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 27th, 2023**

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

Prof. R. D'Andrea

Exam

Exam Duration: **150 minutes**

Number of Problems: **4**

Permitted aids: A single A4 sheet of paper (double sided; hand-written or computer typed).

No calculators allowed.

Problem 1 [25 points]

a) Assume that we have a vessel with a maximum weight capacity of W. We have K different items. For each item, we can either load zero or one copy of the item onto the vessel. Let v_i be the known value of item i, let w_i bet the known weight of item i, and let n_i be the number of copies of item i loaded, $n_i \in \{0, 1\}$. Our goal is to find the most valuable cargo for the vessel while not exceeding the weight limit, i.e., we want to maximize $\sum_{i=1}^{K} n_i v_i$ subject to the constraints $\sum_{i=1}^{K} n_i w_i \leq W$ and $n_i \in \{0, 1\}$.

We want to model this problem as a standard Dynamic Programming problem, where our utility function expresses the total value of our cargo (we maximize the total utility in a similar way to minimizing the cost in the standard definition).

i) Formulate this problem as a Dynamic Programming problem. Define the time horizon, state(s) and state space, initial state, control input(s) and control space, system dynamics, stage utility, and terminal utility.
Hint: Choose the state to be 1-dimensional. *[8 points]*

ii) Can closed-loop control yield a higher total value than open-loop control for this problem? Provide a short explanation. *[2 points]*

iii) Assume that you solved the above problem using the Dynamic Programming Algorithm and have thus obtained an optimal closed-loop policy. Due to unforeseen circumstances, one of the items that would have originally been loaded for optimal value becomes unavailable and may not be loaded onto the vessel. You aim to retain an optimal closed-loop policy; under what circumstance do you NOT have to recompute the policy? Provide a short explanation. *[2 points]*

b) Consider the dynamics

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

for some finite N, where $x_k, u_k, w_k \in \mathbb{R}$. The disturbance w_k is a random walk with Gaussian steps and can be modeled as follows:

$$\begin{aligned} w_k &= y_{k+1}, \\ y_{k+1} &= y_k + \xi_k, \end{aligned}$$

where $y_k, \xi_k \in \mathbb{R}$ and $\xi_k \sim \mathcal{N}(0, 1), k = 0, \ldots, N-1$, are independent Gaussian random variables with mean 0 and variance 1. We assume that at time k, we can observe y_k.

The cost function is given by

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

i) Explain what the above cost function penalizes. *[2 points]*

ii) Let the augmented state be $\tilde{x}_k := (x_k, y_k)$. State the dynamics $\tilde{f}_k(\tilde{x}_k, u_k, \xi_k)$ of the augmented state. *[2 points]*

iii) Initialize the Dynamic Programming Algorithm, i.e., find $J_N(\mathrm{x}, \mathrm{y})$. *[1 point]*

iv) Apply the Dynamic Programming Algorithm recursion step for $k = N-1$, i.e., find $J_{N-1}(\mathrm{x}, \mathrm{y})$ and $\mu_{N-1}^*(\mathrm{x}, \mathrm{y})$. *[8 points]*

Solution 1

a) i)

- The time horizon is the number of items $N = K$.
- The state x_k denotes the total weight of the cargo for items $1, \dots, k$, with $x_0 = 0$.
- The state space is given by

$$\mathcal{S}_k = [0, W], \quad k = 1, \dots, N.$$

- The control input denotes the number of copies of item $k+1$, i.e., $u_k = n_{k+1}$.
- The control space is given by

$$\mathcal{U}_k(x_k) = \begin{cases} \{0,1\}, & \text{if } x_k + w_{k+1} \leq W \\ \{0\}, & \text{otherwise} \end{cases}, \quad k = 0, \dots, N-1.$$

- The dynamics are given by

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

- Stage utility is given by $g_k(x_k, u_k, w_k) = u_k v_{k+1}$.
- The terminal utility is $g_N(x_N) = 0$.

Note: The definitions chosen above are not the only possible solution. For example, the order of items to be loaded onto the vessel can be chosen arbitrarily. Another solution would be to choose the state space as $\mathcal{S}_k = \mathbb{R}^+$ the control space as $\mathcal{U}_k = \{0, 1\}$ and define the terminal utility to be

$$g_N(x_N) = \begin{cases} -\infty, & \text{if } x_N > W \\ 0, & \text{otherwise} \end{cases}.$$

ii) In theory no, since this is a deterministic problem.

iii) Since the dynamics stated above (especially the ordering of loading the items onto the vessel) is not unique, the following general answer holds: If the new problem of loading the vessel with one item becoming unavailable can be posed as a "tail subproblem" of the original problem starting at $k = i$ for some i, the Principle of Optimality applies and we do not have to recompute the policy; the truncated policy $\{\mu_i^*, \dots, \mu_{N-1}^*\}$ is optimal for the subproblem.
If the problem is defined as above, then if item 1 becomes unavailable, loading items $2, \dots, K$ is a tail subproblem of the original problem starting at $k = 1$; then we do not have to recompute the policy.

b) i) The cost penalizes the control effort and the final state error, i.e., the deviation of the final state from zero; the solution will be a trade-off between minimizing the input and minimizing the final state error.

ii) The dynamics of the augmented system are given by

$$\tilde{x}_{k+1} = \begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k + u_k(y_k + \xi_k) \\ y_k + \xi_k \end{bmatrix}.$$

iii) We initialize the DPA with

$$J_N(\mathrm{x}, \mathrm{y}) = \mathrm{x}^2.$$

iv) We apply the DPA for $k = N - 1$:

