Review for Planning & Learning in Robotics

Xu Shang

UC San Diego

Key problem

Optimal control

Horizon and randomness:
Infinite/finite horizon, deterministic/stochastic

- Finite-Horizon Stochastic/Deterministic Optimal Control
- Finite-Horizon Deterministic Optimal Control $\Leftrightarrow$ Shortest path problem
- Infinite-Horizon Stochastic/Deterministic Optimal Control

Important technique: Markov chain and Markov decision process

Section

Markov Chain

- **Stochastic process:** collection of random variables $\{x_0, x_1, \ldots\}$
- **Markov chain:** memoryless stochastic process $\{x_0, x_1, \ldots\}$:
  - x_0 has probability density function $p_0(\cdot)$
  - x_{t+1} conditioned on x_t has probability density function $p_f(\cdot \mid x_t)$ and is independent of the history $x_{0:t-1}$
- **Markov assumption:**
  "The future is independent of the past given the present"
- When the state space is finite, $\mathcal{X} := \{1, \ldots, N\}$, the pdf p_f can be represented by an $N \times N$ transition matrix with elements:

$$P_{ij} := \mathbb{P}(x_{t+1} = j \mid x_t = i) = p_f(j \mid x_t = i).$$

- Properties: transient, recurrent, etc.
- Computations related to the transition matrix and the stationary distribution, *e.g.*, mean first passage time.

Markov Decision Process

Markov Reward Process with controlled transitions defined by a tuple

$$(\mathcal{X}, \mathcal{U}, p_0, p_f, T, \ell, q, \gamma)$$

- $\mathcal{X}$ is a discrete or continuous state space
- $\mathcal{U}$ is a discrete or continuous control space
- $p_0(\cdot)$ is a prior pdf defined on $\mathcal{X}$
- $p_f(\cdot \mid x_t, u_t)$ is a conditional pdf defined on $\mathcal{X}$ for given $x_t \in \mathcal{X}$ and $u_t \in \mathcal{U}$
  (matrices P^u with elements $P^u_{ij} := p_f(j \mid x_t = i, u_t = u)$ in finite-dim case)
- T is a finite or infinite time horizon
- $\ell(x, u)$ is stage cost of applying control $u \in \mathcal{U}$ in state $x \in \mathcal{X}$
- $q(x)$ is terminal cost of being in state x at time T
- $\gamma \in [0, 1]$ is a discount factor

MDP Control Policy and Value Function

- **Control policy:** a function π that maps a time step $t \in \mathbb{N}$ and a state $x \in \mathcal{X}$ to a feasible control input $u \in \mathcal{U}$
- **Value function:** expected cumulative cost of a policy π applied to an MDP with initial state $x \in \mathcal{X}$ at time t:
- **Finite-horizon MDP:** trajectories terminate at fixed $T < \infty$:

$$V_t^\pi(x) := \mathbb{E}\left[q(x_T) + \sum_{\tau=t}^{T-1} \ell(x_\tau, \pi_\tau(x_\tau)) \,\middle|\, x_t = x \right]$$

- **Infinite-horizon MDP:** as $T \to \infty$, optimal policies become stationary, i.e., $\pi := \pi_0 \equiv \pi_1 \equiv \cdots$
  - **First-exit MDP:** trajectories terminate at the first passage time $T = \min\{t \in \mathbb{N} \mid x_t \in \mathcal{T}\}$ to a terminal state $x_t \in \mathcal{T} \subseteq \mathcal{X}$
  - **Discounted MDP:** trajectories continue forever but stage costs are discounted by a factor $\gamma \in [0, 1)$
  - **Average-cost MDP:** trajectories continue forever and the value function is the expected average stage cost

