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 23rd, 2020**

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

Prof. R. D'Andrea

Solutions

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 *[6 points]*

$$x_{k+1} = f(x_k, u_{k-1}) = x_k + u_{k-1}, \qquad k = 0, 1,$$

and the augmented state definition and update equation

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

respectively, where $x_k \in \mathcal{X} = \{-3, -2, -1, 0, 1, 2, 3\}$ is the state, and $u_k \in \mathcal{U}_k(\tilde{x}_k)$ is the input with

$$\mathcal{U}_k(\tilde{x}_k) = \begin{cases} \{0\} & \text{for } k < 0 \\ \{u_k \text{ such that } u_k \neq 0 \text{ and } (f(x_k, y_k) + u_k) \in \mathcal{X}\} & \text{for } k \in \{0, 1\}. \end{cases}$$

The cost function is given by: $x_2 + \sum_{k=0}^{1} x_k^2 + u_k^2$.

i) Given $u_0 = 1$, mark **all** the elements that belong to the control space $\mathcal{U}_1(\tilde{x}_1)$ for the following x_1: *[3 points]*

		-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
(A)	$x_1 = -3$:	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
(B)	$x_1 = 0$:	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
(C)	$x_1 = 2$:	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

ii) Given $u_0 = 1$, mark **all** the optimal control inputs $\mu_1^*(\tilde{x}_1)$ for the following x_1: *[3 points]*

		-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
(A)	$x_1 = -3$:	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
(B)	$x_1 = 0$:	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
(C)	$x_1 = 2$:	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

b) Consider the system *[6 points]*

$$x_{k+1} = f(x_k, u_k) = x_k + u_k, \qquad k = 0, 1,$$

where $x_k \in \mathcal{X} = \{-3, -2, -1, 0, 1, 2, 3\}$ is the state, and $u_k \in \mathcal{U}(x_k)$ is the input with

$$\mathcal{U}(x_k) = \{u_k \text{ such that } u_k \neq 0 \text{ and } f(x_k, u_k) \in \mathcal{X}\}$$

The cost function is given by: $x_2^2 + \sum_{k=0}^{1} x_k^2 + u_k^2$.

For the following questions, the initial state is $x_0 = 0$.

i) In this problem, would an open-loop controller perform better than a closed-loop controller? **Justify** your answer. *[2 points]*

(A) ☐ Yes
(B) ☐ No

Justification:

ii) How many open-loop strategies exist? *[2 points]*

(A) ☐ 12
(B) ☐ 18
(C) ☐ 24
(D) ☐ 36
(E) ☐ 216

iii) How many closed-loop strategies exist? *[2 points]*

(A) ☐ 42^2
(B) ☐ 6^8
(C) ☐ 252
(D) ☐ 6^{14}
(E) ☐ 7^6

c) Consider the system *[9 points]*

$$x_{k+1} = f(x_k, u_k, w_k) = -x_k + 2u_k + w_k, \qquad k = 0, 1, \ldots, N-1,$$

where $x_k \in \mathcal{X} = \mathbb{R}$ is the state, $u_k \in \mathcal{U} = \mathbb{R}$ is the input, and $w_k \in \mathcal{W} \subset \mathbb{R}$ is the disturbance (independent of all previous disturbances, states and inputs).

The cost function is given by: $x_N^2 + \sum_{k=0}^{N-1} x_k^2 + u_k^2$.

i) The disturbance w_k has a mean $\mathrm{E}[w_k] = 0$ and variance $\mathrm{Var}[w_k] = 1$.
Calculate the optimal control input $\mu^*_{N-1}(x_{N-1})$ and cost-to-go $J^*_{N-1}(x_{N-1})$ as a function of the state x_{N-1} for $N = 3$. *[5 points]*

(A) $\mu^*_{N-1}(x_{N-1}) =$

(B) $J^*_{N-1}(x_{N-1}) =$

For the remaining question, we have $w_k = 0$ for all $k \in \{0, 1, \ldots, N-1\}$. Furthermore, assume that a P_k exists for all $k \in \{0, 1, \ldots, N-1\}$ such that the optimal cost-to-go function is

$$J^*_k(x_k) = P_k x_k^2.$$

ii) Calculate the optimal control input $\mu^*_k(x_k)$ as a function of x_k and P_{k+1}. **Justify** your answer. *[4 points]*

d) Consider the system *[9 points]*

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

where $x_k \in \mathcal{X} = \mathbb{R}$ is the state, $u_k \in \mathcal{U} = \mathbb{R}$ is the input, and $w_k \in \mathcal{W} \subset \mathbb{R}$ is the disturbance (independent of all previous disturbances, states and inputs) with a known mean $\mathrm{E}[w_k] = 0$ and variance $\mathrm{Var}[w_k] = 1$.

The cost function is given by: $x_2^2 + \sum_{k=0}^{1} x_k^2$.

We consider two control strategies **a** and **b** defined as:

$$\textbf{a}\text{: } \mu_k^a(x_k) = 2x_k - 1$$
$$\textbf{b}\text{: } \mu_k^b(x_k) = 3x_k.$$

i) Find the value(s) of the initial state x_0 for which the cost-to-go $J_0(x_0)$ is the same for both control strategies. **Justify** your answer. *[6 points]*

ii) In which range must the initial state x_0 lie such that control strategy **b** performs strictly better than control strategy **a**, in terms of the initial cost-to-go.