$$\begin{aligned}
J_{N-1}(\mathrm{x},\mathrm{y}) &= \min_{\mathrm{u}} \mathop{\mathrm{E}}_{\xi_{N-1}} \left[\mathrm{u}^2 + J_N(\mathrm{x}+\mathrm{u}(\mathrm{y}+\xi_{N-1}), \mathrm{y}+\xi_{N-1})\right] \\
&= \min_{\mathrm{u}} \mathop{\mathrm{E}}_{\xi_{N-1}} \left[\mathrm{u}^2 + (\mathrm{x}+\mathrm{u}(\mathrm{y}+\xi_{N-1}))^2\right] \\
&= \min_{\mathrm{u}} \mathop{\mathrm{E}}_{\xi_{N-1}} \left[\mathrm{u}^2 + \mathrm{x}^2 + 2\mathrm{xu}(\mathrm{y}+\xi_{N-1}) + \mathrm{u}^2(\mathrm{y}^2 + 2\mathrm{y}\xi_{N-1} + \xi_{N-1}^2)\right] \\
&= \min_{\mathrm{u}} \mathrm{u}^2 + \mathrm{x}^2 + 2\mathrm{xu}(\mathrm{y}+\mathrm{E}[\xi_{N-1}]) + \mathrm{u}^2(\mathrm{y}^2 + 2\mathrm{y}\,\mathrm{E}[\xi_{N-1}] + E[\xi_{N-1}^2]) \\
&= \min_{\mathrm{u}} \mathrm{u}^2 + \mathrm{x}^2 + 2\mathrm{xuy} + \mathrm{u}^2(\mathrm{y}^2+1) \\
&= \min_{\mathrm{u}} \underbrace{\mathrm{x}^2 + 2\mathrm{xuy} + \mathrm{u}^2(\mathrm{y}^2+2)}_{:=C(\mathrm{u})}
\end{aligned}$$

The minimum is attained at a u for which the gradient with respect to u is zero;

$$\begin{aligned}
\frac{\partial C(\mathrm{u})}{\partial \mathrm{u}} &= 0 \\
2(\mathrm{xy} + \mathrm{u}(\mathrm{y}^2+2)) &= 0 \\
\implies \mathrm{u} &= -\frac{\mathrm{xy}}{\mathrm{y}^2+2}
\end{aligned}$$

Note also that the second partial derivate

$$\frac{\partial^2 C(\mathrm{u})}{\partial \mathrm{u}^2} = 2(\mathrm{y}^2+2) > 0,$$

is strictly positive, which can be seen from the fact that $C(\mathrm{u})$ is quadratic in u. Therefore, $\mathrm{u} = -\frac{\mathrm{xy}}{\mathrm{y}^2+2}$ is optimal and we can substitute it into $C(\mathrm{u})$ to get

$$\begin{aligned}
\mu^*_{N-1}(\mathrm{x},\mathrm{y}) &= -\frac{\mathrm{xy}}{\mathrm{y}^2+2}, \\
J_{N-1}(\mathrm{x},\mathrm{y}) &= \mathrm{x}^2 - \frac{\mathrm{x}^2\mathrm{y}^2}{\mathrm{y}^2+2}.
\end{aligned}$$

Problem 2 [25 points]

a) The state of a stock market x_k at time k can be categorized into up (1), down (2), or closed (0). A trader has three options u_k at each time step: invest (I), perform shorting (S), or do nothing (N). Depending on the state, some options may not be available to the trader. The action of the trader can influence the state of the market. When the stock market closes, the trader is satisfied with his earnings and goes on an eternal vacation. The dynamics of the market can be described as:

$$\begin{aligned} x_{k+1} &= w_k, \\ x_k &\in \{0, 1, 2\}, \\ u_k &\in \mathcal{U}(x_k), \end{aligned}$$

where the sets of admissible control inputs associated with each state are:

$$\mathcal{U}(0) = \{I, S, N\}, \qquad \mathcal{U}(1) = \{I, N\}, \qquad \mathcal{U}(2) = \{S, N\},$$

and the transition probabilities $P_{ij}(\mathrm{u}) := \Pr(w_k = j | x_k = i, u_k = \mathrm{u})$ between the states are:

$$\begin{array}{lll} P_{00}(I) = 1.0 & & \\ P_{10}(I) = 0.25 & P_{11}(I) = 0.5 & P_{12}(I) = 0.25 \\ P_{00}(S) = 1.0 & & \\ P_{20}(S) = 0.1 & P_{21}(S) = 0.4 & P_{22}(S) = 0.5 \\ P_{00}(N) = 1.0 & & \\ P_{10}(N) = 0 & P_{11}(N) = 0.5 & P_{12}(N) = 0.5 \\ P_{20}(N) = 0 & P_{21}(N) = 0.5 & P_{22}(N) = 0.5 \end{array}$$

The cost associated with the state transition and action $g(x_k, u_k, w_k)$ is defined as follows:

$$\begin{array}{lll} g(0, I, 0) = 0 & & \\ g(1, I, 0) = -6 & g(1, I, 1) = 4 & g(1, I, 2) = 6 \\ g(0, S, 0) = 0 & & \\ g(2, S, 0) = -10 & g(2, S, 1) = 8 & g(2, S, 2) = 4 \\ g(0, N, 0) = 0 & & \\ g(1, N, 0) = 1 & g(1, N, 1) = 2 & g(1, N, 2) = 4 \\ g(2, N, 0) = 1 & g(2, N, 1) = 2 & g(2, N, 2) = 4 \end{array}$$

The trader wants to have as much money as possible once the market closes, so the cost function to minimize is the following:

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

i) What is the number of proper policies for this problem? *[2 points]*

ii) Remember the Gauss-Seidel update rule for Value Iteration:

For $i = 1$ to n:

$$V(i) \leftarrow \min_{\mathrm{u} \in \mathcal{U}(i)} \left(q(i, \mathrm{u}) + \sum_{j=1}^{n} P_{ij}(\mathrm{u}) V(j) \right).$$

Compute two iterations of Value Iteration using this update rule, which means, compute $V_2(1)$, $V_2(2)$ given the initial value $V_0(1) = V_0(2) = 0$. Evaluate first the Value Function for state 1, and then for state 2. State the associated best control action for each state. *[8 points]*

iii) In the worst case, what is the total number of iterations that Value Iteration can take before reaching the optimal cost for each state $J^*(i)$, given the problem formulation? *[1 point]*

iv) Is it possible that this problem stems from the reformulation of a Discounted Infinite Horizon problem into its auxiliary Stochastic Shortest Path problem? Provide a short explanation. *[2 points]*

b) Decide whether the following statements regarding Stochastic Shortest Path problems are **True** or **False**. You do not have to provide an explanation. *[6 points]*

i) In the terminal state, any admissible control action is optimal.

ii) When Policy Iteration terminates, it holds that the cost J_{μ^h} associated with policy μ^h in the final policy evaluation step is equal to the optimal cost J^*, i.e., $J_{\mu^h}(i) = J^*(i), \forall i \in \mathcal{S}^+$, where $\mathcal{S}^+$ is the state space excluding the terminal state.

iii) For any Stochastic Shortest Path problem, any admissible policy is also guaranteed to be proper.

iv) In Stochastic Shortest Path problems, the Value Iteration algorithm involves solving a system of linear equations.

v) In Discounted Infinite Horizon problems, Policy Iteration may be initialized with any admissible policy.

vi) The policy μ associated with the state transition graph depicted in Figure 1 is proper.