Section

MDP Finite-horizon Optimal Control

The finite-horizon optimal control problem in an MDP $(\mathcal{X}, \mathcal{U}, p_0, p_f, T, \ell, q, \gamma)$ with initial state $\mathbf{x}$ at time t is:

$$\min_{\pi_{t:T-1}} \quad V_t^{\pi}(\mathbf{x}) := \mathbb{E}_{x_{t+1:T}} \left[\gamma^{T-t} q(\mathbf{x}_T) + \sum_{\tau=t}^{T-1} \gamma^{\tau-t} \ell(\mathbf{x}_\tau, \pi_\tau(\mathbf{x}_\tau)) \,\middle|\, \mathbf{x}_t = \mathbf{x} \right]$$

$$\begin{aligned} \text{s.t.} \quad & \mathbf{x}_{\tau+1} \sim p_f(\cdot \mid \mathbf{x}_\tau, \pi_\tau(\mathbf{x}_\tau)), \qquad \tau = t, \ldots, T-1, \\ & \mathbf{x}_\tau \in \mathcal{X}, \qquad \pi_\tau(\mathbf{x}_\tau) \in \mathcal{U}. \end{aligned}$$

- Deterministic case is a special case for the stochastic problem.
- For the deterministic case, the closed-loop and open-loop control have the same performance.

The Dynamic Programming Algorithm

Algorithm Dynamic Programming

Input: MDP $(\mathcal{X}, \mathcal{U}, p_0, p_f, T, \ell, q, \gamma)$

$V_T(x) = q(x), \qquad \forall x \in \mathcal{X}$

for $t = (T-1) \dots 0$ **do**

$Q_t(x,u) = \ell(x,u) + \gamma \mathbb{E}_{x' \sim p_f(\cdot|x,u)} \left[V_{t+1}(x') \right], \qquad \forall x \in \mathcal{X},\ u \in \mathcal{U}(x)$

$V_t(x) = \min_{u \in \mathcal{U}(x)} Q_t(x,u), \qquad \forall x \in \mathcal{X}$

$\pi_t(x) = \arg \min_{u \in \mathcal{U}(x)} Q_t(x,u), \qquad \forall x \in \mathcal{X}$

end for

return policy $\pi_{0:T-1}$ and value function V_0

- The expected value function at $x' \sim p_f(\cdot \mid x, u)$ is:
  - Discrete $\mathcal{X}$:

$$\mathbb{E}_{x' \sim p_f(\cdot|x,u)} \left[V_{t+1}(x') \right] = \sum_{x' \in \mathcal{X}} V_{t+1}(x') p_f(x' \mid x, u)$$

Dynamic Programming Optimality

Theorem

The policy $\pi_{0:T-1}$ and value function V_0 returned by the Dynamic Programming algorithm are optimal for the finite-horizon optimal control problem.

- **Proof:**
  - Let $V_t^*(x)$ be the optimal cost for the problem with planning horizon $(T-t)$ that starts at time t in state x
  - Proceed by induction
  - **Base-case:** $V_T^*(x) = q(x) = V_T(x)$
  - **Hypothesis:** Assume that for $t+1$, $V_{t+1}^*(x) = V_{t+1}(x)$ for all $x \in \mathcal{X}$
  - **Induction:** Show that $V_t^*(x) = V_t(x)$ for all $x \in \mathcal{X}$

Section

Deterministic Shortest Path (DSP) Problem

- **Path:** a sequence $i_{1:q} := (i_1, i_2, \ldots, i_q)$ of nodes $i_k \in \mathcal{V}$
- **Path length:** sum of edge weights along the path

$$J^{i_{1:q}} = \sum_{k=1}^{q-1} c_{i_k, i_{k+1}}$$

- **All paths from $s \in \mathcal{V}$ to $\tau \in \mathcal{V}$:**

$$\mathcal{P}_{s,\tau} := \{ i_{1:q} \mid i_k \in \mathcal{V},\ i_1 = s,\ i_q = \tau \}$$

- **Objective:** find a path that has the min length from node s to node τ:

$$\textbf{dist}(s, \tau) = \min_{i_{1:q} \in \mathcal{P}_{s,\tau}} J^{i_{1:q}} \qquad i^*_{1:q} \in \arg \min_{i_{1:q} \in \mathcal{P}_{s,\tau}} J^{i_{1:q}}$$