[3 points]

(A) ☐ $2 \leq x_0 \leq 5$

(B) ☐ $-\frac{1}{\sqrt{2}} < x_0 < \frac{1}{\sqrt{2}}$

(C) ☐ $-5 < x_0 < -2$

(D) ☐ $-\infty < x_0 < -\frac{1}{\sqrt{2}}$

(E) ☐ None of the above

Solution 1

a) i) The control spaces are

x_1	$x_2 = x_1 + y_1$	$\mathcal{U}_1(\tilde{x}_1)$
-3	-2	$\{-1, 1, 2, 3, 4, 5\}$
0	1	$\{-4, -3, -2, -1, 1, 2\}$
2	3	$\{-6, -5, -4, -3, -2, -1\}$

ii) The optimal cost-to-go function is

$$\begin{aligned} J_1^*(x_1, y_1 = 1) &= \min_{u_1 \in \mathcal{U}_1(\tilde{x}_1)} x_1^2 + u_1^2 + x_2 \\ &= \min_{u_1 \in \mathcal{U}_1(\tilde{x}_1)} x_1^2 + u_1^2 + (x_1 + y_1) \end{aligned}$$

For $x_1 = -3$ and $x_1 = 0$, this is minimized by $u_1 = -1$ and $u_1 = 1$; for $x_1 = 2$, this is minimized by $u_1 = -1$ only since $1 \notin \mathcal{U}_1(\tilde{x}_1)$.
See the following table for solutions:

x_1	$\mu_1^*(\tilde{x}_1)$
-3	$\{-1, 1\}$
0	$\{-1, 1\}$
2	$\{-1\}$

b) It is sufficient to define the following:

- $N = 2$ - the number of stages
- $N_u = 6$ - number of possible inputs at each time step (for any $x_k \in \mathcal{X}$ and $k \in \{0, 1\}$)
- $N_x = 7$ - number of possible states at each time step (for any $k \in \{0, 1\}$)

i) Open-loop - because the system is deterministic

ii) Option (D)

$$N_u^N = 6^2 = 36$$

iii) Option (B)

$$N_u(N_u^{N_x})^{N-1} = 6(6^7)^1 = 6^8$$

c) Notice the following relation:

$$\mathrm{Var}[w_k] = 1 = \mathrm{E}[w_k^2] - \mathrm{E}[w_k]^2 = \mathrm{E}[w_k^2]$$

i)

$$\begin{aligned}
J_{N-1}(x_{N-1}) &= \min_{u_{N-1}\in\mathbb{R}} x_{N-1}^2 + u_{N-1}^2 + \mathrm{E}\left[(-x_{N-1} + 2u_{N-1} + w_{N-1})^2\right] \\
&= \min_{u_{N-1}\in\mathbb{R}} x_{N-1}^2 + u_{N-1}^2 + x_{N-1}^2 + 4u_{N-1}^2 + \mathrm{E}[w_{N-1}^2] \\
&- 4x_{N-1}u_{N-1} - 2x_{N-1}\mathrm{E}[w_{N-1}] + 4u_{N-1}\mathrm{E}[w_{N-1}] \\
&= \min_{u_{N-1}\in\mathbb{R}} x_{N-1}^2 + u_{N-1}^2 + x_{N-1}^2 + 4u_{N-1}^2 + 1 - 4x_{N-1}u_{N-1} - 0 + 0 \\
&= \min_{u_{N-1}\in\mathbb{R}} 5u_{N-1}^2 - 4x_{N-1}u_{N-1} + 2x_{N-1}^2 + 1 \\
&= (5(0.4)^2 - 4(0.4) + 2)x_{N-1}^2 + 1 \\
&= 1.2 \cdot x_{N-1}^2 + 1
\end{aligned}$$

With

$$\begin{aligned}
\frac{d(\dots)}{du_{N-1}} &= 10u_{N-1} - 4x_{N-1} = 0 \\
\mu_{N-1}^*(x_{N-1}) &= u_{N-1} = 0.4x_{N-1}
\end{aligned}$$

ii)

$$\begin{aligned}
\mu_k^*(x_k) &= \arg\min_{u_k\in\mathbb{R}} x_k^2 + u_k^2 + P_{k+1}(-x_k + 2u_k)^2 \\
&= \arg\min_{u_k\in\mathbb{R}} x_k^2 + u_k^2 + P_{k+1}(x_k^2 - 4x_k u_k + 4u_k^2) \\
&= \frac{2P_{k+1}}{4P_{k+1}+1} x_k
\end{aligned}$$

With

$$\begin{aligned}
\frac{d(\dots)}{du_k} &= 2(1 + 4P_{k+1})u_k - 4P_{k+1}x_k = 0 \\
\mu_k^*(x_k) &= u_k = \frac{2P_{k+1}}{4P_{k+1}+1} x_k
\end{aligned}$$

d) Notice the following relation:

$$\mathrm{Var}[w_k] = 1 = \mathrm{E}[x^2] - \mathrm{E}[x]^2 = \mathrm{E}[x^2]$$

i) For strategy **a**:

$$\begin{aligned} x_{k+1} &= 2x_k - (2x_k - 1) + w_k = 1 + w_k \\ J_1^a(x_1) &= x_1^2 + \mathrm{E}[(1 + w_1)^2] \\ &= x_1^2 + 1 + 2\mathrm{E}[w_1] + \mathrm{E}[w_1^2] \\ &= x_1^2 + 2 \\ J_0^a(x_0) &= x_0^2 + \mathrm{E}[J_1^a(1 + w_0)] \\ &= x_0^2 + \mathrm{E}[(1 + w_0)^2 + 2] \\ &= x_0^2 + 1 + 2\mathrm{E}[w_0] + \mathrm{E}[w_0^2] + 2 \\ &= x_0^2 + 4 \end{aligned}$$