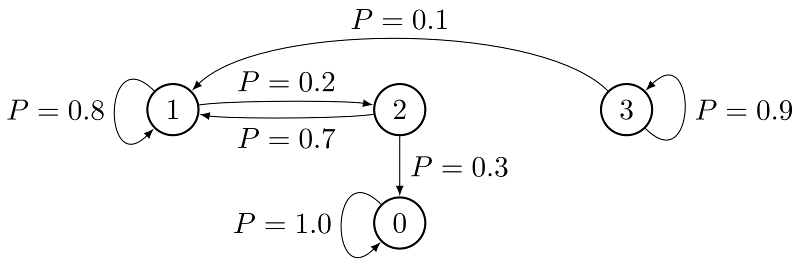

Figure 1: The state transition graph under policy μ for a given Stochastic Shortest Path problem. The probability P associated with a transition is denoted on each arc. 0 is a cost-free termination state.

c) Consider the Discounted Infinite Horizon problem described in Figure 2, with discount factor $\alpha = 0.4$. The cost function that we intend to minimize is the following:

$$\lim_{N\to\infty} \underset{(\tilde{X}_1, \tilde{W}_0 | \tilde{x}_0)}{E} \left[\sum_{k=0}^{N-1} \alpha^k \tilde{g}(\tilde{x}_k, \tilde{\mu}_k(\tilde{x}_k), \tilde{w}_k) \right].$$

i) Write down the Bellman equation for state 1. Substitute the numbers from the graph in Figure 2. You do not have to solve the Bellman equation. *[3 points]*

ii) Formulate the auxiliary Stochastic Shortest Path problem by writing down the transition probabilities and the costs for action $u = L$ in Figure 3. *[3 points]*

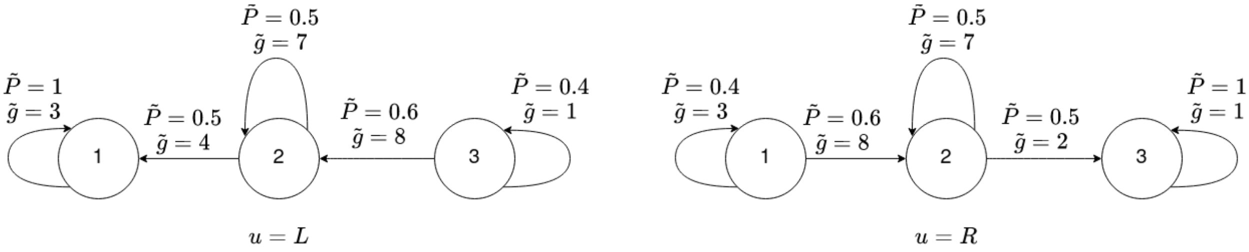

Figure 2: The state transition graphs for actions $u = L$ and $u = R$. The probability $\tilde{P}$ and cost $\tilde{g}$ associated with a transition is denoted on each arc.

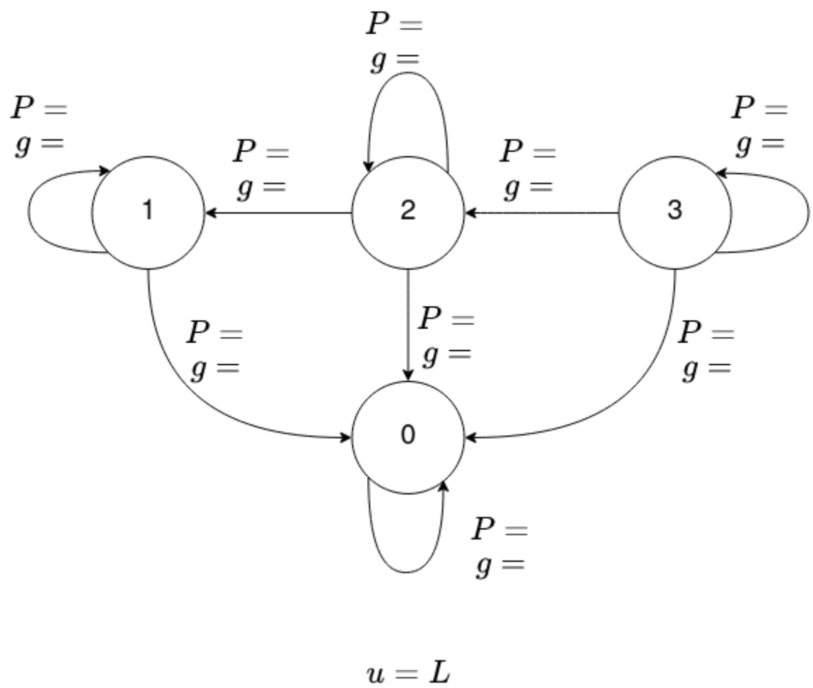

Figure 3: Auxiliary Stochastic Shortest Path problem.

