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, 2021**

Dynamic Programming & Optimal Control (151-0563-01)

Prof. R. D'Andrea

Exam

Exam Duration:	**150 minutes**
Number of Problems:	**4**
Permitted aids:	One A4 sheet of paper. No calculators allowed.

Problem 1 [30 points]

a) Consider the system dynamics

$$x_{k+1} = x_k + u_k, \quad k = 0, ..., N-1, \quad N = 4,$$

with $x_0 = 0$. At each stage k, the control input u_k is in the set

$$\mathcal{U}_k = \{-1, 0, 1\}.$$

i) What is the number of reachable states, $|\mathcal{S}_3|$, at time step $k = 3$? *[1 point]*

ii) How many open loop control policies are there in this problem? *[2 points]*

(A) ☐ 4^3

(B) ☐ 7^3

(C) ☐ 3^7

(D) ☐ 3^4

(E) ☐ None of the above

iii) How many closed loop control policies are there in this problem? *[2 points]*

(A) ☐ 2^{10}

(B) ☐ 2^{30}

(C) ☐ 3^{16}

(D) ☐ 3^{10}

(E) ☐ None of the above

iv) How many closed loop control policies are there for a general N? *[2 points]*

(A) ☐ $3^{\sum_{i=0}^{N-1}(2i+1)}$

(B) ☐ $3^{\sum_{i=0}^{N} i}$

(C) ☐ $2^{\sum_{i=0}^{N} i}$

(D) ☐ $2^{\sum_{i=0}^{N} 3i}$

(E) ☐ None of the above

v) We want to minimize the following generic cost function

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

using the Dynamic Programming Algorithm. How many minimizations in u do we need to perform, from stage $k = 0$ to $k = N - 1$ included?

[2 points]

(A) ☐ $2^{\sum_{i=0}^{N} i}$

(B) ☐ $3^{\sum_{i=0}^{N} i}$

(C) ☐ $\sum_{i=0}^{N} i$

(D) ☐ $\sum_{i=0}^{N-1}(2i+1)$

(E) ☐ None of the above

b) Consider the system dynamics

$$x_{k+1} = x_k + u_k w_k + u_k x_k, \quad k = 0, \ldots, N-1,$$

with $x_k \in \mathbb{R}$ and $u_k \in [-2, 2]$ for each time step k. The disturbance w_k is a random variable with probability density function

$$p_{w_k}(\bar{w}) = \begin{cases} \frac{1}{2\alpha} + \frac{\bar{w}}{2\alpha^2} & \text{if} \quad -\alpha \leq \bar{w} \leq \alpha \\ 0 & \text{otherwise} \end{cases},$$

where $\alpha > 0$ is a parameter. The cost to be minimized is

$$x_N + \sum_{k=0}^{N-1} (x_k + u_k^2).$$

i) What is the expected value of w_k, as a function of the parameter α? *[2 points]*

(A) ☐ 2α

(B) ☐ $\frac{\alpha}{2}$

(C) ☐ $\frac{\alpha}{3}$

(D) ☐ $\frac{2}{3}\alpha$

(E) ☐ None of the above

ii) What are the optimal input and optimal cost to go at time step $N-1$, for the state $x_{N-1} = 1$, when $\alpha = 6$? *[3 points]*

$u_{N-1}(1) =$

$J_{N-1}(1) =$

iii) For what values of $\alpha > 0$ is the optimal control input at state $x_{N-1} = 0$ equal to -2? *[2 points]*

(A) ☐ $\alpha = 12$

(B) ☐ $0 < \alpha \leq 12$

(C) ☐ $\alpha \geq 12$

(D) ☐ $\alpha \geq \frac{1}{12}$

(E) ☐ None of the above

iv) For what values of $\alpha > 0$ is the optimal control input at state $x_{N-1} = 0$ equal to 2? *[2 points]*

(A) ☐ $\alpha = 12$

(B) ☐ $0 < \alpha \leq 12$

(C) ☐ $\alpha \geq 12$

(D) ☐ $\alpha \geq \frac{1}{12}$

(E) ☐ None of the above

v) Let us denote with $\mu^*_{N-1}(0)$ the optimal control input for state 0 at time step $N-1$. What is the maximum value that $\mu^*_{N-1}(0)$ can be equal to, for $\alpha > 0$? *[2 points]*

(A) ☐ 2

(B) ☐ -2

(C) ☐ 0

(D) ☐ $\frac{1}{12}$

(E) ☐ None of the above

vi) Suppose the system dynamics are now

$$x_{k+1} = x_k + u_k w_k, \quad k = 0, \ldots, N-1,$$

where the set of admissible control inputs is $\mathcal{U} = \mathbb{R}$, and the random variable w_k and the cost function are the same as defined before.

Can this problem be solved using forward Dynamic Programming Algorithm? Justify your answer. *[1 point]*

vii) Consider the modified problem described in question vi), with $\alpha = 3$. Let $J_{N-1}(x)$ represent the optimal cost to go for state x at time step $N-1$. Let $\mathrm{x}_1, \mathrm{x}_2$ be two reachable states at time step $N-1$. What is the value of $J_{N-1}(\mathrm{x}_1) - J_{N-1}(\mathrm{x}_2)$? *[2 points]*

(A) ☐ 0

(B) ☐ $\mathrm{x}_1 - \mathrm{x}_2$

(C) ☐ $2(\mathrm{x}_1 - \mathrm{x}_2)$

(D) ☐ $2(\mathrm{x}_1 + \mathrm{x}_2)$

(E) ☐ None of the above

c) Consider the system dynamics

$$x_{k+1} = x_k u_k + x_{k-1} x_{k-2}, \quad k = 2, 3, ..., 7,$$

with $x_0 = 0$, $x_1 = 1$, and $x_2 = 1$. At each time step, the set of admissible control inputs is $\mathcal{U} = \{-1, 1\}$.

i) Which of the following augmented system dynamics models the problem so that you can apply the Dynamic Programming Algorithm, for $k = 2, ..., 7$? *[1 point]*

(A) ☐ $\tilde{x}_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix}$, $\tilde{x}_{k+1} = \begin{bmatrix} x_k u_k + y_k y_{k-1} \\ x_k \end{bmatrix}$

(B) ☐ $\tilde{x}_k = \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix}$, $\tilde{x}_{k+1} = \begin{bmatrix} x_k u_k + y_k z_k \\ x_k \\ y_k \end{bmatrix}$

(C) ☐ $\tilde{x}_k = \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix}$, $\tilde{x}_{k+1} = \begin{bmatrix} x_k u_k + y_k z_k \\ z_k \\ y_k \end{bmatrix}$

(D) ☐ $\tilde{x}_k = \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix}$, $\tilde{x}_{k+1} = \begin{bmatrix} x_k u_k + y_k z_k \\ y_k z_k \\ y_k \end{bmatrix}$

