ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

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

Dynamic Programming and Optimal Control

Fall 2021

Problem Set:
The Dynamic Programming Algorithm

Notes:

- Problems marked with BERTSEKAS are taken from the book *Dynamic Programming and Optimal Control by Dimitri P. Bertsekas, Vol. I, 3rd edition, 2005, 558 pages, hardcover.*
- The solutions were derived by teaching assistants. Please report any error that you may find on Piazza forum.

Problem Set

Problem 1 (Refresher on Probabilities and Expectation, Lecture 1)

Let x be a discrete random variable (DRV) with sample space $\mathcal{X} = \{-1, 0, 1, 2\}$ and probability density function (PDF) $p(x) = ax^2$.

a) Determine a such that $p(x)$ is a valid PDF.

b) Calculate the probability $\Pr(x = 1)$.

c) Calculate the expected value $\mathrm{E}[x]$.

d) Calculate the variance $\mathrm{Var}[x]$.

Problem 2 (Refresher on Probabilities and Expectation, Lecture 1)

Let x and y be continuous random variables (CRVs) with sample space $\mathcal{X} = [-1, 2]$, $\mathcal{Y} = [0, 1]$ and joint PDF $p_{xy}(x, y) = \frac{2}{11}(x + y)^2$.

a) Determine $p_x(x)$.

b) Determine $p_y(y)$.

c) Determine $p_{x|y}(x, y)$.

d) Determine $p_{y|x}(y, x)$.

Problem 3 (BERTSEKAS, p. 51, exercise 1.1 a, c, Lecture 2)

Consider the system

$$x_{k+1} = x_k + u_k + w_k, \qquad k = 0, 1, 2, 3, \tag{1}$$

with initial state $x_0 = 5$, and the cost function

$$\sum_{k=0}^{3}(x_k^2 + u_k^2). \tag{2}$$

Apply the DP algorithm for the following two cases:

a) The control constraint set $U_k(x_k)$ is $\{u \mid 0 \leq x_k + u \leq 5, u : \text{integer}\}$ for all x_k and k, and the disturbance w_k is equal to zero for all k.

b) The control constraint is as in part (a) and the disturbance w_k takes the values -1 and 1 with equal probability $\frac{1}{2}$ for all x_k and u_k, except if $x_k + u_k$ is equal to 0 or 5, in which case $w_k = 0$ with probability 1.

Problem 4 (Chess Game Cost Function, Lecture 2)

In lecture 1 example 3, we maximize the probability of winning the chess game, P_{win} by using the following cost function:

$$g_2(x_2) + \sum_{k=0}^{1} g_k(x_k, u_k, w_k)$$

where

$$g_k(x_k, u_k, w_k) = 0, \quad \forall k \in \{0, 1\}$$
$$g_2(x_2) = \begin{cases} -1 & \text{if } x_2 = (\frac{3}{2}, \frac{1}{2}) \text{ or } (2, 0), \\ -p_w & \text{if } x_2 = (1, 1), \\ 0 & \text{if } x_2 = (\frac{1}{2}, \frac{3}{2}) \text{ or } (0, 2). \end{cases}$$

Show that minimizing this cost function indeed maximizes P_{win}.

Problem 5 (BERTSEKAS, p. 52, exercise 1.3, Lecture 2)

Suppose we have a machine that is either running or is broken down. If it runs throughout one week, it makes a gross profit of \$100. If it fails during the week, gross profit is zero. If it is running at the start of the week and we perform preventive maintenance, the probability that it will fail during the week is 0.4. If we do not perform such maintenance, the probability of failure is 0.7. However, maintenance will cost \$20. When the machine is broken down at the start of the week, it may either be repaired at cost of \$40, in which case it will fail during the week with a probability of 0.4, or it may be replaced at a cost of \$150 by a new machine that is guaranteed to run through its first week of operation. Find the optimal repair, replacement, and maintenance policy that maximizes total profit over four weeks, assuming a new machine at the start of the first week.

Problem 6 (Discounted Cost per Stage, BERTSEKAS, p. 53, exercise 1.6, Lecture 2)

In the framework of the basic problem, consider the case where the cost is of the form

$$\underset{\substack{w_k \\ k=0,1,\ldots,N-1}}{\mathrm{E}} \left\{ \alpha^N g_N(x_N) + \sum_{k=0}^{N-1} \alpha^k g_k(x_k, u_k, w_k) \right\}, \tag{3}$$

where α is a discount factor with $0 < \alpha < 1$. Show that an alternative form of the DP algorithm is given by

$$V_N(x_N) = g_N(x_N), \tag{4}$$

$$V_k(x_k) = \min_{u_k \in U_k(x_k)} \underset{w_k}{\mathrm{E}} \left\{ g_k(x_k, u_k, w_k) + \alpha V_{k+1}\Big(f_k(x_k, u_k, w_k) \Big) \right\}. \tag{5}$$

Clarification: here V_k is used in place of J_k in the DPA as the requirement of the exercise is to find an alternative form. They are two different cost-to-go functions for the same problem.