Solution 2

a) i) There exist **9** proper policies for this problem, as follows:

$$\mu(0) = I, \mu(1) = I, \mu(2) = S$$
$$\mu(0) = I, \mu(1) = I, \mu(2) = N$$
$$\mu(0) = I, \mu(1) = N, \mu(2) = S$$

$$\mu(0) = S, \mu(1) = I, \mu(2) = S$$
$$\mu(0) = S, \mu(1) = I, \mu(2) = N$$
$$\mu(0) = S, \mu(1) = N, \mu(2) = S$$

$$\mu(0) = N, \mu(1) = I, \mu(2) = S$$
$$\mu(0) = N, \mu(1) = I, \mu(2) = N$$
$$\mu(0) = N, \mu(1) = N, \mu(2) = S$$

ii) Initialization:

$$V^0(1) = 0$$
$$V^0(2) = 0$$

Iteration 1:

$$\begin{aligned} V^1(1) = \min_{u\in\{I,N\}} [&-6 \cdot 0.25 + 4 \cdot 0.5 + 6 \cdot 0.25 + 0.5 \cdot V^0(1) + 0.25 \cdot V^0(2), \\ &2 \cdot 0.5 + 4 \cdot 0.5 + 0.5 \cdot V^0(1) + 0.25 \cdot V^0(2)] = \min_{u\in\{I,N\}} [2, 3] = 2 \\ V^1(2) = \min_{u\in\{S,N\}} [&-10 \cdot 0.1 + 8 \cdot 0.4 + 4 \cdot 0.5 + 0.4 \cdot V^1(1) + 0.5 \cdot V^0(2), \\ &2 \cdot 0.5 + 4 \cdot 0.5 + 0.5 \cdot V^1(1) + 0.5 \cdot V^0(2)] = \min_{u\in\{S,N\}} [5, 4] = 4 \end{aligned}$$

Iteration 2:

$$\begin{aligned} V^2(1) &= \min_{u\in\{I,N\}} [-6 \cdot 0.25 + 4 \cdot 0.5 + 6 \cdot 0.25 + 0.5 \cdot V^1(1) + 0.25 \cdot V^1(2), \\ &\qquad 2 \cdot 0.5 + 4 \cdot 0.5 + 0.5 \cdot V^1(1) + 0.5 \cdot V^1(2)] \\ &= \min_{u\in\{I,N\}} [2 + 0.5 \cdot 2 + 0.25 \cdot 4, 3 + 0.5 \cdot 2 + 0.5 \cdot 4] = 4 \\ V^2(2) &= \min_{u\in\{S,N\}} [-10 \cdot 0.1 + 8 \cdot 0.4 + 4 \cdot 0.5 + 0.4 \cdot V^2(1) + 0.5 \cdot V^1(2), \\ &\qquad 2 \cdot 0.5 + 4 \cdot 0.5 + 0.5 \cdot V^2(1) + 0.5 \cdot V^1(2)] \\ &= \min_{u\in\{S,N\}} [4.2 + 0.4 \cdot 4 + 0.5 \cdot 4, 3 + 0.5 \cdot 4 + 0.5 \cdot 4] = 7 \end{aligned}$$

Optimal policy at the end of the second iteration:

$$\mu^2(1) = I$$
$$\mu^2(2) = N$$

iii) Value Iteration generally requires an infinite number of iterations to converge.

iv) No, this problem is not derived from the reformulation of an equivalent Discounted Infinite Horizon problem; there is no discount factor α for which the probability of going from all the states to the terminal state is equal to $1-\alpha$ regardless of the action.

b) i) True: regardless of the action, we will stay at 0 indefinitely with 0 incurred cost.

ii) True

iii) False

iv) False

v) True

vi) True, there is a positive probability path from each state to the terminal state.

c) i) The Bellman equation for state 1 is the following:

$$\begin{aligned} J^*(x=1) &= \min\{q(1,L) + \alpha\sum_{j=1}^{3}\tilde{P}_{ij}(L)J^*(j), q(1,R) + \alpha\sum_{j=1}^{3}\tilde{P}_{ij}(R)J^*(j)\} \\ &= \min\left\{3 + 0.4J^*(1), 6 + 0.16J^*(1) + 0.24J^*(2)\right\}. \end{aligned}$$

ii) The solution is stated in Figure 4.

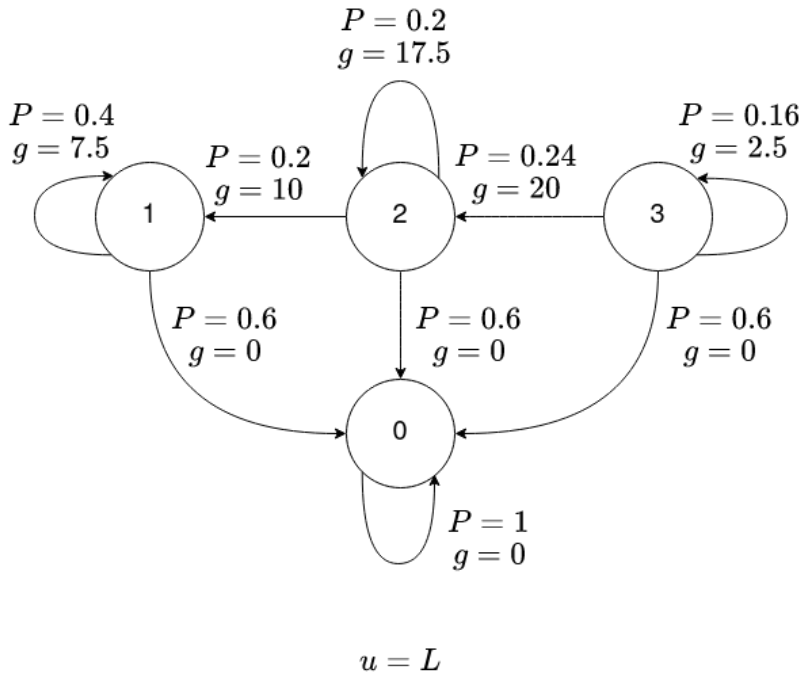

Figure 4: Auxiliary Stochastic Shortest Path problem.