(E) ☐ $\tilde{x}_k = \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix}$, $\tilde{x}_{k+1} = \begin{bmatrix} x_k y_k + y_k z_k \\ x_k \\ y_k \end{bmatrix}$

ii) How many possible values can x_4 take? *[1 point]*

iii) Let $\tilde{x}_k$ be the correct augmented state formulation from question i). How many possible values can $\tilde{x}_4$ take? *[1 point]*

iv) Let $\tilde{x}_k$ be the correct augmented state formulation from question i). The stage cost function $g_3(\tilde{x}_k, u_k)$ of the augmented problem is defined as

$$g_3(\tilde{x}_3, u_3) = c \quad \forall \tilde{x}_3 \in \tilde{\mathcal{S}}_3,\ \forall u_3 \in \mathcal{U},$$

where the constant c is $-1 \leq c \leq 5$. Furthermore, we know that

$$J_4\left(\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}\right) = -4, \qquad J_4\left(\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}\right) = 3,$$

where the function $J_k(\tilde{x}_k)$ represents the optimal cost to go at stage k. Which values could $J_3\left(\begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}\right)$ take? Select all possible correct answers. *[3 points]*

-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

v) Using the augmented state formulation, is it possible to solve this problem with the forward Dynamic Programming Algorithm? Justify your answer. *[1 point]*

Solution 1

a) i) The set of reachable states at time step 3 is $\mathcal{S}_3 = \{-3, -2, \ldots, 2, 3\}$, so $|\mathcal{S}_3| = 7$.

ii) The number of open loop control policies is

$$N_u^N = 3^4.$$

iii) Here we cannot use the formula $N_u^{N_x(N-1)+1}$ to determine the number of closed loop control policies, since the state space is not the same at each time step. The number of closed loop control policies is the product of the number of possible policies at each time step. The number of possible closed loop control policies at time step k is itself $N_u^{|\mathcal{S}_k|}$. It is very easy to note that $|\mathcal{S}_k| = 2k + 1$. Thus, in general the number of closed loop control policies is equal to

$$N_u^{\sum_{i=0}^{N-1}(2i+1)} = 3^{\sum_{i=0}^{N-1}(2i+1)}.$$

In this specific case the number of closed loop control policies would be equal to 3^{16}.

iv) The general answer is

$$3^{\sum_{i=0}^{N-1}(2i+1)},$$

as explained in the previous point.

v) In the Dynamic Programming Algorithm, we perform a minimization in u for each reachable state at each time step. Thus, the total number of minimizations is equal to

$$\sum_{i=0}^{N-1} |\mathcal{S}_k| = \sum_{i=0}^{N-1} (2i + 1).$$

b) i) The expected value of the continuous random variable w can easily be computed as

$$\mathbb{E}[w] = \int_{-\infty}^{+\infty} w p_w dw = \int_{-\alpha}^{\alpha} \left[\frac{w}{2\alpha} + \frac{w^2}{2\alpha^2} \right] dw = \frac{\alpha}{3}$$

ii) The Dynamic Programming Algorithm initialization and recursion implies that

$$J_N(x_N) = x_N \quad \forall x_N \in \mathcal{S}_N,$$

$$\begin{aligned} J_{N-1}(x_{N-1}) &= \min_u x_{N-1} + u^2 + x_{N-1} + u\mathbb{E}[w] + u x_{N-1} \\ &= \min_u u^2 + u(x_{N-1} + \frac{\alpha}{3}) + 2x_{N-1} \end{aligned} \tag{1}$$

Equation (1) is a parabola in u, thus has its argmin at $u = -\frac{(x_{N-1} + \frac{\alpha}{3})}{2}$ (this can also be seen by differentiating in u). When $\alpha = 6$ and $x_{N-1} = 1$, the minimum of equation (1) is achieved at $u = -\frac{3}{2} \in \mathcal{U}$, with corresponding optimal cost $J_{N-1}(1) = -\frac{1}{4}$.

iii) When $x_{N-1} = 0$, the coordinate in u of the vertex of the parabola in (1) would be attained at $u = -\frac{\alpha}{6}$. Since $\alpha > 0$, such vertex coordinate is always negative. This means that if $-\frac{\alpha}{6} \in \mathcal{U} = [-2, 2]$, then the minimum is attained for $u = -\frac{\alpha}{6}$. Otherwise it must be $-\frac{\alpha}{6} < -2$. In this case, the convex parabola is increasing in the interval $[-2, 2]$ and always attains its minimum at -2. Thus, the minimum is -2 when $-\frac{\alpha}{6} \leq -2$, i.e. $\alpha \geq 12$.

iv) There is no value of alpha for which this happens, as explained in the previous point. Either the best u is $-\frac{\alpha}{6} < 0$ (if this value is in the admissible control set), or the best control input is -2.

v) As $\alpha \to 0$ the maximum optimal control input becomes as large as possible, i.e. $-\frac{\alpha}{6} \to 0$, but never touches that value. Thus 0 is the superior limit, but it is never achieved for any value of $\alpha > 0$.

vi) No, it cannot be solved with forward Dynamic Programming Algorithm because of the presence of the disturbances.

vii) The key here is noticing that the optimal u does not depend on the state x. Indeed, the optimal u at stage $N-1$ is equal to $u^* = -\frac{\mathbb{E}(w)}{2}$, for each admissible state. It can be easily seen that, for each x_{N-1}, it holds

$$J_{N-1}(x_{N-1}) = 2x_{N-1} + (u^*)^2 + u^*\mathbb{E}(w) = 2x_{N-1} + k,$$

where k is the same constant for each x_{N-1}. This implies that $J_{N-1}(\mathrm{x}_1) - J_{N-1}(\mathrm{x}_2) = 2(\mathrm{x}_1 - \mathrm{x}_2)$, since the constants cancel out.

c) i) Formulation (B) is the only correct standard reformulation.

ii) x_4 can take two possible values i.e. 0 or 2.

iii) The state $\tilde{x}_4$ is a triplet and can take four possible values, namely $\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$, $\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$.

iv) From the state $\tilde{x}_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ we can reach two possible states at the next time step, i.e. state $\begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$ if we use control input 1, and $\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix}$, if we use control input -1. In this problem there are no disturbances. Given that the stage cost $g_3 = c$ is constant with respect to u, the Dynamic Programming Algorithm recursion tells us that the best u is the one that leads to the minimum J_4. Thus for $\tilde{x}_3 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$ the optimal

control input is 1, since $J_4\left(\begin{bmatrix}2\\-1\\1\end{bmatrix}\right) = -4 < 3 = J_4\left(\begin{bmatrix}0\\-1\\1\end{bmatrix}\right)$, and the optimal associated cost is $c - 4$. Since $-1 \leq c \leq 5$, the optimal cost must be between $-1 - 4 = -5$ and $5 - 4 = 1$, inclusive. So the admissible values are $-5, -4, \ldots, 1$.

v) Yes, the problem can be solved through standard Dynamic Programming Algorithm because there are no disturbances and the state space is finite.

Problem 2 [25 points]

For the multiple-choice questions in this problem, you will get the points if only the correct answers are marked.

a) You allocated 30 hours in the semester break to prepare for the Dynamic Programming and Optimal Control exam. The content of this course consists of 4 sections, *A*, *B*, *C*, and *D*. Each section corresponds to one problem in the exam. You have estimated how many points each problem is worth, and the time it takes to prepare for each section, as shown in Table 1.