For strategy **b**:

$$\begin{aligned} x_{k+1} &= 2x_k - (3x_k) + w_k = -x_k + w_k \\ J_1^b(x_1) &= x_1^2 + \mathrm{E}[(-x_1 + w_1)^2] \\ &= x_1^2 + x_1^2 - 2x_1\mathrm{E}[w_1] + \mathrm{E}[w_1^2] \\ &= 2x_1^2 + 1 \\ J_0^b(x_0) &= x_0^2 + \mathrm{E}[J_1^b(-x_0 + w_0)] \\ &= x_0^2 + \mathrm{E}[2(-x_0 + w_0)^2 + 1] \\ &= x_0^2 + 2x_0^2 - 4x_0\mathrm{E}[w_0] + 2\mathrm{E}[w_0^2] + 1 \\ &= 3x_0^2 + 3 \end{aligned}$$

Combining the two:

$$\begin{aligned} J_0^a(x_0^*) &= J_0^b(x_0^*) \\ (x_0^*)^2 + 4 &= 3(x_0^*)^2 + 3 \\ 2(x_0^*)^2 &= 1 \\ x_0^* &= \pm\frac{1}{\sqrt{2}} \end{aligned}$$

ii) Strategy **b** performs better than strategy **a** when $J_0^a(x_0) - J_0^b(x_0) > 0$ Reusing the result of the previous question, this results in:

$$\begin{aligned} &- 2x_0^2 + 1 > 0 \\ \Leftrightarrow &-\frac{1}{\sqrt{2}} < x_0 < \frac{1}{\sqrt{2}} \end{aligned}$$

Correct answer is: Option B)

Problem 2 [20 points]

For all questions related to the Label Correcting Algorithm, if two nodes enter the OPEN bin in the same iteration, the one with the lowest node index is added first.

a) Consider the deterministic Shortest Path problem with parameters $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$, and $\gamma \in \mathbb{R}$, whose graph is depicted in Figure 1. *[4 points]*

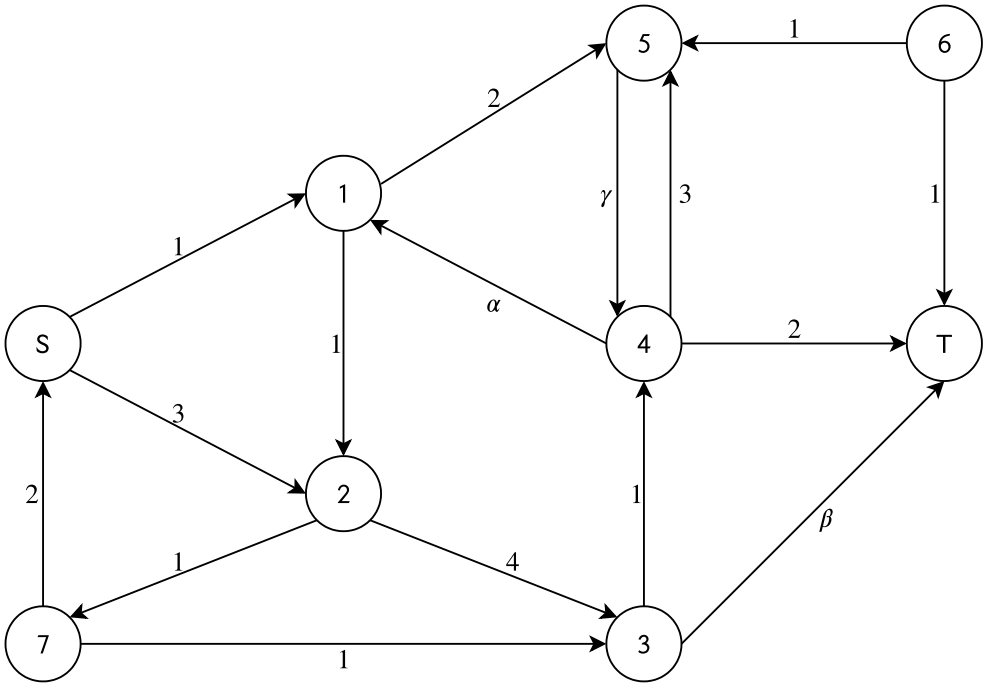

Figure 1: Graph for the deterministic Shortest Path problem a).

In what range must the parameters α, β, and γ reside to have a valid Shortest Path problem? **Justify** your answer.

b) Consider the deterministic Shortest Path problem, whose graph is depicted in Figure 2. *[12 points]*