Problem 3 [25 points]

a) You are the CEO of a company that currently owns one factory and would like to plan for new investments before your retirement. The following Deterministic Finite State System describes the evolution of your business:

- The state x of your company is captured by the number of factories you own. Therefore, $x_0 = 1$. You cannot have less than 0 factories.
- You can decide to either buy ($u_k = 0.5$) or sell ($u_k = -0.3$) one factory or to do nothing ($u_k = 0$). It takes 2 years to implement your decision. You can take the next action after that time period.
- Since you plan to retire in four years, you have already decided to sell all factories in year four, no matter how many factories you own at that point.
- The cost can be modeled as follows:

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

where

$$g_k = -0.2x_k + u_k,$$
$$g_N = -0.5x_N.$$

Convert the system stated above to a Shortest Path problem by drawing the corresponding graph including all edge weights. You do not have to solve the problem. *[6 points]*

b) Consider the Shortest Path problem depicted in Figure 5.

i) Find the shortest path from S to T by filling Table 1 using the Label Correcting Algorithm with the breadth-first search method. *[7 points]*

ii) How would the Label Correcting Algorithm behave if a new edge from E to D with a weight of -10 was introduced? *[2 points]*

iii) Would using the A*-algorithm with the heuristic $h_j = d_j$, where d_j is the current distance to node j, be a good choice to accelerate the convergence of the Label Correcting Algorithm? Provide a short explanation. *[2 points]*

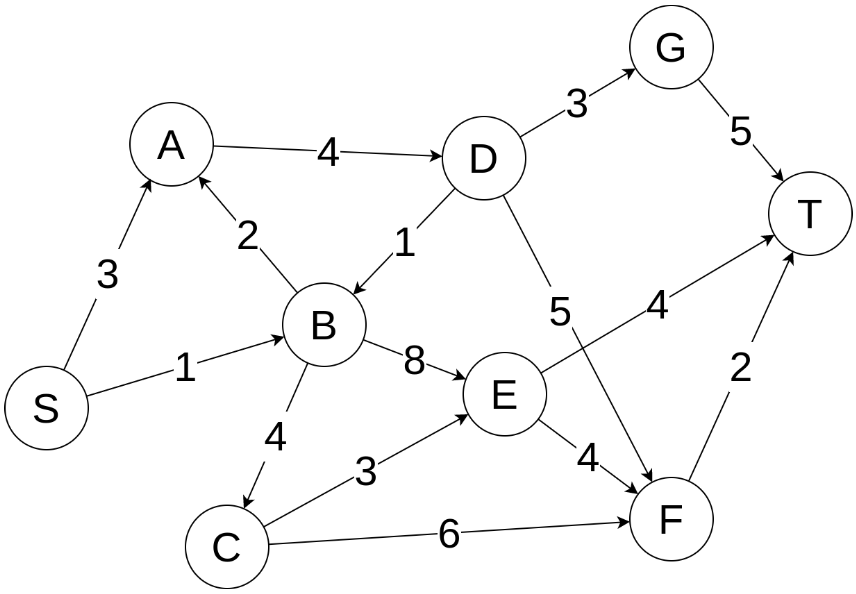

Figure 5

Iteration	Remove	OPEN	d_S	d_A	d_B	d_C	d_D	d_E	d_F	d_G	d_T
0	-	s	0	∞	∞	∞	∞	∞	∞	∞	∞
1	s										

Table 1: Label Correcting Algorithm Table for question b) (breadth-first search).

Optimal Path: ....

Instructions: Recall that only one instance of a node can be in OPEN at any time. If a node that is already in the OPEN bin enters the OPEN bin again, the node retains its original position in the OPEN bin. If multiple nodes enter the OPEN bin in the same iteration, add the nodes in alphabetical order. Example: OPEN bin: C, B, F; Node exiting OPEN: C (nodes entering OPEN: E, F, G); new OPEN bin: B, F, E, G; Node exiting OPEN: B.

c) Consider a system with state space $\mathcal{S} = \{A, B\}$ and transition probabilities given as:

$$\Pr(x_k = A | x_{k-1} = B) = 0.2,$$
$$\Pr(x_k = B | x_{k-1} = A) = 0.4.$$

Every time step k, you receive a measurement $z_k \in \{0, 1\}$. The likelihood of taking measurement z_k given x_{k-1} is captured by:

$$\Pr(z_k = 0 | x_{k-1} = A) = 0.3,$$
$$\Pr(z_k = 0 | x_{k-1} = B) = 0.7.$$

Further, z_k is conditionally independent with x_k, and prior variables $x_l, l < k-1$ and $z_l, l < k$, given x_{k-1}. The probability that $x_0 = A$ is 0.3 and the first two measurements are $z_1 = 1$ and $z_2 = 0$.

i) Construct the equivalent Shortest Path problem as per the Viterbi algorithm, by filling in the blanks in Figure 6 below, such that if solved, it yields the maximum a-posteriori estimate of the state sequence from $k = 0$ to $k = 2$. If $\ln(\cdot)$ is required, leave your answer in the form $\pm\ln(\#.\#\#)$ (e.g. $+\ln(1.35)$, $-\ln(1.00)$). *[6 points]*

ii) Assume that you will receive more measurements in the future. What algorithm would you apply to already solve for the most likely state trajectory up until now? Can future measurements influence the most likely state trajectory for past time steps? *[2 points]*

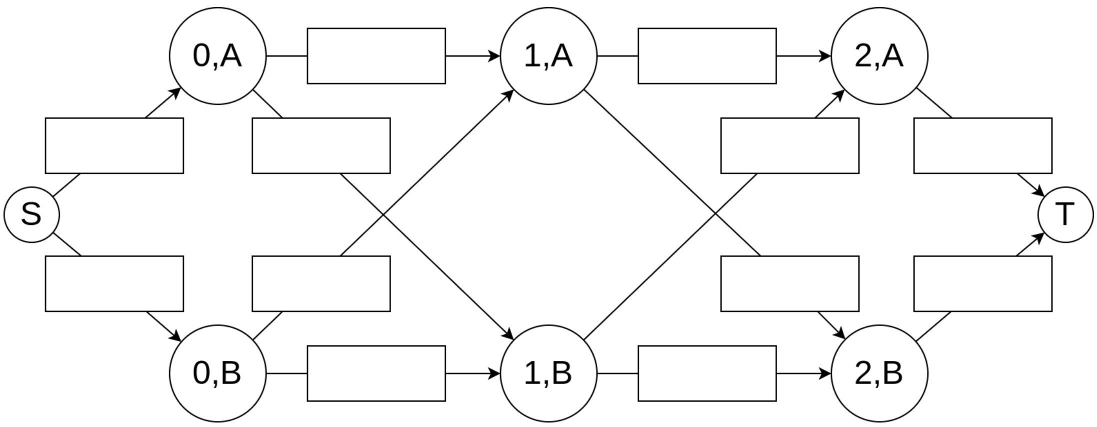

Figure 6

Solution 3

a) The graph corresponding to the Deterministic Finite State System is shown in Figure 7.