Since preparing all sections would take longer than your time budget for the course, you would like to find out which sections are worth preparing to maximize your final grade. You have made the following assumptions:

- You get full points for the sections you have prepared and half of the points for the sections you have not prepared;
- You do not start preparing a section if you would not have enough time to finish it within the 30h time budget;
- You have to study section *A* before *B*, *C*, or *D*. You have to study section *B* before *D*;
- Before starting the preparation, you may spend 4 hours to do a math review *M*, which would help you understand the course faster. The time to prepare for each section after completing the math review is shown in Table 1.

Section	Points	Time needed to prepare (h)	Time needed to prepare after math review (h)
A	20	8	6
B	30	12	10
C	20	6	6
D	30	14	10

Table 1: Points of each section and time to prepare.

Answer the following questions based on the provided information.

i) This task can be formulated as a deterministic finite state problem. We define the tuple *(list of reviewed contents, time remaining)* as the state. For example, *(MAB, 10)* means you have performed the math review and studied for sections A and B, and thus you have 10 hours left. When you finish reviewing a section, you get the corresponding points as a reward; the points for the sections you do not review are received in the last step.
Complete the state transition diagram in Figure 1 by filling in the states in each node and writing the **rewards** next to each edge. *[2 points]*

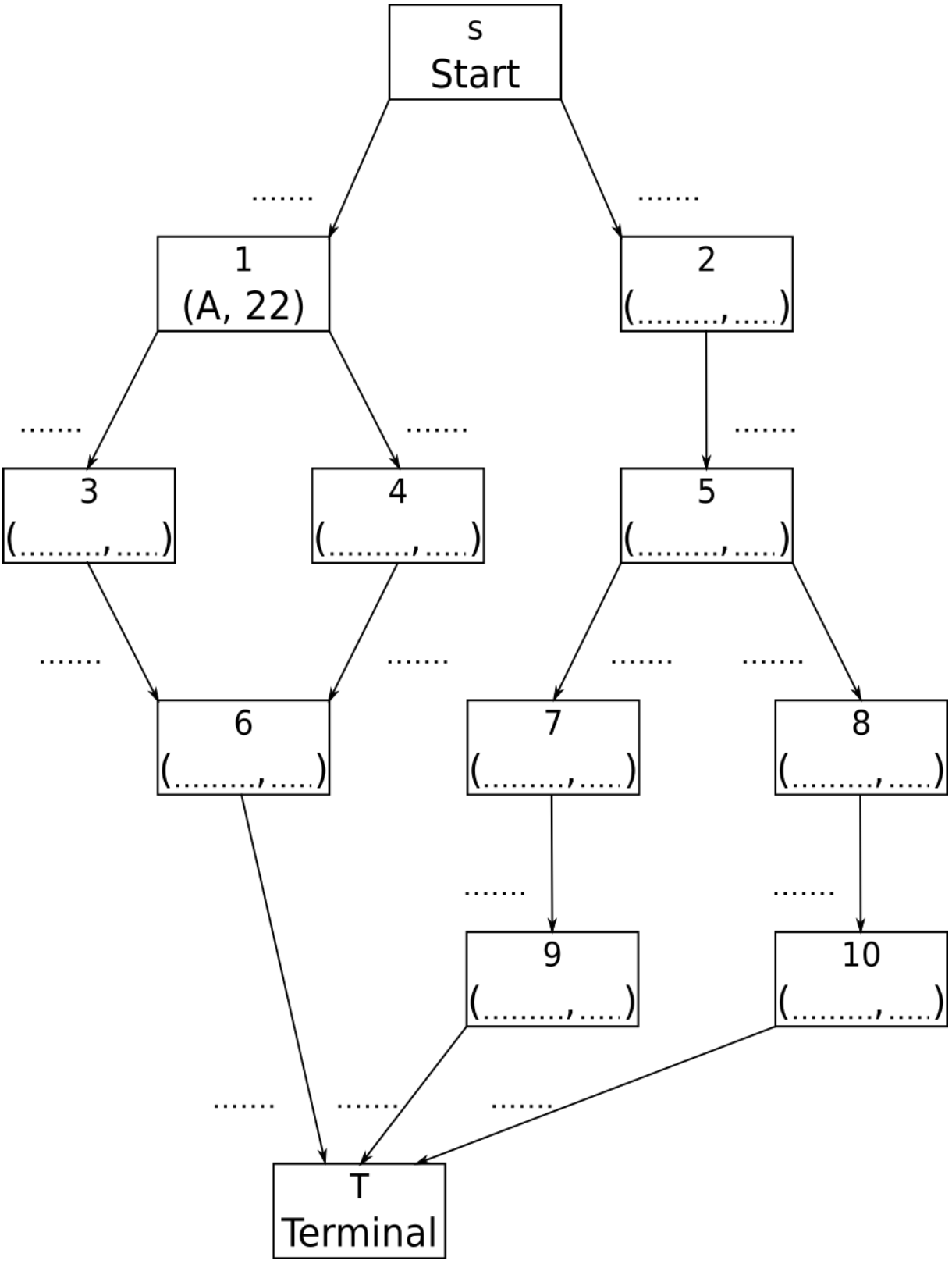

Figure 1: State transition diagram to be filled.

ii) Finding the path from S to T with the highest reward in Figure 1 is very similar to the standard Shortest Path problem and is sometimes called the Longest Path problem. Which of the following statements about this problem are correct? *[2 point]*

(A) ☐ This Longest Path problem is well-defined because there are no negative cycles in Figure 1.

(B) ☐ This Longest Path problem is well-defined because there are no positive cycles in Figure 1.

(C) ☐ In Figure 1, there is no edge from node 2 to node 4, so the reward from node 2 to node 4 is ∞.

(D) ☐ In Figure 1, there is no edge from node 4 to node 7, so the reward from node 4 to node 7 is $-\infty$.

iii) When solving this problem with the **standard** Dynamic Programming Algorithm, which of the following statements are correct? *[2 point]*
In the following statements, $r_{i,j}$ represents the reward associated with the edge from node i to node j. N is the time horizon, and $J_k(i)$ is the optimal total reward obtained when getting from node i to node T in $N-k$ steps.

(A) ☐ $N = 10$.

(B) ☐ The Dynamic Programming Algorithm should be initialized with $J_k(\mathsf{T}) = 0$ for the terminal node and $J_k(i) = \infty$ for any other node i.

(C) ☐ At each iteration, the update rule is $J_k(i) = \min\limits_{j \neq \mathsf{T}} (r_{i,j} + J_{k+1}(j))$, for all node i except the terminal node.

(D) ☐ At each iteration, the update rule is $J_k(i) = \max\limits_{j \neq \mathsf{T}} (r_{i,j} + J_{k+1}(j))$, for all node i except the terminal node.

iv) Perform the **forward** Dynamic Programming Algorithm and fill in this table. Note that $J_k^F(i)$ is the optimal total reward obtained when going from node `s` to node i in k steps. *[2 points]*

Node Index i	$J_1^F(i)$	$J_2^F(i)$	$J_3^F(i)$	$J_4^F(i)$	$J_5^F(i)$
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
`T`					

v) Would you get the same result when solving this problem with the **forward** and the **backward** Dynamic Programming Algorithm? Justify your answer. *[2 points]*

b) You come to the campus to study for the exam every day. Thus it is critical to find the shortest path from your home to the university. The following graph represents a map of Zurich showing the traveling time between your home (node `s`), the university (node `t`), and major bus stations (nodes 1 to 6). The numbers on the edges represent the time to travel between the nodes.
In this problem, for the questions dealing with the Label Correcting Algorithm, if two nodes are added into the OPEN bin at the same iteration, the node with the smaller index is added first.