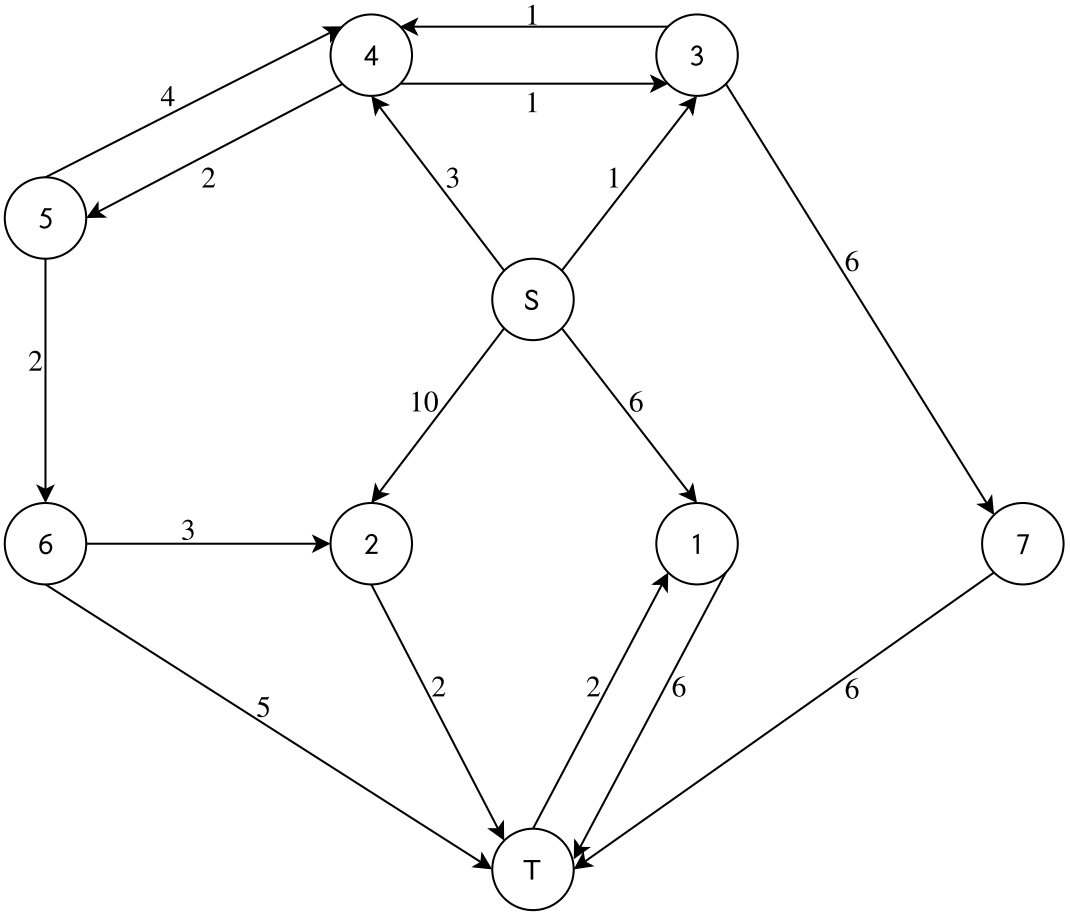

Figure 2: Graph for the deterministic Shortest Path problem b).

i) What is the shortest path from node s to node t and its associated cost? **Justify** your answer by filling in the Label Correcting Algorithm Table below using a **breadth-first** search strategy to remove nodes from the OPEN bin.

[6 points]

Iteration	Remove	OPEN	d_{s}	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_{t}

Path:

Cost:

ii) We would like to convert the given Shortest Path problem into an equivalent Deterministic Finite State problem. N is the smallest possible time horizon for the resulting Deterministic Finite State formulation.
Give the values of $J_N(i)$, $J_{N-1}(i)$, and $J_{N-2}(i)$ of the resulting Deterministic Finite State problem for every node $i \in \mathcal{V} \backslash \{\mathsf{T}\}$, where $\mathcal{V}$ is the set containing all the nodes of the graph. *[6 points]*

c) Answer the following True or False questions. If the answer is False, you need to **justify why**. You will get 1 point for each correct answer, 0 points for each blank answer, and -0.5 point for each incorrect or ambiguous answer. The minimum total number of points is 0.
For the questions related to the Label Correcting Algorithm, assume that all edges have non-negative cost.

i) In Deterministic Finite State problems, the disturbances can have a uniform distribution.

☐ True
☐ False: _____

ii) For a given Shortest Path problem, the A* algorithm always requires fewer iterations than the Label Correcting Algorithm employing a Best-First Search strategy.

☐ True
☐ False: _____

iii) You employ the Label Correcting Algorithm on a graph that has a tree-like structure, i.e. there is a unique path from s to every other node $i \in \mathcal{V}\backslash\{\mathsf{t}\}$. If you use the depth-first strategy to remove the nodes from the OPEN bin, then nodes will never enter the OPEN bin more than once.

☐ True
☐ False: _____

iv) Consider the following removal strategy of nodes from the OPEN bin of the Label Correcting Algorithm: remove node $i^* = \arg\max_{i \in \text{OPEN}} d_i$. Then the Label Correcting Algorithm terminates in a finite number of steps.
☐ True
☐ False: _____

Solution 2

a) i) In order for the SP problem to be valid, there can't be negative cycles in the graph. As for the parameter α, since the edge $4 \to 1$ appears in both the cycle $1 \to 2 \to 3 \to 4$ and $1 \to 5 \to 4$, the condition on α is given by the following system of equations.

$$\begin{cases} c_{12} + c_{27} + c_{73} + c_{34} + c_{41} \geq 0 \\ c_{12} + c_{23} + c_{34} + c_{41} \geq 0 \\ c_{15} + c_{54} + c_{41} \geq 0 \\ c_{45} + c_{54} \geq 0 \end{cases}$$

$$\begin{cases} 1 + 1 + 1 + 1 + \alpha \geq 0 \\ 1 + 4 + 1 + \alpha \geq 0 \\ 2 + \gamma + \alpha \geq 0 \\ 3 + \gamma \geq 0 \end{cases}$$

$$\begin{cases} \alpha \geq -4 \\ \alpha \geq -6 \\ \gamma + \alpha \geq -2 \\ \gamma \geq -3 \end{cases}$$

$$\begin{cases} \alpha \geq -4 \\ \gamma + \alpha \geq -2 \\ \gamma \geq -3 \end{cases}$$

Above are the conditions on α and γ. On the other hand, there are no conditions on β, since the edge $3 \to \mathsf{T}$ doesn't appear in any cycle. Hence, $\beta \in \mathbb{R}$.

b) i) Solve the SP problem with the LCA algorithm, BFS strategy.

