cost to go from node i to node T is given by

$$l_i = |3 - x_i| + |3 - y_i|, \tag{1}$$

where (x_i, y_i) are the coordinates of node i.

Use the lower bound (1) to strengthen the condition on whether a node enters OPEN (this is known as the A^* *algorithm*). Use *Best-First Search* to determine at each iteration which node to remove from OPEN. Solve the problem by populating a table of the form given in Table 1.



3
2
1
0
2
2
7
2
1
3
2
2
1
3
6
T
1
4
5
S
2
0
1
2
3
x
y
obstacle
possible path
i
node i

Figure 2

Solution 2

a) For the first part, the (standard) Label Correcting method is applied: at each iteration the condition

$$d_i + a_{ij} < d_T$$

is used to decide if the node j enters the `OPEN` list.

Iteration	Node exiting `OPEN`	`OPEN`	d_S	d_1	d_2	d_3	d_4	d_5	d_6	d_T
0	–	S	0	∞	∞	∞	∞	∞	∞	∞
1	S	1,2	0	2	1	∞	∞	∞	∞	∞
2	2	1	0	2	1	∞	∞	∞	∞	8
3	1	3,4	0	2	1	4	5	∞	∞	8
4	3	4,6	0	2	1	4	5	∞	6	8
5	4	6,5	0	2	1	4	5	7	6	8
6	6	5	0	2	1	4	5	7	6	7
7	5	–	0	2	1	4	5	7	6	7

The shortest path is $S \to 1 \to 3 \to 6 \to T$ with a total length of 7.

b) The extra information, the lower bound l_i on the cost to go from state i to T, can be used to strengthen the condition on whether a node enters the `OPEN` bin:

$$d_i + a_{ij} + l_j < d_T,$$

which is known as the A^* *algorithm*. With this stricter condition, the state 4 and 5 can be excluded and do not enter `OPEN`.

Iteration	Node exiting `OPEN`	`OPEN`	d_S	d_1	d_2	d_3	d_4	d_5	d_6	d_T
0	–	S	0	∞	∞	∞	∞	∞	∞	∞
1	S	1,2	0	2	1	∞	∞	∞	∞	∞
2	2	1	0	2	1	∞	∞	∞	∞	8
3	1	3	0	2	1	4	∞	∞	∞	8
4	3	6	0	2	1	4	∞	∞	6	8
5	6	–	0	2	1	4	∞	∞	6	7

Problem 3 **25%**

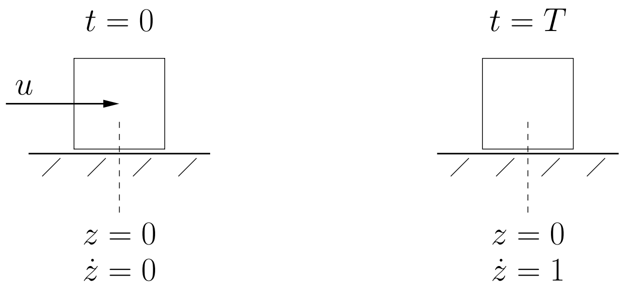


Figure 3

At time $t = 0$, a unit mass is at rest at location $z = 0$. The mass is on a frictionless surface and it is desired to apply a force $u(t)$, $0 \le t \le T$, such that at time $t = T$ the mass is at the same location $z = 0$, but with unit velocity $\dot{z} = 1$. The force is constrained by $-1 \le u(t) \le 1$. The objective is to perform the maneuver in minimum time.

The dynamics are given by

$$\ddot{z}(t) = u(t), \quad 0 \le t \le T,$$

with initial and terminal conditions:

$$\begin{aligned} z(0) &= 0, \quad \dot{z}(0) = 0, \\ z(T) &= 0, \quad \dot{z}(T) = 1. \end{aligned}$$

Find the optimal input $u(t)$ that minimizes the terminal time T,

$$T = \int_0^T 1 \; dt.$$

State the minimizing input $u^*(t)$ and the minimum time T^*.

Solution 3

Introducing the state vector

$$x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} z(t) \\ \dot{z}(t) \end{bmatrix},$$

the dynamics read

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix} = \begin{bmatrix} x_2(t) \\ u(t) \end{bmatrix}$$

with initial and terminal conditions,

$$\begin{aligned} x_1(0) &= 0, \quad x_2(0) = 0, \\ x_1(T) &= 0, \quad x_2(T) = 1. \end{aligned}$$

We apply the Minimum Principle.