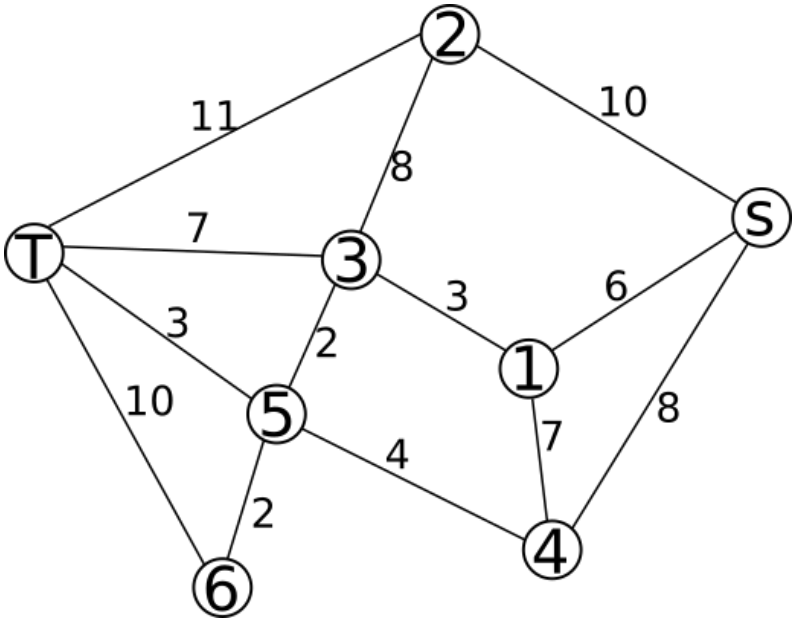

Figure 2: Map of Zurich transportation.

Iteration	Remove	OPEN	d_s	d_1	d_2	d_3	d_4	d_5	d_6	d_t
0	-	s	0	∞	∞	∞	∞	∞	∞	∞
1	s	1,2,4	0	6	10	∞	8	∞	∞	∞
2	1	2,4,3	0	6	10	9	8	∞	∞	∞
3	4	2,3,5	0	6	10	9	8	12	∞	∞

Table 2: Label Correcting Algorithm Table for questions i) and ii).

i) Three iterations of the Label Correcting Algorithm have been performed in Table 2. Which search method is used to produce these results? *[1 point]*

(A) ☐ Breadth-first Search

(B) ☐ Depth-first Search

(C) ☐ Best-first Search

ii) Complete Table 2 using the Label Correcting Algorithm with the same method selected in the previous question. *[2 points]*

iii) Write down the optimal path based on the completion of Table 2. *[1 point]*

iv) Recall that in the A^* algorithm you consider $d_j + h_j$ instead of d_j when adding j into the OPEN bin.
Justify why the length of the shortest edge connecting j to another node, i.e. $\min_{k \in V \setminus \{j\}} c_{jk}$, is a good choice for the A^* heuristic h_j. *[1 point]*

v) Complete the A^* algorithm Table below using the heuristic $h_j = \min_{k \in V \setminus \{j\}} c_{jk}$.
At each iteration, remove the node j with the smallest $d_j + h_j$ from the OPEN bin. *[3 points]*

Iteration	Remove	OPEN	d_s	d_1	d_2	d_3	d_4	d_5	d_6	d_t
0	-	s	0	∞	∞	∞	∞	∞	∞	∞
1	s	1,2,4	0	6	10	∞	8	∞	∞	∞

Table 3: A^* Table for question v).

c) Select the correct answer(s) for the following questions.

i) When applying the standard Dynamic Programming Algorithm on a Shortest Path problem, suppose that $J_k(i) = 10$ was obtained for a node i. Then it must hold that $J_{k-1}(i) \leq 10$. *[1 point]*

☐ True
☐ False

ii) In a Shortest Path problem, if a positive number $l \in \mathbb{R}$ is added to the length of any edge that does not belong to the optimal path, then Dijkstra's Algorithm might take fewer iterations to converge. *[1 point]*

☐ True
☐ False

iii) In the A^* algorithm, suppose that we select node j with the smallest $d_j + h_j$ to be removed from the OPEN bin at each iteration.
If h_j is equal to the length of the shortest path from node j to node T, then d_{T} will only be updated once. *[1 point]*

☐ True
☐ False

iv) The Shortest Path problem can be solved by either the standard Dynamic Programming Algorithm or the Label Correcting Algorithm. Which statements about the two methods are true? *[1 point]*

(A) ☐ The standard Dynamic Programming Algorithm terminates with the length of the shortest path from each node to the terminal node.

(B) ☐ The Label Correcting Algorithm terminates with the length of the shortest path from the start node to each node.

(C) ☐ Both (A) and (B) are true.

(D) ☐ Neither (A) or (B) is true.

v) The state of a Hidden Markov Model is estimated using the Viterbi Algorithm. The observation sequence is $Z_1 = (z_1, z_2, \ldots, z_N)$ and the most likely state trajectory found is $X^* = (x_1^*, x_2^*, \ldots, x_N^*)$. The length of the shortest path found in the formulated shortest path problem is 0.6. What is the value of the joint probability of the state trajectory and observation, $p(X^*, Z_0)$? *[1 points]*

(A) ☐ 0.6

(B) ☐ $e^{-0.6}$

(C) ☐ $ln(\frac{5}{3})$

(D) ☐ None of the above, but can be determined

(D) ☐ Cannot be determined

vi) In a Hidden Markov Model with time horizon N, the state space is $\mathcal{S}$ and the measurement space is $\mathcal{Z}$ at every time step. Both the state transition and the measurement are uniformly distributed, i.e.

$$P_{ij} = 1/|\mathcal{S}|, \ \forall i \in \mathcal{S}, \ \forall j \in \mathcal{S},$$

$$M_{ij}(z) = 1/|\mathcal{Z}|, \ \forall z \in \mathcal{Z}, \ \forall i \in \mathcal{S}, \ \forall j \in \mathcal{S},$$

where $|\cdot|$ denotes the number of elements in a set.
The Viterbi Algorithm is used to estimate the state trajectory of this model. Which of the following statements are correct? *[2 points]*

(A) ☐ The resulting Shortest Path problem has $N|\mathcal{S}|$ shortest paths of equal length.

(B) ☐ Any path from s to t in the formulated Shortest Path problem is optimal.

(C) ☐ The formulated Shortest Path problem has $N|\mathcal{S}|$ nodes in total.

(D) ☐ The formulated Shortest Path problem has $N|\mathcal{S}|^2+2|\mathcal{S}|$ edges in total.