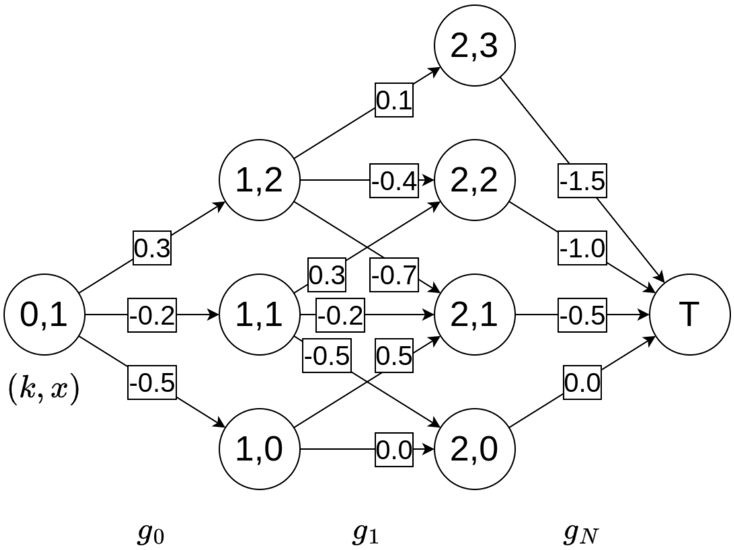

Figure 7

b) i) Table 2 shows all iterations obtained from applying the Label Correcting Algorithm.

Iteration	Remove	OPEN	d_S	d_A	d_B	d_C	d_D	d_E	d_F	d_G	d_T
0	-	s	0	∞	∞	∞	∞	∞	∞	∞	∞
1	s	A,B	0	3	1	∞	∞	∞	∞	∞	∞
2	A	B,D	0	3	1	∞	7	∞	∞	∞	∞
3	B	D,C,E	0	3	1	5	7	9	∞	∞	∞
4	D	C,E,F,G	0	3	1	5	7	9	12	10	∞
5	C	E,F,G	0	3	1	5	7	8	11	10	∞
6	E	F,G	0	3	1	5	7	8	11	10	12
7	F	G	0	3	1	5	7	8	11	10	12
8	G	-	0	3	1	5	7	8	11	10	12

Table 2: Label Correcting Algorithm Table for question b) (Breadth-first search)

Optimal Path: S-B-C-E-T

ii) The algorithm would not terminate as the graph includes a negative cycle.

iii) The heuristic $h_j = d_j$ is not a lower bound and the algorithm is not guaranteed to find the optimal solution. It is therefore not a good choice.

c) i) The equivalent Shortest Path problem is shown in Figure 8.

ii) The forward Dynamic Programming Algorithm can be used to find the most likely state trajectory. Future measurements can change the most likely trajectory for past time steps.

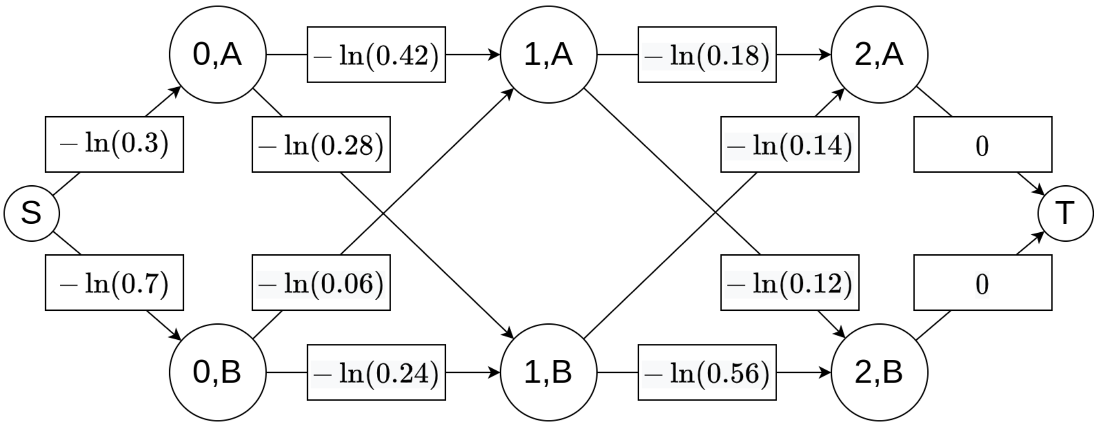
0,A
1,A
2,A
S
T
0,B
1,B
2,B
$-\ln(0.42)$
$-\ln(0.18)$
$-\ln(0.3)$
$-\ln(0.28)$
$-\ln(0.14)$
0
$-\ln(0.7)$
$-\ln(0.06)$
$-\ln(0.12)$
0
$-\ln(0.24)$
$-\ln(0.56)$

Figure 8

Problem 4 [25 points]

a) Consider a standard deterministic continuous time optimal control problem with dynamics

$$\dot{x}(t) = f(x(t), u(t)), \quad t \in [0, T],$$

and cost function

$$h(x(T)) + \int_0^T g(x(\tau), u(\tau)) d\tau,$$

where $x(t) \in \mathbb{R}$ and $u(t) \in \mathbb{R}$ for all $t \in [0, T]$. The cost function has the following form:

$$\begin{aligned} h(\mathrm{x}) &= 0, \quad \forall \mathrm{x} \in \mathbb{R}, \\ g(\mathrm{x}, \mathrm{u}) &= |\mathrm{x}|, \quad \forall \mathrm{x} \in \mathbb{R}, \forall \mathrm{u} \in \mathbb{R}. \end{aligned}$$

For an initial time t and state $\mathrm{x} \in \mathbb{R}$, the cost-to-go function associated with some control law $\mu(t, \mathrm{x}), \forall t \in [0, T], \mathrm{x} \in \mathbb{R}$ is

$$J_\mu(t, \mathrm{x}) = h(x(T)) + \int_t^T g(x(\tau), \mu(\tau, \mathrm{x})) d\tau,$$

subject to the dynamics and initial condition.