Iteration	Remove	OPEN	d_{S}	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_{T}
0	–	S	0	∞	∞	∞	∞	∞	∞	∞	∞
1	S	1,2,3,4	0	6	10	1	3	∞	∞	∞	∞
2	1	2,3,4	0	6	10	1	3	∞	∞	∞	12
3	2	3,4	0	6	10	1	3	∞	∞	∞	12
4	3	4,7	0	6	10	1	2	∞	∞	7	12
5	4	7,5	0	6	10	1	2	4	∞	7	12
6	7	5	0	6	10	1	2	4	∞	7	12
7	5	6	0	6	10	1	2	4	6	7	12
8	6	2	0	6	9	1	2	4	6	7	11
9	2	–	0	6	9	1	2	4	6	7	11

The shortest path from S to T is $\mathsf{S} \to 3 \to 4 \to 5 \to 6 \to \mathsf{T}$. It has a cost of 11.

ii) Denoting by $\mathcal{V}$ the vertex space, $N := |\mathcal{V}|-1 = 9-1 = 8$. You were asked to calculate $J_8(x_8)$, $J_7(x_7)$, and $J_6(x_6)$.$J_8(x_8)$ is equal to the terminal cost of the DFS problem $g_N(x_8) := 0$, $\forall x_8 \in \mathcal{S}_8 = \{\mathsf{T}\}$. To calculate J_7 and J_6 we will use the following relation:

$$J_k(i) = \min_{j \in \mathcal{V}\backslash\{\mathsf{T}\}} (c_{i,j} + J_{k+1}(j)), \quad \forall i \in \mathcal{V}\backslash\{\mathsf{T}\},\ k = 7, 6$$

i	$J_7(i)$	$\mu_7(i)$
S	∞	T
1	6	T
2	2	T
3	∞	T
4	∞	T
5	∞	T
6	5	T
7	6	T

i	$J_6(i)$	$\mu_6(i)$
S	12	1,2
1	6	1
2	2	2
3	12	7
4	∞	3, 4, 5
5	7	6
6	5	2,6
7	6	7

c) i) False. In a DFS, there can't be any disturbances.

ii) False.

iii) True.

iv) True.

Problem 3 [25 points]

Consider the discounted Stochastic Shortest Path problem we discussed in class, which is represented below.
The system dynamics are

$$\begin{aligned} &x_{k+1} = w_k, \qquad x_k \in \mathcal{S}, \\ &\Pr\left(w_k = j | x_k = i, u_k = \mathrm{u}\right) = P_{ij}(\mathrm{u}), \qquad \mathrm{u} \in \mathcal{U}(i), \end{aligned} \tag{1}$$

where $\mathcal{S}$ is a finite set and $\mathcal{U}(\mathrm{x})$ is a finite set for all $\mathrm{x} \in \mathcal{S}$.
Given an initial state $i \in \mathcal{S}$, the expected closed loop cost of starting at i associated with a policy $\pi = (\mu_0(\cdot), \mu_1(\cdot), \ldots, \mu_{N-1}(\cdot))$ is

$$J_\pi(i) = \underset{(X_1, W_0 | x_0 = i)}{E} \left[\sum_{k=0}^{N-1} \alpha^k g(x_k, \mu_k(x_k), w_k) \right], \quad \textit{subject to } (1), \tag{2}$$

where $g(\cdot,\cdot,\cdot)$ is some given function and $0 < \alpha \leq 1$ is the discount factor.
The goal is to construct an optimal policy π^* such that for all $i \in \mathcal{S}$,

$$\pi^* = \arg\min_{\pi \in \Pi} J_\pi(i),$$

and explore what happens as N goes to infinity.
Furthermore, for the case where $\alpha = 1$, two assumptions on the problem data are made:

Assumption 1. *There exists a cost-free termination state, which we designate as state 0. In particular, there are $n+1$ states with $\mathcal{S} = \{0, 1, \ldots, n\}$, where*

$$P_{00}(u) = 1 \text{ and } g(0, \mathrm{u}, 0) = 0 \quad \text{for all } \mathrm{u} \in \mathcal{U}(0).$$

Assumption 2. *There exists at least one proper policy $\mu \in \Pi$. Furthermore, for every improper policy μ', the corresponding cost function $J_{\mu'}(i)$ is infinity for at least one state $i \in \mathcal{S}$.*

For questions a) and b) consider the undiscounted problem, i.e. $\alpha = 1$.

a) Answer the following True or False questions. If the answer is False, you need to **justify** why. You will get 1 point for each correct answer, 0 points for each blank answer, and -0.5 points for each incorrect or ambiguous answer. The minimum total number of points is 0.
Assume that an optimal solution exists and that unless stated otherwise, the algorithms are initialized properly.

i) Policy Iteration and Value Iteration always converge towards the same cost.

☐ True
☐ False: _____

ii) Policy Iteration and Value Iteration always converge towards the same policy.

☐ True
☐ False: _____

iii) Using Value Iteration, $J^{k+1}(i) \leq J^k(i)$ for all i, where k is the iteration of the algorithm.