- **Assumption:** There are no negative cycles in the graph, i.e., $J^{i_{1:q}} \geq 0$, for all $i_{1:q} \in \mathcal{P}_{i,i}$ and all $i \in \mathcal{V}$
- The finite-state DSP problem is equivalent to a finite-horizon finite-state deterministic optimal control (DOC) problem
- Apply dynamic programming or label correcting (variant of a "forward" DPA) to the equivalent DOC problem

Label Correcting Algorithm

Algorithm Label Correcting Algorithm

OPEN $\leftarrow \{s\}$, $g_s = 0$, $g_i = \infty$ for all $i \in \mathcal{V} \setminus \{s\}$
while OPEN is not empty **do**
    Remove i from OPEN
    for $j \in \text{Children}(i)$ **do**
        if $(g_i + c_{ij}) < g_j$ **and** $(g_i + c_{ij}) < g_\tau$ **then**
            $g_j = g_i + c_{ij}$
            $\text{Parent}(j) = i$
            if $j \neq \tau$ **then**
                $\text{OPEN} = \text{OPEN} \cup \{j\}$
            end if
        end if
    end for
end while

If you have time: Check the proof for label correcting algorithm can find the shortest path if the problem is feasible.

A^* Algorithm with an ϵ-consistent Heuristic

Algorithm Weighted A* Algorithm

OPEN $\leftarrow \{s\}$, CLOSED $\leftarrow \{\}$, $\epsilon \geq 1$
$g_s = 0$, $g_i = \infty$ for all $i \in \mathcal{V} \setminus \{s\}$
while $\tau \notin$ CLOSED **do**
    Remove i with smallest $f_i := g_i + \epsilon h_i$ from OPEN
    Insert i into CLOSED
    for $j \in \text{Children}(i)$ and $j \notin$ CLOSED **do**
        if $g_j > (g_i + c_{ij})$ **then**
            $g_j \leftarrow (g_i + c_{ij})$
            $\text{Parent}(j) \leftarrow i$
            if $j \in$ OPEN **then**
                Update priority of j
            else
                OPEN $\leftarrow$ OPEN $\cup \{j\}$

Optimality, efficiency, requirement of the heuristic function

Rapidly Exploring Random Tree (RRT)

- Starting with an initial configuration x_s, build a tree until the goal configuration x_τ is part of it

Algorithm RRT

$V \leftarrow \{x_s\}$; $E \leftarrow \emptyset$
for $i = 1 \ldots n$ **do**
    $x_{rand} \leftarrow \text{SampleFree}()$
    $x_{nearest} \leftarrow \text{Nearest}((V, E), x_{rand})$
    $x_{new} \leftarrow \text{Steer}(x_{nearest}, x_{rand})$
    if $\text{CollisionFree}(x_{nearest}, x_{new})$ **then**
        $V \leftarrow V \cup \{x_{new}\}$; $E \leftarrow E \cup \{(x_{nearest}, x_{new})\}$
    end if
end for
return $G = (V, E)$ =0

Comment

The deterministic shortest path is more about the algorithm; remember to also write down the variants of A^*, such as $LRTA^*, RTAA^*$, etc.

Section

Infinite-Horizon Stochastic Optimal Control

- In this lecture, we consider what happens with the stochastic optimal control problem as the planning horizon T goes to infinity
- We will consider two formulations of the infinite-horizon stochastic optimal control problem
  - **Discounted Problem:** obtained by letting $T \to \infty$ in the finite-horizon stochastic optimal control problem with $\gamma < 1$
  - **First-Exit Problem:** obtained by considering stochastic transitions in the shortest path problem and terminating when the goal region is reached
- Just like the DOC and DSP problems, the Discounted Problem and the First-Exit Problem are equivalent, i.e., one can be converted into the other

Bellman Equations Summary

- **Value Function:**

$$V^{\pi}(x) = \ell(x, \pi(x)) + \gamma \mathbb{E}_{x' \sim p_f(\cdot|x,\pi(x))} \left[V^{\pi}(x') \right], \qquad \forall x \in \mathcal{X}$$