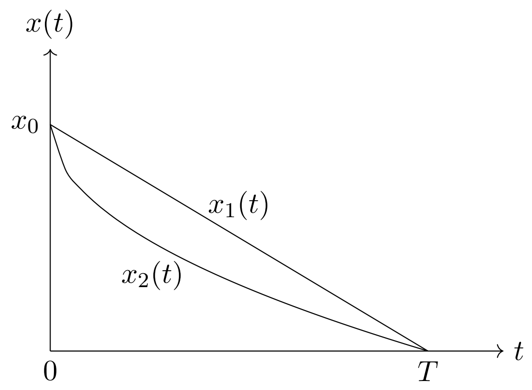

Figure 9: State trajectories $x_1(t)$ and $x_2(t)$.

Consider two admissible feedback laws $\mu_1(\cdot, \cdot)$ and $\mu_2(\cdot, \cdot)$, which result in trajectories $x_1(t)$ and $x_2(t)$ respectively, with the same initial condition $x(0) = x_0$. The state trajectories are shown in Figure 9.

Decide whether the following statements are **True** or **False** and provide short explanations.
[6 points]

i) It is possible that trajectory $x_1(t)$ is optimal.

ii) If the feedback law $\mu_2(\cdot, \cdot)$ is found using the Hamilton-Jacobi-Bellman equation, then it is optimal.

iii) The following equation regarding $J_{\mu_1}(t, x_1(t))$ holds.

$$0 = |x_1(t)| + \frac{\partial}{\partial t} J_{\mu_1}(t, x_1(t)) + \frac{\partial}{\partial x_1} J_{\mu_1}(t, x_1(t)) f(x_1(t), \mu_1(t, x_1(t))), \quad \forall t \in [0, T],$$

where $J_{\mu_1}(t, x_1(t))$ is the evaluation of the cost-to-go function $J_{\mu_1}(t, \mathrm{x}_1)$ at the state trajectory $x_1(t)$ at time t.

b) Consider a deterministic continuous time optimal control problem with dynamics

$$\dot{x}(t) = f(x(t), u(t)), \quad t \in [0, T],$$

free terminal time T, and cost function

$$h(x(T)) + \int_0^T g(x(\tau), u(\tau))d\tau,$$

where $x(t) \in \mathbb{R}$ and $u(t) \in \mathbb{R}$ for all $t \in [0, T]$. The cost function has the following form:

$$h(\mathrm{x}) = 0, \quad \forall \mathrm{x} \in \mathbb{R},$$
$$g(\mathrm{x}, \mathrm{u}) = |\mathrm{x}|+|\mathrm{u}|, \quad \forall \mathrm{x} \in \mathbb{R}, \forall \mathrm{u} \in \mathbb{R}.$$

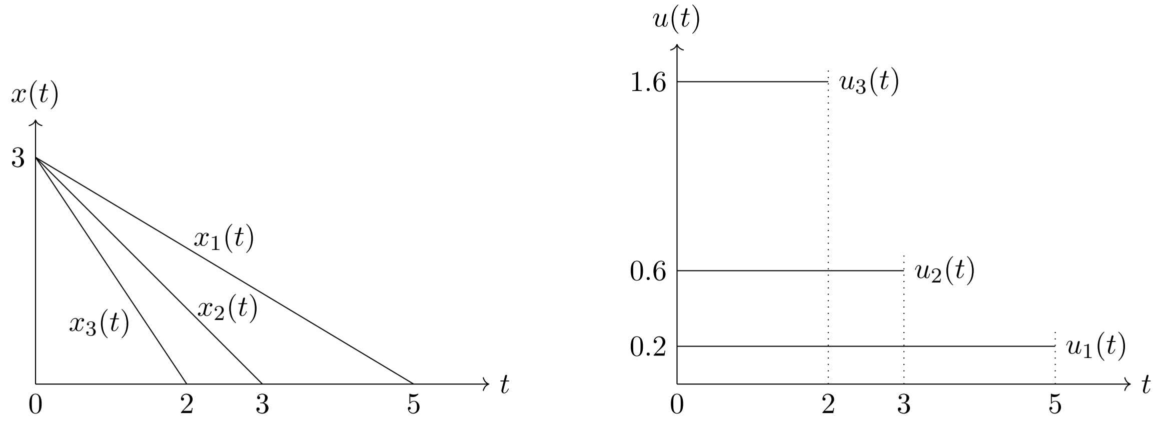

Figure 10: State trajectories (left) and input trajectories (right).

By applying Pontryagin's Minimum Principle, you get exactly three input sequences $u_1(\cdot)$, $u_2(\cdot)$ and $u_3(\cdot)$, which result in trajectories $x_1(t)$, $x_2(t)$ and $x_3(t)$ respectively, with the same fixed initial condition $x(0) = 3$ and fixed terminal state $x(T) = 0$. The state and input trajectories are shown in Figure 10.

i) Calculate the cost for each of the three trajectories. *[3 points]*

ii) Is the trajectory with the lowest cost optimal? Provide a short explanation. *[2 points]*

c) Consider the system dynamics

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

where $z(t)$ is the position of a cart on a one-dimensional track and $u(t)$ is the control input at time t, with $|u(t)| \leq \phi, \phi > 0$. The goal of this problem is to drive the cart to track a given reference trajectory $r(t)$ using Pontryagin's Minimum Principle. The cart is at $z(0) = 0$ with velocity $\dot{z}(0) = 0$ at time 0, and it is required that the cart is able to follow the trajectory exactly at the terminal time T, which means $z(T) = r(T)$ and $\dot{z}(T) = \dot{r}(T)$.