☐ True
☐ False: _____

iv) Using Policy Iteration, $J_{\mu^{k+1}}(i) \leq J_{\mu^k}(i)$ for all i, where k is the iteration of the algorithm.

☐ True
☐ False: _____

v) Solving a given problem with Policy Iteration is always more computationally efficient than solving it with Value Iteration.

☐ True
☐ False: _____

vi) Value Iteration can be initialized arbitrarily.

☐ True
☐ False: _____

vii) Policy Iteration can be initialized with any admissible policy.

☐ True
☐ False: _____

b) Explain why you have to remove the terminal state from the state space in the Policy Iteration algorithm. *[4 points]*

c) The goal of this question is to model a Stochastic Shortest Path problem. *[14 points]* The problem is a simplified one-vs-one basketball match where the players move in 1 dimension. You control one of the players. The game has the following mechanics:

- At each time step, both players can move (by one cell in a discretized grid) left, right, or, if they have the ball, attempt to shoot in the hoop. They move simultaneously and can occupy the same cell.
- The probability of getting a point when shooting is $p_{score}(d_1, d_2)$, a function of d_1, the distance between the player and the corresponding hoop, and d_2, the distance between the players. After a shooting attempt, the other player receives the ball.
- At each time step, the referee can decide to stop the match with probability p_{stop}.

You can consider the other player to be a random agent (it selects one of the actions randomly at each time step).
The objective is to maximize, in your favor, the difference in points between your player and the other player.

The problem is depicted in Figure 3.

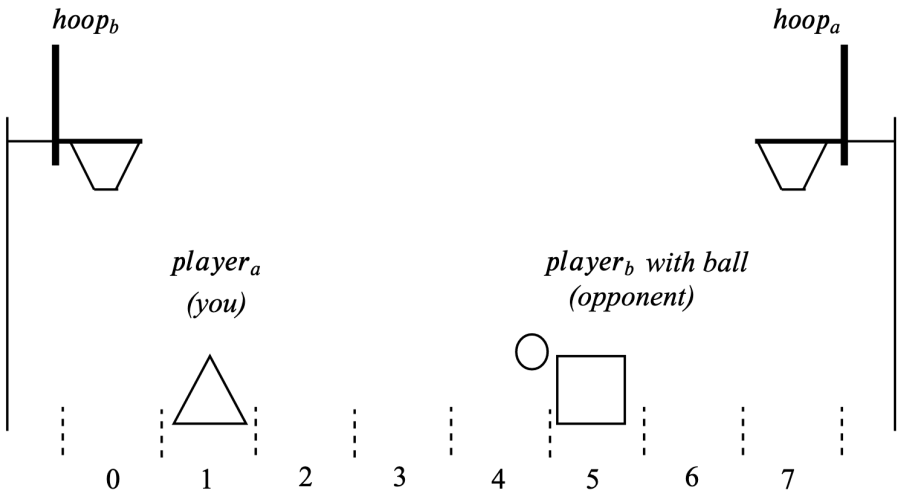

Figure 3: Basketball match set-up. You are controlling player$_a$, and player$_b$ is the opponent. player$_a$ starts at cell 1. player$_b$ starts at cell 5 with the ball.

i) State a possible state space $\mathcal{X}$ for this problem. *Note*: Multiple solutions are possible, give only one. *[1 point]*

ii) State the input space $\mathcal{U}(i)$ for all $i \in \mathcal{X}$ for this problem. *[1 point]*

iii) State a possible expected stage cost $q(i, u)$ for all $i \in \mathcal{X}$ and $u \in \mathcal{U}(i)$ for this problem. *Note*: Multiple solutions are possible, give only one. *[2 points]*

iv) Figure 3 describes the initial conditions at the beginning of time step 0 and you select the input $u_0 = \textit{move right}$. Write the states that can be reached at the beginning of time step 1 **and** the probabilities of the corresponding transitions. *[3 points]*

v) What is the range of discount factors α that guarantees that the optimal solution has finite expected cost when $p_{stop} = 0$? **Justify** your answer. *[1 point]*

vi) What is the range of discount factors α that guarantees that the optimal solution has finite expected cost when $0 < p_{stop} < 1$? **Justify** your answer *[1 point]*

For the remaining questions, consider two variations of the problem:

- Problem A: we have $\alpha = 0.9$ and $p_{stop} = 0$
- Problem B: we have $\alpha = 1$ and $p_{stop} = 0.1$

vii) Answer the following True or False questions. You will get 1 point for each correct answer, 0 points for each blank answer, and -0.5 points for each incorrect or ambiguous answer. The minimum total number of points is 0.

1. Problems A and B have the same optimal cost-to-go.

   ☐ True
   ☐ False

2. Problems A and B have the same optimal policy.

   ☐ True
   ☐ False

3. Problems A and B have the same transition probability matrix.

   ☐ True
   ☐ False

viii) Do you need to carefully select the initial policy when using Policy Iteration to solve problem B? **If yes**, provide a correct choice of policy. *[2 points]*

Solution 3

a) i) True

ii) False - potentially multiple optimal policies

iii) False - we can initialize J with any values, so it might increase or decrease depending on the initial values.

iv) True

v) False - depends on the problem

vi) True

vii) False - only proper policies

b) Policy iteration fails because we are inverting $(I - P)$, but $\det(I - P) = 0$ since in P the row of the terminal state has all zeros except a single 1 on the diagonal.

c) i) $\mathcal{S} = \{position_{player_a}, position_{player_b}, who\ has\ the\ ball\}$ Other correct answers are possible. The state space must completely define the stage cost and the transition probability matrix. Since we are trying to maximize the point difference regardless of the score itself, the number of points of the players is not important and should not be included in the state space.