Solution 2

a) i) **Node 3 and 4 interchangeable!! Nodes {7,9} and {8,10} interchangeable!!**

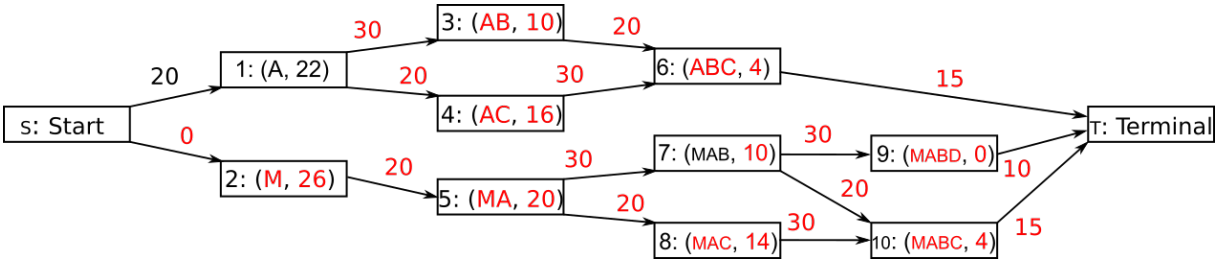

Figure 3: State transition diagram filled

ii) (B) and (D)

iii) (D)

iv) **Node 3 and 4 interchangeable!!**

Node Index i	$J_1^F(i)$	$J_2^F(i)$	$J_3^F(i)$	$J_4^F(i)$	$J_5^F(i)$
1	20	20	20	20	20
2	0	0	0	0	0
3	-∞	50	50	50	50
4	-∞	40	40	40	40
5	-∞	20	20	20	20
6	-∞	-∞	70	70	70
7	-∞	-∞	50	50	50
8	-∞	-∞	40	40	40
9	-∞	-∞	-∞	80	80
10	-∞	-∞	-∞	70	70
T	-∞	-∞	-∞	85	90

v) Yes. They are the same due to the fact that SP problems are symmetric. Forward DPA is the same as a the standard backward DPA applied to the problem with the arcs flipped maintaining the same arc length.

b) i) (C)

ii) The completed LCA is shown in Table 4.

iii) The shortest path is: $S \to 1 \to 3 \to 5 \to T$.

iv) The heuristic chosen is appropriate because h_j is positive and shorter than or equal to the shortest path from node j to T.

v) A^* iterations are shown in Table 5.

c) i) True

ii) True

iii) True

iv) (A)

Iteration	Remove	OPEN	d_s	d_1	d_2	d_3	d_4	d_5	d_6	d_t
0	-	s	0	∞	∞	∞	∞	∞	∞	∞
1	s	1,2,4	0	6	10	∞	8	∞	∞	∞
2	1	2,4,3	0	6	10	9	8	∞	∞	∞
3	4	2,3,5	0	6	10	9	8	12	∞	∞
4	3	2,5	0	6	10	9	8	11	∞	16
5	2	5	0	6	10	9	8	11	∞	16
6	5	6	0	6	10	9	8	11	13	14
7	6	-	0	6	10	9	8	11	13	14

Table 4: Completed table using LCA

Iteration	Remove	OPEN	d_s	d_1	d_2	d_3	d_4	d_5	d_6	d_t
0	-	s	0	∞	∞	∞	∞	∞	∞	∞
1	s	1,2,4	0	6	10	∞	8	∞	∞	∞
2	1	2,4,3	0	6	10	9	8	∞	∞	∞
3	3	2,4,5	0	6	10	9	8	11	∞	16
4	4	2,5	0	6	10	9	8	11	∞	16
5	5	2	0	6	10	9	8	11	13	14
6	2	-	0	6	10	9	8	11	13	14

*Table 5: Completed table using A**

v) (B)

vi) (B) and (D)

Problem 3 [20 points]

In this problem, every question is worth 1 point. For the multiple-choice and True/False questions, you get 0 points for each blank answer and -0.5 points for each incorrect or ambiguous answer. The minimum total number of points of the problem is 0.

a) Answer the following questions for the standard Stochastic Shortest Path problem. We use VI and PI in order to abbreviate Value Iteration and Policy Iteration, respectively.

i) In Stochastic Shortest Path problems, the VI algorithm involves solving a system of linear equations.

☐ True
☐ False

ii) Select the correct statement.

(A) ☐ PI and VI always converge to the same optimal cost and policy.

(B) ☐ PI and VI always converge to the same optimal cost but do not always converge to the same optimal policy.

(C) ☐ PI and VI do not always converge to the same optimal cost and policy.

(D) ☐ PI and VI do not always converge to the same optimal cost but always converge to the same optimal policy.

iii) In order to have a unique solution in the policy evaluation phase of PI, it is necessary to remove the terminal state from the state space.

☐ True
☐ False

iv) The optimal policy π^* is time-invariant.

☐ True
☐ False

v) VI can be initialized arbitrarily.

☐ True
☐ False

vi) PI can be initialized arbitrarily.

☐ True
☐ False

vii) It is necessary to update the policy for every state at every iteration for PI to converge to an optimal policy.

- ☐ True
- ☐ False

viii) VI is always less computationally efficient than PI, when used to solve the same problem.

- ☐ True
- ☐ False

ix) In order to solve the Stochastic Shortest Path problem there must be a state, denoted as 0, such that $P_{00}(u) = 0$ and $g(0, \mathrm{u}, 0) = 1, \forall \mathrm{u} \in \mathcal{U}(0)$.

- ☐ True
- ☐ False

b) Answer the following questions for the discounted Stochastic Shortest Path problem with discount factor α, with $0 < \alpha < 1$. We use VI and PI in order to abbreviate Value Iteration and Policy Iteration, respectively.

i) In discounted Stochastic Shortest Path problems, the expected closed loop cost incurred by a stationary policy always exists, i.e. it is finite for every initial state.

☐ True
☐ False

ii) In discounted Stochastic Shortest Path problems, the PI algorithm cannot be initialized with an arbitrary admissible policy.

☐ True
☐ False

iii) Consider the auxiliary problem used to find the solution of a discounted Stochastic Shortest Path problem, as defined in the lecture. There is a one-to-one mapping between a policy π of the auxiliary problem and a policy $\tilde{\pi}$ of the discounted problem.

☐ True
☐ False

iv) Discounted Stochastic Shortest Path problems cannot have a termination state.

☐ True
☐ False

v) Explain in your own words why the discounted Stochastic Shortest Path problem does not require a terminal state to be solved.

c) Answer the following questions considering the Stochastic Shortest Path problem represented in Figure 4, where at any state $i \in \{0, 1, 2\}$ the control action u can either be A or B.