- **Optimal Value Function:**

$$V^{*}(x) = \min_{u \in \mathcal{U}} \left\{ \ell(x, u) + \gamma \mathbb{E}_{x' \sim p_f(\cdot|x,u)} \left[V^{*}(x') \right] \right\}, \qquad \forall x \in \mathcal{X}$$

- **Q Function:**

$$Q^{\pi}(x, u) = \ell(x, u) + \gamma \mathbb{E}_{x' \sim p_f(\cdot|x,u)} \left[Q^{\pi}(x', \pi(x')) \right], \qquad \forall x \in \mathcal{X},\ u \in \mathcal{U}$$

- **Optimal Q Function:**

$$Q^{*}(x, u) = \ell(x, u) + \gamma \mathbb{E}_{x' \sim p_f(\cdot|x,u)} \left[\min_{u' \in \mathcal{U}} Q^{*}(x', u') \right], \qquad \forall x \in \mathcal{X},\ u \in \mathcal{U}$$

Important procedures

- Policy Evaluation
- Value Iteration
- Policy Iteration

Remark:

- When the state is finite, we have special approaches to solve these problems. (L10: P26, P54)
- Model-free case $\Rightarrow$ Monte-Carlo

Incremental vs Batch optimization

- **Incremental optimization:**
  - **Gradient descent:**

$$\delta\theta = -\nabla_\theta J(\theta) = \mathbb{E}\left[\left(V^\pi(x) - \hat{V}(x;\theta)\right)\nabla_\theta \hat{V}(x;\theta)\right]$$

$$\delta\theta = -\nabla_\theta J(\theta) = \mathbb{E}\left[\left(Q^\pi(x,u) - \hat{Q}(x,u;\theta)\right)\nabla_\theta \hat{Q}(x,u;\theta)\right]$$

  - **Stochastic gradient descent:** uses samples x_t, u_t from π rather than computing the exact expectation:

$$\delta\theta_t = \left(V^\pi(x_t) - \hat{V}(x_t;\theta)\right)\nabla_\theta \hat{V}(x_t;\theta)$$

$$\delta\theta_t = \left(Q^\pi(x_t,u_t) - \hat{Q}(x_t,u_t;\theta)\right)\nabla_\theta \hat{Q}(x_t,u_t;\theta)$$

- **Batch optimization:** the expected update $\mathbb{E}[\delta\theta_t]$ must be zero at the minimizer θ^* of $J(\theta)$. Determine θ^* directly by solving:

$$\mathbb{E}[\delta\theta_t] = 0$$

Remark: Take care of the case when $\hat{V}(x;\theta)$ is in the form of $\theta^\mathsf{T}\phi(x)$.

Deterministic Optimal Control

- Deterministic optimal control with initial state x_0:

$$\begin{aligned} \min_{u_{0:T-1}} \quad & V_0^{u_{0:T-1}}(x_0) = q(x_T) + \sum_{t=0}^{T-1} \ell(x_t, u_t) \\ \text{s.t.} \quad & x_{t+1} = f(x_t, u_t), \quad t = 0, \ldots, T-1. \end{aligned}$$

- The problem has a closed-form solution when the costs are **quadratic** and the dynamics are **linear**:

$$\begin{aligned} q(x) &= \frac{1}{2} x^\top Q_0 x + a^\top x + a, \qquad Q_0 \succeq 0,\ Q \succeq 0,\ R \succ 0, \\ \ell(x, u) &= \frac{1}{2} x^\top Q x + \frac{1}{2} u^\top R u + x^\top P u + q^\top x + r^\top u + q, \\ f(x, u) &= Ax + Bu + c. \end{aligned}$$