ii) The control space is $\mathcal{U} = \{move\ left, move\ right, shoot\}$ if $player_a$ has the ball, and $\mathcal{U} = \{move\ left, move\ right\}$ if $player_a$ does not have the ball.

iii) The general rule is: cost of 1 if the $player_b$ scores, -1 if you score. If $player_a$ has the ball, you can only score if you decide to shoot: $q(i, u) = -p_{score_a}$ if $u = shoot$ and 0 otherwise.
If player b has the ball, there is a 1/3 probability that it will decide to shoot: $q(i, u) = 1/3 p_{score_b}$

iv) There are 4 possible states at the beginning of step 1, 3 states if the match continues, all with equal probability $p = (1 - p_{stop})/3$:
x_1 =[2, 5, $player_a$ has the ball] $player_b$ decided to shoot
x_1 =[2, 4, $player_b$ has the ball] $player_b$ decided to move left
x_1 =[2, 6, $player_b$ has the ball] $player_b$ decided to move right
And the case where the match is stopped, with probability p_{stop}:
$x_1 = TERIMNAL$

v) $0 < \alpha < 1$ There is no terminal state in this case. The cost-to-go will be infinite for $alpha = 1$.

vi) $0 < \alpha \leq 1$ There is always a non-zero probability of going to the terminal state, as such the cost-to-go is finite.

vii) 1. False - need re-scaling with α
2. True
3. False - Probabilities are lower in B since there is a chance of going to the terminal state.

viii) Any admissible policy is proper and, as such, correct since the termination is independent of the policy, there is always a chance of going to the terminal state. There is no need to carefully select the initial policy.

Problem 4 [25 points]

Consider a control maneuver for a rocket in space. The rocket can move in a straight line by applying a force with its thrusters. The simplified dynamics are

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

where $\ddot{x}(t)$ is the acceleration, $\dot{x}(t)$ the velocity, $x(t)$ the position, and $u(t)$ the normalized thrust of the spacecraft with $|u(t)| \leq u_{\max}$, $u_{\max} > 0$. We define the state of the system as $\mathbf{x}(t) = [x(t), \dot{x}(t)]$.
In this maneuver, the rocket has to be brought from the initial state

$$\mathbf{x}(0) = [0, 0]$$

to the terminal state

$$\mathbf{x}(T) = [2, 0]$$

in a fixed amount of time T, while minimizing the total control effort

$$\int_0^T |u(t)| dt.$$

The goal of this problem is to derive a control trajectory for that maneuver using Pontryagin's Minimum Principle.
The co-state of that problem is defined as $\mathbf{p}(t) = [p_1(t), p_2(t)]$ for all $t \in [0, T]$.

a) Write the system dynamics $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), u(t))$. *[2 points]*

b) Write the Hamiltonian $H(\mathbf{x}(t), u(t), \mathbf{p}(t))$ of the problem. *[2 points]*

c) Write the co-state derivative $\dot{\mathbf{p}}(t)$ of the problem. *[3 points]*

d) Draw a possible curve of $p_2(t)$ for all $t \in [0, T]$ directly in Figure 4. You do not need to give the values of the intersections of $p_2(t)$ with the axes. **Justify** your answer. *[9 points]*

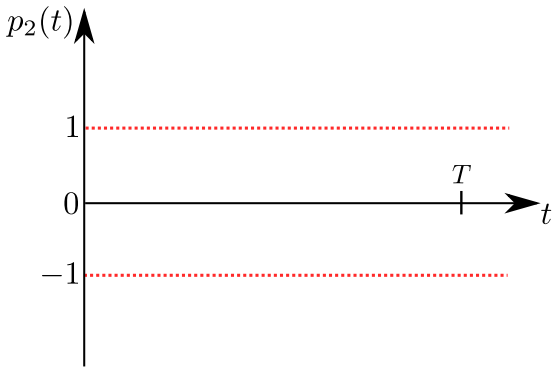

Figure 4: Plot of $p_2(t)$.

Justification:

e) Assume that $u_{max} = 2$, $T = 4$.
Derive the control input $u(t)$ for all $t \in [0, T]$ that satisfies the minimum principle. **Justify** your answer. *[9 points]*

Solution 4

a) We define $\mathbf{x}(t) = [x_1(t), x_2(t)]$.
The dynamics are $\dot{x}_1(t) = x_2(t)$, $\dot{x}_2(t) = u(t)$

b) The Hamiltonian is

$$\begin{aligned} H(\mathbf{x}, u, \mathbf{p}) &= |u(t)| + p_1(t)\dot{x}_1(t) + p_2(t)\dot{x}_2(t) \\ &= |u(t)| + p_1(t)x_2(t) + p_2(t)u(t) \end{aligned}$$

c) The co-state derivatives are defined by

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

so we have $\dot{p}_1(t) = 0$, $\dot{p}_2(t) = -p_1(t)$.

d) We first solve the co-state differential equations

$$\begin{aligned} p_1(t) &= C_1 \\ p_2(t) &= -C_1 t + C_2 \end{aligned}$$

According to the minimum principle

$$\begin{aligned} u(t) &= \underset{|u| \le u_{\max}}{\arg\min}\, H(\mathbf{x}, u, \mathbf{p}) \\ &= \underset{|u| \le u_{\max}}{\arg\min}\, |u(t)| + p_2(t)u(t) \end{aligned}$$