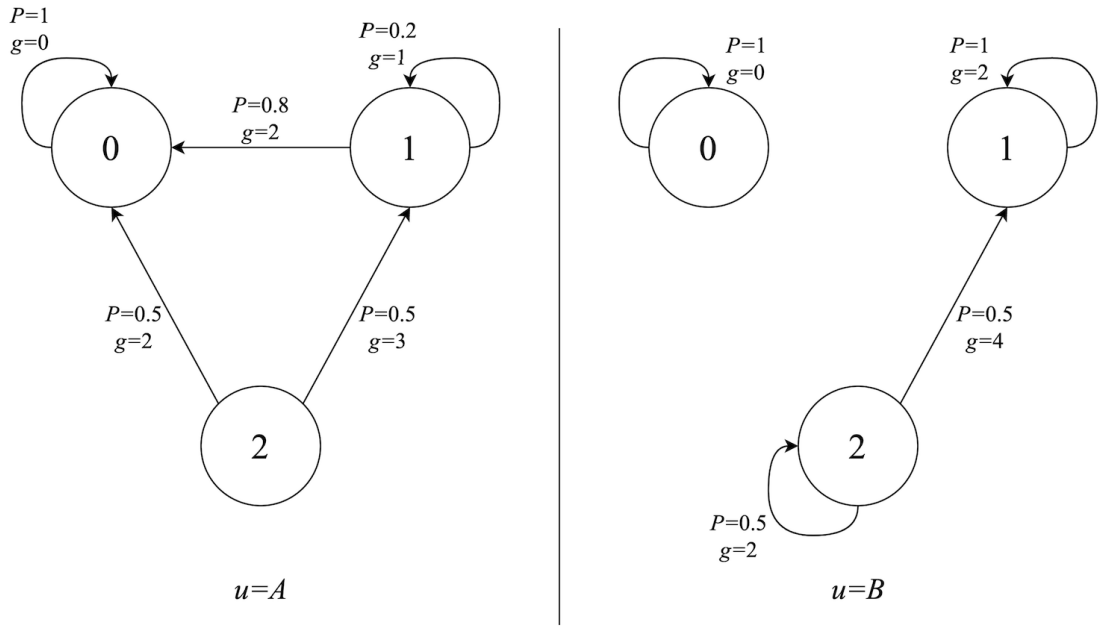

Figure 4: State transition graph with associated probabilities and costs denoted on each edge.

The associated probabilities and costs are:

$$
\begin{aligned}
P_{00}(A) &= 1 \\
P_{10}(A) &= 0.8 \\
P_{11}(A) &= 0.2 \\
P_{20}(A) &= 0.5 \\
P_{21}(A) &= 0.5 \\
g(0, A, 0) &= 0 \\
g(0, A, 1) &= 2 \\
g(1, A, 1) &= 1 \\
g(2, A, 0) &= 2 \\
g(2, A, 1) &= 3
\end{aligned}
\qquad
\begin{aligned}
P_{00}(B) &= 1 \\
P_{11}(B) &= 1 \\
P_{21}(B) &= 0.5 \\
P_{22}(B) &= 0.5 \\
g(0, B, 0) &= 0 \\
g(1, B, 1) &= 2 \\
g(2, B, 1) &= 4 \\
g(2, B, 2) &= 2
\end{aligned}
$$

i) When solving this problem using the Policy Iteration algorithm, which of the following policies can be used to initialize the algorithm?

(A) ☐ $\mu(0) = A, \quad \mu(1) = A, \quad \mu(2) = A$

(B) ☐ $\mu(0) = A, \quad \mu(1) = B, \quad \mu(2) = B$

(C) ☐ $\mu(0) = B, \quad \mu(1) = A, \quad \mu(2) = B$

(D) ☐ $\mu(0) = B, \quad \mu(1) = A, \quad \mu(2) = A$

(E) ☐ $\mu(0) = A, \quad \mu(1) = B, \quad \mu(2) = A$

(F) ☐ None of the above

ii) Which of the following statements about the Stochastic Shortest Path problem represented in Figure 4 is correct?

(A) ☐ 8 proper policies, 0 improper policies.

(B) ☐ 6 proper policies, 2 improper policies.

(C) ☐ 4 proper policies, 4 improper policies.

(D) ☐ 2 proper policies, 6 improper policies.

(E) ☐ None of the above.

iii) Let $\mu(1) = A$ and $\mu(2) = B$. Select the correct transition probability matrix $P_\mu \in \mathbb{R}^{2\times 2}$, whose $(i,j)^{th}$ entry is $P_{ij}(\mu(i))$ with $i, j \in \{1, 2\}$.

(A) ☐ $P_\mu = \begin{bmatrix} 0.2 & 0.8 \\ 0.5 & 0.5 \end{bmatrix}$

(B) ☐ $P_\mu = \begin{bmatrix} 0.2 & 0 \\ 0.5 & 0.5 \end{bmatrix}$

(C) ☐ $P_\mu = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.5 \end{bmatrix}$

(D) ☐ $P_\mu = \begin{bmatrix} 0.2 & 0 \\ 0.5 & 0 \end{bmatrix}$

(E) ☐ None of the above

d) Answer the following questions considering the discounted Stochastic Shortest Path problem represented in Figure 5, where at any state $i \in \{1, 2\}$ the control action u can either be C or D.

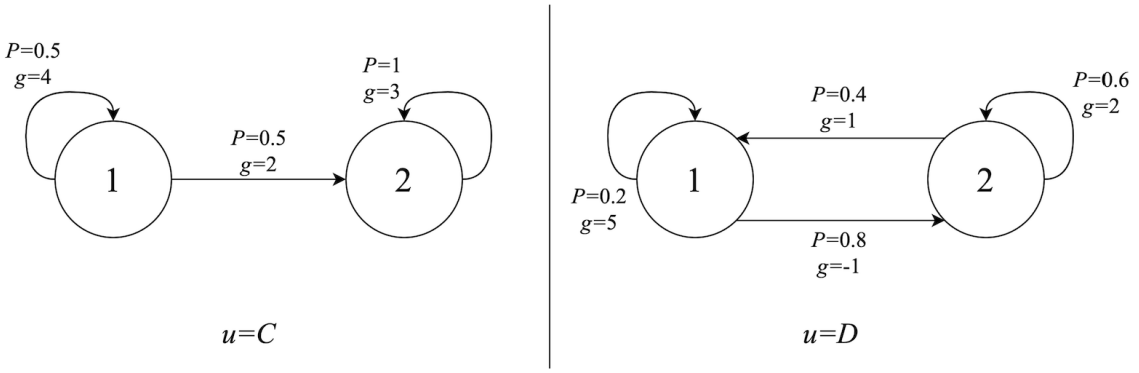

Figure 5: State transition graph with associated probabilities and costs denoted on each edge.

The associated probabilities and costs are:

$$\begin{aligned} P_{11}(C) &= 0.5 \\ P_{12}(C) &= 0.5 \\ P_{22}(C) &= 1 \\ g(1, C, 1) &= 4 \\ g(1, C, 2) &= 2 \\ g(2, C, 2) &= 3 \end{aligned} \qquad \begin{aligned} P_{11}(D) &= 0.2 \\ P_{12}(D) &= 0.8 \\ P_{21}(D) &= 0.4 \\ P_{22}(D) &= 0.6 \\ g(1, D, 1) &= 5 \\ g(1, D, 2) &= -1 \\ g(2, D, 1) &= 1 \\ g(2, D, 2) &= 2 \end{aligned}$$

i) Let $q(i, \mathrm{u}) := \underset{(w|x=i,u=u)}{E}[g(x, u, w)]$. Which of the following equations are correct for the problem represented in Figure 5?