Deterministic Optimal Control

- **Cost and dynamics:**

$$q(x) = \frac{1}{2}x^\top Q x + a^\top x + a, \qquad Q \succeq 0,\ Q_t \succeq 0,\ R_t \succ 0,$$

$$\ell_t(x,u) = \frac{1}{2}x^\top Q_t x + \frac{1}{2}u^\top R_t u + x^\top P_t u + q_t^\top x + r_t^\top u + q_t,$$

$$f_t(x,u) = A_t x + B_t u + c_t.$$

- **Optimal value:**

$$V_t^*(x) = \frac{1}{2}x^\top M_t x + \mathbf{m}_t^\top x + m_t$$

- **Optimal policy:**

$$\pi_t^*(x) = -H_{uu,t}^{-1}\left(H_{xu,t}^\top x + \mathbf{h}_{u,t}\right)$$

Continued

Riccati equations:

$$M_T = Q, \qquad \mathbf{m}_T = \mathbf{a}, \qquad m_T = a$$

$$M_t = H_{xx,t} - H_{xu,t} H_{uu,t}^{-1} H_{xu,t}^\top$$

$$\mathbf{m}_t = A_t^\top (\mathbf{m}_{t+1} + M_{t+1} c_t) + \mathbf{q}_t - H_{xu,t} H_{uu,t}^{-1} \mathbf{h}_{u,t}$$

$$m_t = -\frac{1}{2} \mathbf{h}_{u,t}^\top H_{uu,t}^{-1} \mathbf{h}_{u,t} + \frac{1}{2} c_t^\top M_{t+1} c_t + \mathbf{m}_{t+1}^\top c_t + m_{t+1} + q_t$$

$$\mathbf{h}_{u,t} = B_t^\top (\mathbf{m}_{t+1} + M_{t+1} c_t) + \mathbf{r}_t$$

$$H_{xx,t} = Q_t + A_t^\top M_{t+1} A_t$$

$$H_{uu,t} = R_t + B_t^\top M_{t+1} B_t$$

$$H_{xu,t} = P_t + A_t^\top M_{t+1} B_t$$

Do not ignore the discount factor!

Section

Problem 1

Consider an experiment described by a two-state Markov chain with transition matrix:

$$P = \begin{bmatrix} 0.5 & 0.5 \\ p & 1-p \end{bmatrix}, \tag{1}$$

where $p \in (0, 1)$ is an unknown parameter. When the experiment is performed many times, the chain ends in state one approximately 20% of the time and in state two approximately 80% of the time. Provide an estimate for p and explain how it was determined.

Solution

Solution. The stationary distribution is $\mathbf{w} = [1/5,\ 4/5]^\top$ and satisfies $\mathbf{w}^\top P = \mathbf{w}^\top$. Hence:

$$\mathbf{w}^\top = \begin{bmatrix} \frac{1}{5} & \frac{4}{5} \end{bmatrix} = \mathbf{w}^\top P = \begin{bmatrix} \frac{1}{10}(1+8p) & \frac{1}{10}(9-8p) \end{bmatrix} \quad \Longrightarrow \quad p = \frac{1}{8}. \quad (2)$$

Problem 2

Consider a vessel with 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 $v_i \in \mathbb{R}_{>0}$ be the known value of item i, let $w_i \in \mathbb{R}_{>0}$ be the known weight of item i, and let $n_i \in \{0, 1\}$ indicate whether item i is loaded or not. 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 constraint

$$\sum_{i=1}^{K} n_i w_i \leq W.$$

Formulate this problem as a finite-horizon optimal control problem, where the cumulative cost expresses the negative total value of our cargo. Define the time horizon, state space, control space, motion model, stage cost, and terminal cost.

Solution

Solution. The time horizon is equal to the number of items $T = K$. The state x_t denotes the total weight of the cargo for items $1, \ldots, t$, with $x_0 = 0$. The state space is $\mathcal{X} = \mathbb{R}_{>0}$. The control input $u_k \in \{0, 1\}$ indicates whether item k is loaded or not. The control space is $\mathcal{U} = \{0, 1\}$. The motion model is:

$$x_{t+1} = x_t + u_t w_{t+1}, \qquad t = 0, \ldots, T-1. \tag{3}$$

The stage cost is

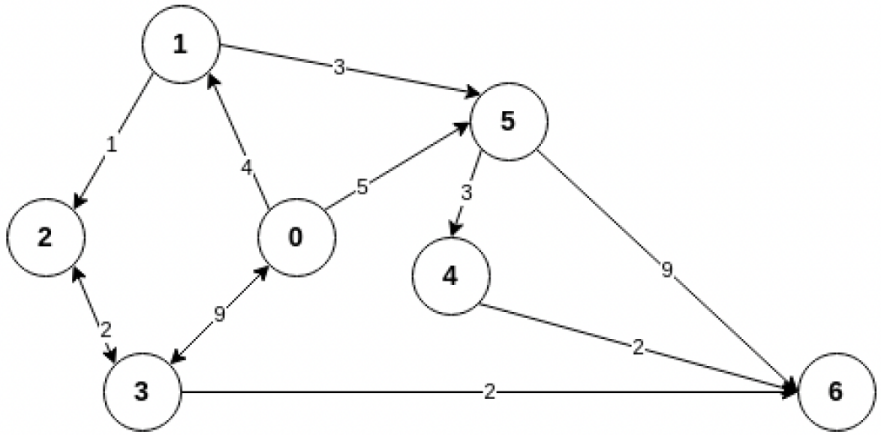
$$\ell_t(x, u) = -u v_{t+1}.$$

The terminal cost is:

$$q(x) = \begin{cases} \infty, & \text{if } x > W, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

Problem 3

Consider a deterministic shortest path problem for the graph in Fig. 1 with initial node 0 and goal node 6.



(a) Perform 3 iterations of the weighted A* algorithm with heuristic function provided in the Table and $\epsilon = 1$. Do *not* reopen nodes already in the CLOSED list. Report the g values computed by A*.

	h_0	h_1	h_2	h_3	h_4	h_5	h_6
h	0	4	3	2	1	5	0

Solution

Iteration	Node exiting OPEN	OPEN	g_0	g_1	g_2	g_3	g_4	g_5	g_6
0	–	$\{0\}$	0	∞	∞	∞	∞	∞	∞
1	0	$\{1,3,5\}$	0	4	∞	9	∞	5	∞
2	1	$\{2,3,5\}$	0	4	5	9	∞	5	∞
3	2	$\{3,5\}$	0	4	5	7	∞	5	∞
4	3	$\{6,5\}$	0	4	5	7	∞	5	9

Problem 3

(b) Without performing any iterations, guess if the result in Table 2 may change if we allow reopening nodes in the CLOSED list. Justify your answer.

Solution. The result may change because the heuristic function is inconsistent:

$$5 = h_5 > c_{5,4} + h_4 = 3 + 1 = 4. \tag{5}$$

Problem 3

(c) Consider the graph in Fig. 1 and use the final g values obtained in part (a). Suppose that a new node 7 is added to the graph. Use the RRT* algorithm to extend and rewire the graph. Specifically, consider all nodes $\mathcal{N} = \{0, 1, 2, 3, 4, 5, 6\}$ as potential neighbors for the new node 7 with edge costs shown in Table 4. Determine which node $i \in \mathcal{N}$ will be connected to node 7 in the extend step and compute g_7. Determine which nodes $i \in \mathcal{N}$ will be rewired and compute their updated g_i values.

Table: Edge costs from and to the new node 7 with respect to the graph in Fig. 1.

	0	1	2	3	4	5	6
$c_{7,i}$	4	3	2	1	5	0	0
$c_{i,7}$	4	3	2	1	5	0	0

Solution

Solution. We can compute potential g values for the new node 7.

	0	1	2	3	4	5	6
g_i	0	4	5	7	∞	5	∞
$c_{i,7}$	4	3	2	1	5	0	0
g_7	4	7	7	8	∞	5	∞

Thus, the nearest node n to node 7 is:

$$n = \arg\min_{i \in \mathcal{N}} (g_i + c_{i,7}) = 0. \tag{6}$$

Now consider the rewiring step. Nodes 3,4,5,6 will be rewired.