The cost function of the tracking problem is given as

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

We consider $T = 1$ and the reference trajectory $r(t) = t, t \geq 0$.

i) Write down the dynamics of the system and the boundary conditions using the state $x(t) = [z(t), \dot{z}(t)]^T$. *[2 points]*

ii) Write down the Hamiltonian $H(\mathrm{x}, \mathrm{u}, \mathrm{p}, t)$ and the co-state derivative $\dot{p}(t)$ of the problem. *[3 points]*

iii) In order to track the reference trajectory perfectly for a nontrivial amount of time before the terminal time T, $z(t) = r(t), \dot{z}(t) = \dot{r}(t), \forall t \in [t_e, T], t_e < T$, what condition does the control input limit ϕ have to satisfy? *[9 points]*

Solution 4

a) i) False, because trajectory $x_2(t)$ has a lower cost.

ii) True, because the HJB equation is a sufficient condition for optimality.

iii) True, for a given policy $\mu(t,\mathrm{x})$ and its associated state trajectory $x(t)$, the total derivative of the cost-to-go evaluated at this state trajectory is

$$\begin{aligned}\frac{d}{dt}J_\mu(t,x(t)) &= \frac{d}{dt}\left(h(x(T)) + \int_t^T g(x(\tau),\mu(\tau,x(\tau)))d\tau\right)\\ &= -g(x(t),\mu(t,x(t)))\\ &= \frac{\partial}{\partial t}J_\mu(t,x(t)) + \frac{\partial}{\partial x}J_\mu(t,x(t))\cdot f(x(t),u(t)).\end{aligned}$$

This shows that

$$g(x(t),\mu(t,x(t))) + \frac{\partial}{\partial t}J_\mu(t,x(t)) + \frac{\partial}{\partial x}J_\mu(t,x(t))\cdot f(x(t),u(t)) = 0, \quad \forall t\in[0,T]$$

Note: This is not a contradiction to the HJB equation, because when solving the HJB equation, we do not assume that a control input $u(t)$ already exists.

b) i) $J_1 = 8.5, J_2 = 6.3, J_3 = 6.2$

ii) Yes. The three trajectories are from the Pontryagin Minimum Principle, which means that the optimal trajectory must be one of the three trajectories, since the Pontryagin Minimum Principle is a necessary condition of optimality. Trajectory $x_3(t)$ has the lowest cost, thus it is optimal.

c) i)

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

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

The boundary conditions are

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

ii) The Hamiltonian is

$$H(\mathrm{x},\mathrm{u},\mathrm{p},t) = \frac{1}{2}(\mathrm{x}_1 - t)^2 + \mathrm{p}_1\mathrm{x}_2 + \mathrm{p}_2\mathrm{u}$$

The derivative of the co-state is

$$\begin{aligned}\dot p_1(t) &= t - x_1(t)\\ \dot p_2(t) &= -p_1(t)\end{aligned}$$

iii) First, let us apply the minimum principle

$$\begin{aligned} u^*(t) &= \arg\min_{\mathrm{u}\in[-\phi,\phi]} H(x(t),\mathrm{u},p(t),t) \\ &= \arg\min_{\mathrm{u}\in[-\phi,\phi]} \frac{1}{2}(x_1(t)-t)^2 + p_1(t)x_2(t) + p_2(t)\mathrm{u} \end{aligned}$$

Notice that the Hamiltonian is linear in u, so we have

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

In the third case, we have $p_2(t) = 0$ for a nontrivial interval of time

$$\begin{aligned} p_2(t) &= 0 \\ \dot{p}_2(t) &= -p_1(t) = 0 \\ \dot{p}_1(t) &= t - x_1(t) = 0 \\ x_1(t) &= t \\ x_2(t) &= \dot{x}_1(t) = 1 \\ u(t) &= \dot{x}_2(t) = 0 \end{aligned}$$

The optimal control input is bang-bang control:

$$u^*(t) = \begin{cases} \phi & p_2(t) < 0 \\ -\phi & p_2(t) > 0 \\ 0 & p_2(t) = 0 \end{cases}$$

The singular arc corresponds to the situation when the cart is tracking the reference trajectory perfectly, and we want a nontrivial interval of singular arc before the terminal time T. The intuition is that we apply $u(t) = \phi$ for $0 \le t \le t_s$, and switch to $u(t) = -\phi$ for $t_s \le t \le t_e$. At $t = t_e$, our cart should be at the same position as the reference with the same velocity. After that, we set $u(t) = 0$ until the terminal time T.

1) $0 \le t \le t_s$:

$$x_2(t) = \phi t + C$$

$C = 0$ because of the boundary condition $x_2(0) = 0$.

$$x_1(t) = \frac{1}{2}\phi t^2 + D$$

$D = 0$ because of the boundary condition $x_1(0) = 0$. So we have

$$\begin{aligned} x_1(t) &= \frac{1}{2}\phi t^2 \\ x_2(t) &= \phi t, \quad 0 \le t \le t_s \end{aligned}$$

At time $t = t_s$ we have

$$\begin{aligned} x_1(t_s) &= \frac{1}{2}\phi t_s^2 \\ x_2(t_s) &= \phi t_s \end{aligned}$$

2) $t_s \leq t \leq t_e$:

$$x_2(t) = -\phi t + E$$

$E = 2\phi t_s$ because of the boundary condition $x_2(t_s) = \phi t_s$.

$$x_1(t) = -\frac{1}{2}\phi t^2 + 2\phi t_s t + F$$

$F = -\phi t_s^2$ because of the boundary condition $x_1(t_s) = \frac{1}{2}\phi t_s^2$. Thus we have

$$\begin{aligned} x_1(t) &= -\frac{1}{2}\phi t^2 + 2\phi t_s t - \phi t_s^2 \\ x_2(t) &= -\phi t + 2\phi t_s, \quad t_s \leq t \leq t_e \end{aligned}$$

Since at time $t = t_e$ the singular arc starts, we have

$$\begin{aligned} x_1(t_e) &= -\frac{1}{2}\phi t_e^2 + 2\phi t_s t_e - \phi t_s^2 = t_e \\ x_2(t_e) &= -\phi t_e + 2\phi t_s = 1 \end{aligned}$$

By combining the two equations, we have

$$(\phi t_e)^2 - 2(\phi t_e) - 1 = 0$$

Since $\phi t_e > 0$, we have

$$\phi t_e = 1 + \sqrt{2}$$

In order to have a nontrivial interval of singular arc, we require $t_e < T = 1$, so we can get

$$\phi > 1 + \sqrt{2}$$