- The Hamiltonian is given by

$$\begin{aligned} H(x, u, p) = &\ g(x, u) + p^T f(x, u) \\ &= 1 + p_1 x_2 + p_2 u. \end{aligned}$$

- The adjoint equations

$$\begin{aligned} \dot{p}_1(t) &= 0 \\ \dot{p}_2(t) &= -p_1(t) \end{aligned}$$

  are integrated and result in the following equations for the co-states:

$$\begin{aligned} p_1(t) &= c_1, && c_1 \text{ constant} \\ p_2(t) &= -c_1 t + c_2, && c_2 \text{ constant.} \end{aligned}$$

- The optimal input $u^*(t)$ is obtained by minimizing the Hamiltonian along the optimal trajectory

$$u^*(t) = \arg \min_{-1 \leq u \leq 1} \left(1 + p_1 x_2^* + p_2 u\right).$$

  Since the Hamiltonian is linear in u, the minimum is attained on the boundaries of the control space $C = [-1, 1]$. (The solution is thus called a *bang-bang* solution.) In particular,

$$u^*(t) = \begin{cases} -1 & \text{if } p_2(t) > 0 \\ 1 & \text{if } p_2(t) < 0 \\ \text{undefined} & \text{if } p_2(t) = 0. \end{cases}$$

  It remains to compute the switching conditions using knowledge about the co-states and initial and terminal conditions.

- Since $p_2(t)$ is affine-linear in t, $p_2(t)$ has at most one zero crossing. Therefore the input $u^*(t)$ switches at most once. From physical intuition, it is clear that the mass has to decelerate first ($\ddot{z} < 0$) and then (after the switch) accelerate ($\ddot{z} > 0$) so that it can end up at the same location as before with a positive velocity.[1] Therefore, we will integrate the system equation with input $u(t) = -1$ for $0 \le t < \tau$, where τ is the switching time to be determined, and $u(t) = 1$ for $\tau < t \le T$. We will then use initial and terminal conditions on x to determine the switching time τ and the terminal time T.

- For $0 \le t \le \tau$,

$$\begin{aligned}
\dot{x}_2 &= -1 &&\Rightarrow\quad x_2(t) = -t + c_3, && c_3 \text{ constant,}\\
\dot{x}_1 &= x_2 = -t + c_3 &&\Rightarrow\quad x_1(t) = -\tfrac{1}{2}t^2 + c_3 t + c_4, && c_4 \text{ constant.}
\end{aligned}$$

  Using $x_1(0) = x_2(0) = 0$, it follows that c_3 and c_4 are both 0. At the switching time τ the states are thus

$$x_1(\tau) = -\tfrac{1}{2}\tau^2 \quad \text{and} \quad x_2(\tau) = -\tau. \tag{2}$$

  For $\tau \le t \le T$,

$$\dot{x}_2 = 1 \quad\Rightarrow\quad x_2(t) = t + c_5, \qquad c_5 \text{ constant,} \tag{3}$$
$$\dot{x}_1 = x_2 = t + c_5 \quad\Rightarrow\quad x_1(t) = \tfrac{1}{2}t^2 + c_5 t + c_6, \qquad c_6 \text{ constant.} \tag{4}$$

  To ensure continuity of the states x (a discontinuity in position or velocity would not make physical sense), we require that the equations (3), (4) satisfy (2). It follows that $c_5 = -2\tau$ and $c_6 = \tau^2$. Furthermore, at time T, the terminal conditions $x_1(T) = 0$ and $x_2(T) = 1$ need to be satisfied, i. e.

$$x_2(T) = T - 2\tau = 1 \tag{5}$$
$$x_1(T) = \tfrac{1}{2}T^2 - 2\tau T + \tau^2 = 0. \tag{6}$$

  Solving (5) for τ and inserting this in (6), we obtain a quadratic equation for T

$$\begin{aligned}
\tfrac{1}{2}T^2 - 2\left(-\tfrac{1}{2} + \tfrac{T}{2}\right)T + \left(-\tfrac{1}{2} + \tfrac{T}{2}\right)^2 &= 0\\
T^2 - 2T - 1 &= 0,
\end{aligned}$$

  which is solved by

$$T = 1 \pm \sqrt{2}.$$

  Since $T > 0$, the minimum terminal time is $T^* = 1 + \sqrt{2}$. Using (5), we obtain the switching time $\tau = \frac{\sqrt{2}}{2}$, and thus, the optimal input

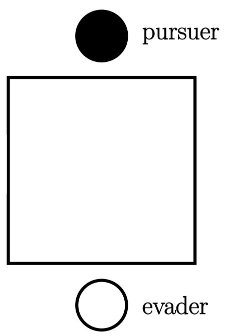
$$u^*(t) = \begin{cases} -1 & \text{if } 0 \le t < \frac{\sqrt{2}}{2} \\ 1 & \text{if } \frac{\sqrt{2}}{2} \le t < 1 + \sqrt{2}. \end{cases}$$

[1] Without this physical interpretation one could alternatively integrate the system equation with $u = c$ for $0 \le t < \tau$, where $c = \pm 1$, and with $u = -c$ for $\tau < t \le T$. One will then find that $c = -1$ is the only solution satisfying $\tau \le T$.