	0	1	2	3	4	5	6
$c_{7,i}$	4	3	2	1	5	0	0
g_i'	8	7	6	5	9	4	4
g_i	0	4	5	7	∞	5	∞

Problem 4

Consider a discrete-time system with state $x_t \in \mathbb{R}$, input $u_t \in \mathbb{R}$, and motion model:

$$x_{t+1} = x_t + u_t + w_t, \tag{7}$$

where w_t is Gaussian noise with

$$\mathbb{E}[w_t] = 0, \qquad \mathbb{E}[w_t^2] = 1, \qquad \mathbb{E}[w_t^3] = 0.$$

(a) Determine the optimal input u_0^* at $x_0 = \frac{13}{6}$ at time $t = 0$ that minimizes the following stage cost and terminal cost:

$$\ell(x, u) = x^3 + u^2, \qquad q(x) = x^2 \tag{8}$$

for planning horizon $T = 2$ with discount factor $\gamma = 1$. What is the value of u_0^*?

(b) Determine the optimal value function $V^*(x)$ that minimizes the following stage cost:

$$\ell(x, u) = x^2 + u^2 \tag{14}$$

for planning horizon $T = \infty$ with discount factor $\gamma = 0.5$. What is the function $V^*(x)$?

Solution

Solution. We will use the Dynamic Programming algorithm. At $t = T = 2$, we have

$$V_2^*(x) = q(x) = x^2.$$

At $t = 1$, we have:

$$\begin{aligned} V_1^*(x) &= \min_u \{\ell(x,u) + \gamma \mathbb{E}\left[V_2^*(x+u+w)\right]\} \\ &= \min_u \left\{x^3 + u^2 + \mathbb{E}\left[(x+u+w)^2\right]\right\} \\ &= \min_u \left\{x^3 + u^2 + (x+u)^2 + 1\right\} \\ &= x^3 + \frac{1}{2}x^2 + 1. \end{aligned} \tag{9}$$

At $t = 0$, we have:

$$\begin{aligned} V_0^*(x) &= \min_u \{\ell(x,u) + \gamma \mathbb{E}\left[V_1^*(x+u+w)\right]\} \\ &= \min_u \left\{x^3 + u^2 + \mathbb{E}\left[(x+u+w)^3 + \frac{1}{2}(x+u+w)^2 + 1\right]\right\} \\ &= \min_u \left\{x^3 + u^2 + (x+u)^3 + 3(x+u) + \frac{1}{2}(x+u)^2 + \frac{3}{2}\right\}. \end{aligned} \tag{10}$$

Solution

Letting

$$Q_0^*(x,u) = x^3 + u^2 + (x+u)^3 + 3(x+u) + \frac{1}{2}(x+u)^2 + \frac{3}{2},$$

we have:

$$\begin{aligned} 0 &= \frac{d}{du} Q_0^*(x,u) \\ &= 2u + x + u + 3(x+u)^2 + 3 \\ &= 3u^2 + 3(2x+1)u + 3x^2 + x + 3. \end{aligned} \tag{11}$$

Choosing $x = \frac{13}{6}$, we have:

$$u = \frac{-16 \pm 5}{6}. \tag{12}$$

However, since $Q_0^*(13/6, u)$ is a cubic function in u, the value $\bar{u}_0 = -\frac{11}{6}$ is just a local minimizer. Since the input u is not constrained, the global minimum is $-\infty$ at

$$u_0^* = -\infty.$$

Solution

Solution. This is a Linear Quadratic Gaussian (LQG) problem for which the optimal value function $V^*(x)$ satisfies:

$$V^*(x) = \frac{1}{2}Mx^2 + m$$

$$M = 2 + \frac{1}{2}M - \frac{M^2}{8 + 2M} \tag{15}$$

$$m = \frac{1}{2}M.$$

Solving the second equation above for M, we get:

$$M^2 = 8 \quad \Rightarrow \quad M = 2\sqrt{2} \quad \Rightarrow \quad V^*(x) = \sqrt{2}x^2 + \sqrt{2}. \tag{16}$$

Thank you for your great effort and enthusiasm throughout ECE 276B!

Q & A