(A) ☐ $q(1, D) = 0.2$

(B) ☐ $q(1, D) = 1.8$

(C) ☐ $q(2, C) = 3$

(D) ☐ $q(2, D) = 2$

ii) What is the Bellman Equation for the discounted Stochastic Shortest Path problem represented in Figure 5, with discount factor $\alpha = 0.5$?

(A) ☐ $\begin{cases} J(1) = \min\{3 + 0.5J(1) + 0.5J(2), 1.8 + 0.2J(1) + 0.8J(2)\} \\ J(2) = \min\{3 + J(2), 2 + 0.4J(1) + 0.6J(2)\} \end{cases}$

(B) ☐ $\begin{cases} J(1) = \min\{3 + 0.5J(1) + 0.5J(2), 0.2 + 0.2J(1) + 0.8J(2)\} \\ J(2) = \min\{3 + J(2), 1.6 + 0.4J(1) + 0.6J(2)\} \end{cases}$

(C) ☐ $\begin{cases} J(1) = \min\{3 + 0.25J(1) + 0.25J(2), 0.2 + 0.1J(1) + 0.4J(2)\} \\ J(2) = \min\{3 + 0.5J(2), 1.6 + 0.2J(1) + 0.3J(2)\} \end{cases}$

(D) ☐ $\begin{cases} J(1) = \min\{3 + 0.25J(1) + 0.25J(2), 1.8 + 0.1J(1) + 0.4J(2)\} \\ J(2) = \min\{3 + 0.5J(2), 2 + 0.2J(1) + 0.3J(2)\} \end{cases}$

(E) ☐ None of the above

iii) Let $\mu(1) = C$ and $\mu(2) = C$. What is the associated expected infinite horizon closed loop cost for the discounted Stochastic Shortest Path problem represented in Figure 5, with discount factor $\alpha = 0.5$?

(A) ☐ $\begin{cases} J(1) = 7.5 \\ J(2) = 1.5 \end{cases}$

(B) ☐ $\begin{cases} J(1) = 6 \\ J(2) = 6 \end{cases}$

(C) ☐ $\begin{cases} J(1) = 4.5 \\ J(2) = 6 \end{cases}$

(D) ☐ Cannot be determined

(E) ☐ None of the above, but can be determined with the given information

Solution 3

a) i) **False**, policy iteration involves solving a system of linear equations.

ii) **(B)**, PI and VI always converge to the same optimal cost (that is unique) but there could be multiple possible optimal policies.

iii) **True**. There is a unique solution for the policy evaluation stage of the PI if and only if $(I - P)$ is invertible, that means that we need to remove the terminal state from the state space. If we do not remove the terminal state $(I - P)$ is not invertible, since in P the row of the terminal state has all zeros except for a 1 on the diagonal.

iv) **True**, the solution of the Stochastic Shortest Path problem is time-invariant.

v) **True**, VI works with any initialization.

vi) **False**, we can only use proper policies to initialize PI.

vii) **False**, see asynchronous PI.

viii) **False**, depends on the given problem.

ix) **False**, see assumption on cost-free termination state.

b) i) **True**, it is one of the properties of the discounted Stochastic Shortest Path problems.

ii) **False**, it can be initialized with any admissible policy.

iii) **True**, it is one of the properties between the discounted Stochastic Shortest Path problem and its associated auxiliary problem (see lecture on discounted Stochastic Shortest Path problems).

iv) **False**, the addition of a terminal state would not change the feasibility of the discounted Stochastic Shortest Path problem.

v) The discounted Stochastic Shortest Path problem can be converted to an equivalent standard Stochastic Shortest Path (as seen in lectures) problem, since in the discounted Stochastic Shortest Path problem the costs $\to 0$ with $N \to \infty$. For this reason there is no need for a terminal state and the problem is still solvable using Stochastic Shortest Path techniques.

c) i) **(A), (C) and (D)**, PI needs to be initialized by only proper policies:

- $\mu(0) = A, \quad \mu(1) = A, \quad \mu(2) = A$ **proper**,
- $\mu(0) = B, \quad \mu(1) = A, \quad \mu(2) = A$ **proper**,
- $\mu(0) = A, \quad \mu(1) = B, \quad \mu(2) = A$ **improper**,
- $\mu(0) = B, \quad \mu(1) = B, \quad \mu(2) = A$ **improper**,
- $\mu(0) = A, \quad \mu(1) = A, \quad \mu(2) = B$ **proper**,
- $\mu(0) = B, \quad \mu(1) = A, \quad \mu(2) = B$ **proper**,
- $\mu(0) = A, \quad \mu(1) = B, \quad \mu(2) = B$ **improper**,
- $\mu(0) = B, \quad \mu(1) = B, \quad \mu(2) = B$ **improper**.

ii) **(C)**, see previous answer.

iii) **(B)**, just substitute the values in the formula given by the text.

d) i) **(A) and (C)**:

- $q(1, D) = -1 \cdot 0.8 + 5 \cdot 0.2 = 0.2$,
- $q(2, C) = 3 \cdot 1 = 3$,
- $q(2, D) = 1 \cdot 0.4 + 2 \cdot 0.6 = 1.6$.

ii) **(C)**, the Bellman Equation formula is

$$J(i) = \min_{\mathrm{u} \in \mathcal{U}(i)} \left[q(i, \mathrm{u}) + \alpha \cdot \sum_{j=1}^{n} P_{ij}(\mathrm{u}) J(j) \right] \quad , \quad \forall i \in S^{+} \quad .$$

iii) **(B)**, the expected infinite horizon closed loop cost formula is:

$$J(i) = q(i, \mu(i)) + \alpha \cdot \sum_{j=1}^{n} P_{ij}(\mu(i)) J(j) \quad , \quad \forall i \in S^{+} \quad .$$

Problem 4 [25 points]

Consider the system dynamics

$$\dot{x}(t) = u(t),$$

where $x(t)$ is the state of the system and $u(t)$ the control input at time t, with $|u(t)| \leq \beta$, $\beta \geq 0$. The goal of this problem is to derive a control policy to track a given reference trajectory $h(t)$ using the Pontryagin Minimum Principle, while enforcing the system's initial and terminal conditions. The system starts at $x(0) = 1$ and ends at $x(T) = 2$, and the cost function of the tracking problem is given as

$$\int_0^T \frac{1}{2}(x(t) - h(t))^2 dt.$$

For the remainder of the problem, we set $T = 1$ and the reference trajectory as $h(t) = t$.

a) Write the Hamiltonian $H(x, u, p, t)$ and the co-state derivative $\dot{p}(t)$ of the problem. *[2 points]*

b) For what values of β does the problem have no solution? **Justify** your answer. *[2 points]*

c) What is the optimal control input when $\beta = 1$? **Justify** your answer. *[1 point]*

d) For the values of β with $1 \leq \beta < \beta^*$, the resulting state trajectory $x(t)$ derived from the Minimum Principle never intersects $h(t)$. Show that $\beta^* = 1 + \sqrt{2}$. *[14 points]*
Hint: You have to solve the problem using the Minimum Principle and prove that the control trajectory must start with the control input $u(t) = -\beta$.

e) Derive the control input $u(t)$ for all $t \in [0, T]$ for $\beta > \beta^*$ that satisfies the Minimum Principle as a function of β. **Justify** your answer. *[6 points]*