We thus have

$$u(t) = \begin{cases} -u_{\max} & \text{if } p_2(t) > 1 \\ 0 & \text{if } -1 \le p_2(t) \le 1 \\ u_{\max} & \text{if } p_2(t) < -1 \end{cases} \tag{3}$$

The intuitive behaviour is the following: the spacecraft first accelerates towards the goal, then applies no thrust, and finally decelerates to reach zero velocity at the goal. Combining this intuition and using Eq. 3, we must first have $p_2(t) < -1$, then $-1 \le p_2(t) \le 1$, and finally $p_2(t) > 1$. Anything else would result in not being able to meet the terminal conditions. As a result, and since $p_2(t)$ is an affine function, it has the following curve:

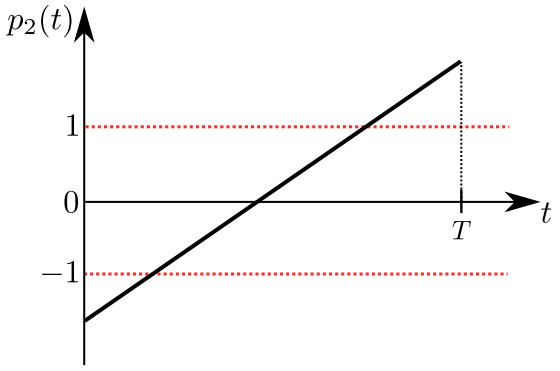

Figure 5: Plot of $p_2(t)$.

e) As shown before we have three regular arcs where

$$u(t) = \begin{cases} u_{\max} & \text{for } t \in [0, t_1] \\ 0 & \text{for } t \in [t_1, t_2] \\ -u_{\max} & \text{for } t \in [t_2, T] \end{cases} \tag{4}$$

and $t_1 \leq t_2$. We have to find t_1 and t_2.

- First arc $t \in [0, t_1]$
  We have $\dot{x}_2(t) = u_{\max}$ with the boundary conditions $x_1(0) = 0$ and $x_2(0) = 0$.
  As a result we have

$$x_1(t) = \frac{1}{2} u_{\max} t^2$$
$$x_2(t) = u_{\max} t$$

- Second arc $t \in [t_1, t_2]$
  We have $\dot{x}_2(t) = 0$. Because of continuity between arc 1 and 2, the boundary conditions are $x_1(t_1) = \frac{1}{2} u_{\max} t_1^2$ and $x_2(t_1) = u_{\max} t_1$. As a result we have

$$x_1(t) = u_{\max} t_1 \left(t - \frac{1}{2} t_1 \right)$$
$$x_2(t) = u_{\max} t_1$$

- Third arc $t \in [t_2, T]$
  We have $\dot{x}_2(t) = -u_{\max}$ and thus

$$x_1(t) = -\frac{1}{2} u_{\max} t^2 + C_3 t + C_4$$
$$x_2(t) = -u_{\max} t + C_3.$$

  Due to the terminal conditions, we have $x_1(T) = 2$ and $x_2(T) = 0$.
  Calculate C_3:
  $x_2(T) = 0 \Leftrightarrow C_3 = u_{\max} T$
  $x_1(t) = -\frac{1}{2} u_{\max} t^2 + u_{\max} T t + C_4$
  $x_2(t) = u_{\max}(T - t)$

  Calculate C_4:
  $x_1(T) = 2 \Leftrightarrow C_4 = 2 - \frac{1}{2} u_{\max} T^2$
  $x_1(t) = 2 + u_{\max}(-\frac{1}{2} t^2 + Tt - \frac{1}{2} T^2)$

  Because of continuity between arc 2 and 3, we have $x_1(t_2) = u_{\max} t_1 \left(t_2 - \frac{1}{2} t_1 \right)$ and $x_2(t_2) = u_{\max} t_1$. We use these boundary conditions to find t_1 and t_2
  $u_{\max} t_1 = u_{\max}(T - t_2) \Leftrightarrow t_1 = (T - t_2)$

$$
\begin{aligned}
& u_{\max} t_1 \left(t_2 - \frac{1}{2} t_1 \right) = 2 + u_{\max}(-\frac{1}{2} t_2^2 + T t_2 - \frac{1}{2} T^2) \\
\Leftrightarrow & u_{\max}(T - t_2) \left(t_2 - \frac{1}{2}(T - t_2) \right) = 2 + u_{\max}(-\frac{1}{2} t_2^2 + T t_2 - \frac{1}{2} T^2) \\
\Leftrightarrow & (4 - t_2) \left(t_2 - \frac{1}{2}(4 - t_2) \right) = 1 + (-\frac{1}{2} t_2^2 + 4 t_2 - 8) \\
\Leftrightarrow & 4 t_2 - 8 + 2 t_2 - t_2^2 + 2 t_2 - \frac{1}{2} t_2^2 = -7 - \frac{1}{2} t_2^2 + 4 t_2 \\
\Leftrightarrow & t_2^2 - 4 t_2 + 1 = 0 \\
\Leftrightarrow & t_2 = 2 \pm \sqrt{3}
\end{aligned}
$$

The only possibility is that $t_2 = 2 + \sqrt{3}$, otherwise we would have $t_2 < t_1$.

Solution:

$$
u(t) = \begin{cases} u_{\max} & \text{for } t \in [0, 2 - \sqrt{3}] \\ 0 & \text{for } t \in [2 - \sqrt{3}, 2 + \sqrt{3}] \\ -u_{\max} & \text{for } t \in [2 + \sqrt{3}, 4] \end{cases} \tag{5}
$$