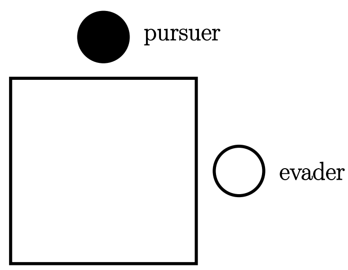
Problem 4 25%

Consider the following two-player game, played around a table with four corners. One player, the so-called pursuer, is attempting to catch the other player, called evader. The game evolves in stages where, in each stage, both players implement actions simultaneously. In each stage, the evader randomly moves one corner clockwise with probability p, one corner counter-clockwise with probability p, or stays where he is with probability $(1-2p)$, $p < 0.5$. In the situation when the players are across from one another (see Figure 4(a)), the pursuer has the option to stay where he is, move one corner clockwise, or move one corner counter-clockwise. When the two players are adjacent to one another (see Figure 4(b)), again, the pursuer decides whether to stay where he is, move toward the other's current location, or move away from the other's current location. The pursuer catches the evader only by arranging to land on the same side of the table as the evader at the end of a period. (The possibility exists that, when they are adjacent, they can both move toward each other's current location. This does not result in the evader being caught in "mid-air".)

The pursuer's objective is to capture the evader in minimum expected time. The game ends when the evader is caught. (There is no cost involved for moving around the table.)



(a) Across from one another.



(b) Adjacent to one another.

Figure 4: Table with the two players.

Formulate the problem as a stochastic shortest path problem:

$$x_{k+1} = w_k, \quad k = 0, 1, \dots ,$$

with $x_k, w_k \in S$, $u_k \in U(x_k)$ and transition probabilities $p_{ij}(u) = P(w_k = j \mid x_k = i, u_k = u)$, where the objective is to minimize

$$\lim_{N\to\infty} E\left\{ \sum_{k=0}^{N-1} g(x_k, u_k) \right\} .$$

That is, define the set of states S, the control sets $U(x_k)$, the corresponding transition probabilities $p_{ij}(u)$, and the stage costs $g(x_k, u_k)$. Explain your steps. **You do not have to solve the problem.**

Hint: Try to use as few states as possible.

Solution 4

Set of states:

$$S = \{0,\, 1,\, 2\}\,,$$

where 0 represents the termination state 'evader caught', 1 stands for 'players adjacent to one another', and 2 is 'players across from one another'.

Set of control sets:

$$\begin{aligned} U(1) &= \{0,\, 1,\, 2\}\,, \\ U(2) &= \{0,\, 1\}\,. \end{aligned}$$

Different input sets are defined for the two situations: adjacent and across from one another. If the two players are adjacent to one another, there are three options: 0 which means 'not move', 1 which denotes 'move towards evader', and 2 which is 'move away from evader'. If the two players are across from each other, we only have the input values 0 or 1. The input set $U(0)$ is not relevant for the problem solution and could be anything.

Transition probabilities:

$$\begin{aligned} p_{10}(0) &= p\,, \\ p_{10}(1) &= 1-2p\,, \\ p_{10}(2) &= 0\,, \\ p_{11}(0) &= 1-2p\,, \\ p_{11}(1) &= 2p\,, \\ p_{11}(2) &= 2p\,, \\ p_{12}(0) &= p\,, \\ p_{12}(1) &= 0\,, \\ p_{12}(2) &= 1-2p\,, \end{aligned}$$

$$\begin{aligned} p_{20}(0) &= 0\,, \\ p_{20}(1) &= p\,, \\ p_{21}(0) &= 2p\,, \\ p_{21}(1) &= 1-2p\,, \\ p_{22}(0) &= 1-2p\,, \\ p_{22}(1) &= p\,, \end{aligned}$$

and the probability of staying in the termination state is $p_{00}(u) = 1\,,\ \forall u \in U(0)$. All other probabilities are zero.

Cost:

$$\begin{aligned} g(i,u) &= c > 0 \qquad i = 1,2\,,\ u \in U(i)\,, \\ g(i,u) &= 0 \qquad\quad\;\; i = 0\,,\ u \in U(0)\,. \end{aligned}$$

Since the goal is to minimize the expected time for the pursuer to catch the evader, c can be any constant positive value but, for simplicity, might be chosen to be 1. As soon as the evader is caught, we assume zero cost to get a well-defined problem with finite overall cost.