Solution 4

a) The Hamiltonian is

$$H(x,u,p,t) = \frac{1}{2}(x(t)-t)^2 + p(t)u(t),$$

and the co-state derivative

$$\dot{p}(t) = -\frac{\partial H(x,u,p,t)}{\partial x} = t - x(t).$$

b) From the system dynamics, it is apparent that the terminal condition cannot be met for $0 \leq \beta < 1$, so the problem is not solvable for that interval.

c) For $\beta = 1$, the control input has to be $u(t) = 1$ to be able to reach the terminal condition, but we cannot take the cost function into consideration.

d) For any $\beta > 1$, we have

$$\begin{aligned} u(t) &= \arg\min_{|u|\leq\beta} H(x,u,p,t) \\ &= \arg\min_{|u|\leq\beta} \frac{1}{2}(x(t)-t)^2 + p(t)u(t), \end{aligned}$$

and thus

$$u(t) = \begin{cases} -\beta & \text{if } p(t) > 0 \\ \beta & \text{if } p(t) < 0 \\ u_s(t) & \text{if } p(t) = 0, \end{cases} \tag{2}$$

with $-\beta \leq u_s(t) \leq \beta$ a singular control input.

For $\beta > 1$, intuitively, the system will have the following behaviour. It will first apply some negative control input to reduce $x(t)$ and thus minimize the term $(x(t)-t)^2$. Then, it will apply some positive control to be able to reach the terminal condition. If β is big enough, the system will even be able to reach the position $x(t_1) = t_1$ and track the reference for a non-trivial amount of time, i.e $x(t) = t \; \forall \; t \in [t_1, t_2]$ with $t_1 < t_2 < T$.

We can formally prove this. From the co-state derivative, we can see that

$$\begin{cases} \dot{p}(t) < 0 & \text{for } t < x \\ \dot{p}(t) > 0 & \text{for } t > x \\ \dot{p}(t) = 0 & \text{for } t = x \end{cases} \tag{3}$$

Suppose the control trajectory starts with $u = \beta$. We see from Eq. 2 that $p(0) < 0$. Since $t < x$, $\dot{p}(t) < 0$ and $p(t)$ will remain negative for $t \in [0, T]$. Therefore, we have $u(t) = \beta$ for $t \in [0, T]$ and we will miss the terminal condition. We cannot start with $u = \beta$.

Suppose the control trajectory starts with $u = u_s(t)$, i.e. $p(0) = 0$. Since at the beginning, $t < x$, $\dot{p}(t) < 0$, $p(t)$ will immediately turn negative. Therefore, we cannot start with a singular input for a non-trivial amount of time.

We must start with $u = -\beta$, i.e. $p(0) > 0$. Since at the beginning, $t < x$, $\dot{p}(t) < 0$, and $p(t)$ decreases over time. However, $p(t)$ cannot stay positive for $t \in [0, T]$, otherwise it will miss the terminal condition, because $p(t) > 0$ implies $u(t) = -\beta$ for $t \in [0, T]$. At some point t_1, the system must switch the control input to either $u_s(t)$ or β. We can show that this depends on the value of β.

The control input $u_s(t)$ can only occur for a non-trivial amount of time when $p(t) = 0$ and $\dot{p}(t) = 0$, i.e. $x(t) = t$. However, for a too small value of β, $x(t) = t$ cannot be reached from the first regular arc and the terminal condition still met. For these values of β, we have $u(t) = -\beta$ for $t \in [0, t_1]$ and $u(t) = \beta$ for $t \in [t_1, T]$. Note that $u(t) = \beta$ for $t \in [t_1, T]$ implies $p(t) < 0$ and $\dot{p}(t) < 0$ because $t < x$ so that no more switch can occur until the end of the trajectory. We have $x(t) = 1 - \beta t$ for $t \in [0, t_1]$ and $x(t) = 1 - \beta t_1 + \beta(t - t_1)$ for $t \in (t_1, T]$. Using the terminal condition

$$\begin{aligned} & x(T) = 2 \\ \Leftrightarrow\ & 1 - \beta t_1 + \beta(1 - t_1) = 2 \\ \Leftrightarrow\ & t_1 = \frac{\beta - 1}{2\beta}. \end{aligned}$$

This occurs until β is large enough that the system can reach $x(t_1) = t_1$:

$$\begin{aligned} & x(t_1) = t_1 \\ \Leftrightarrow\ & 1 - \beta t_1 = t_1 \\ \Leftrightarrow\ & t_1 = \frac{1}{1 + \beta} \\ \Leftrightarrow\ & \frac{\beta - 1}{2\beta} = \frac{1}{1 + \beta} \\ \Leftrightarrow\ & \beta^2 - 2\beta - 1 = 0 \\ \Leftrightarrow\ & \beta = 1 \pm \sqrt{2}. \end{aligned}$$

As a result, we have $\beta^* = 1 + \sqrt{2}$ and for $1 < \beta \leq 1 + \sqrt{2}$, the resulting control strategy is

$$u(t) = \begin{cases} -\beta & \text{for } t \in \left[0, \dfrac{\beta - 1}{2\beta}\right] \\ \beta & \text{for } t \in \left[\dfrac{\beta - 1}{2\beta}, 1\right] \end{cases}$$

e) For any $\beta > 1 + \sqrt{2}$, this control strategy cannot hold anymore because otherwise the terminal condition cannot be met. A singular arc must occur for $t_1 < t \leq t_2 \leq T$. As mentioned above, along the singular arc, we have $x(t) = t$, and thus $u(t) = \dot{x}(t) = 1$. We therefore have

$$x(t) = \begin{cases} 1 - \beta t & \text{for } t \in [0, t_1] \\ t & \text{for } t \in (t_1, t_2] \\ 2 + \beta(t - 1) & \text{for } t \in (t_2, 1] \end{cases}$$

Using the boundary condition we have

$$\begin{aligned} & x(t_2) = t_2 \\ \Leftrightarrow\ & 2 + \beta(t_2 - 1) = t_2 \\ \Leftrightarrow\ & t_2 = \frac{\beta - 2}{\beta - 1} \end{aligned}$$

Recap: The control trajectory resulting from the Pontryagin minimum principle is the following:

- For $\beta < 1$, the problem has no solution,
- for $\beta = 1$: $u(t) = 1$ for $t \in [0, T]$,
- for $1 < \beta \leq 1 + \sqrt{2}$:

$$u(t) = \begin{cases} -\beta & \text{for } t \in \left[0, \dfrac{\beta - 1}{2\beta}\right] \\ \beta & \text{for } t \in \left[\dfrac{\beta - 1}{2\beta}, 1\right], \end{cases}$$

- for $\beta > 1 + \sqrt{2}$:

$$x(t) = \begin{cases} 1 - \beta t & \text{for } t \in [0, t_1] \\ t & \text{for } t \in (t_1, t_2] \\ 2 + \beta(t-1) & \text{for } t \in (t_2, 1], \end{cases}$$

with $t_1 = \dfrac{1}{1+\beta}$ and $t_2 = \dfrac{\beta - 2}{\beta - 1}$.