Problem 7 (Exponential Cost Function, BERTSEKAS, p. 53, exercise 1.7, Lecture 2)

In the framework of the basic problem, consider the case where the cost is of the form

$$\underset{\substack{w_k \\ k=0,1,\dots,N-1}}{\mathrm{E}} \left\{ \exp \left(g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k) \right) \right\}. \tag{6}$$

.

a) Show that the optimal cost and optimal policy can be obtained from the DP-like algorithm

$$J_N(x_N) = \exp\big(g_N(x_N)\big), \tag{7}$$

$$J_k(x_k) = \min_{u_k \in U_k(x_k)} \underset{w_k}{\mathrm{E}} \Big\{ J_{k+1}\big(f_k(x_k, u_k, w_k)\big) \exp\big(g_k(x_k, u_k, w_k)\big) \Big\}. \tag{8}$$

b) Define the function $V_k(x_k) = \ln J_k(x_k)$. Assume also that g_k is a function of x_k and u_k only (and not of w_k). Show that the above algorithm can be rewritten as

$$V_N(x_N) = g_N(x_N), \tag{9}$$

$$V_k(x_k) = \min_{u_k \in U_k(x_k)} \left\{ g_k(x_k, u_k) + \ln \underset{w_k}{\mathrm{E}} \Big\{ \exp\Big(V_{k+1}\Big(f_k(x_k, u_k, w_k) \Big) \Big) \Big\} \right\}. \tag{10}$$

Note: the exponential cost function is an example of a *risk-sensitive cost function* that can be used to encode a preference for policies with a small variance of the cost $g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)$. The associated problems have a lot of interesting properties, which are discussed in several sources, e.g. Whittle [Whi90], Fernandez-Gaucherand and Markus [FeM94], James, Baras, and Elliott [JBE94]. Basar and Bernhard [BaB95].

Problem 8 (Terminating Process, BERTSEKAS, p. 54, exercise 1.8, Lecture 3)

In the framework of the basic problem, consider the case where the system evolution terminates at time i when a given value $\overline{w}_i$ of the disturbance at time i occurs, or when a termination decision $\overline{u}_i$ is made by the controller. If termination occurs at time i, the resulting cost is

$$T + \sum_{k=0}^{i} g_k(x_k, u_k, w_k), \tag{11}$$

where T is a termination cost. If the process has not terminated up to the final time N, the resulting cost is $g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, w_k)$. Reformulate the problem into the framework of the basic problem. *Hint*: Augment the state space with a special termination state.

Problem 9 (Inscribed Polygon of Maximal Perimeter, BERTSEKAS, p. 59, exercise 1.22, Lecture 2)

Consider the problem of inscribing an N-side polygon in a given circle, so that the polygon has maximal perimeter.

a) Formulate the problem as a DP problem involving sequential placement of N points in the circle.

b) Use DP to show that the optimal polygon is regular (all sides are equal).

Problem 10 (Road Trip with a Cabriolet and Weather Forecast, Lecture 3)

Consider a road trip that is split into two intervals. For each interval we can choose a travel distance u_k. We define the state x_k to be the distance traveled so far, hence the system dynamics can be written as

$$x_{k+1} = x_k + u_k, \qquad k = 0, 1. \tag{12}$$

The cost function is given by

$$(x_2 - D)^2 + \sum_{k=0}^{1} w_k u_k^2, \tag{13}$$

where D is the distance we want to travel in total, and u_k appears quadratically to penalize our cost (e.g. too long day rides or fuel consumption). The disturbance w_k takes the value 4 if it is raining, and the value 1 otherwise. This means that, since we are driving a cabriolet, we want to avoid traveling long distances if it is raining. At each time step k, a weather forecast is published that indicates if we are in a cold front or in a warm front. In a cold front the probability of rain is $q_c = \frac{2}{3}$, and in a warm front the probability of rain is $q_w = \frac{1}{3}$. A priori, a cold front and a warm front have equal probability $p_c = p_w = \frac{1}{2}$. Apply the Dynamic Programming Algorithm to find the optimal policy for each possible situation.

Problem 11 (Matlab Programming Exercise, Lecture 2)

Consider the system

$$x_{k+1} = x_k + u_k w_k, \qquad k = 0, 1, 2, \ldots, 9, \tag{14}$$

with initial state x_0 being an integer and constrained to $0 \leq x_0 \leq 10$. The control constraint set is given by $U_k(x_k) := \{u \mid 0 \leq x_k + u \leq 10, u : \text{ integer}\}$ for all x_k and k, and the disturbance w_k takes the value 1 with probability $\frac{1}{3}$ and the value 0 with probability $\frac{2}{3}$ for all x_k and u_k. The cost function is

$$x_{10}^2 + \sum_{k=0}^{9} \left((x_k - x_{ref}(k))^2 + u_k^2 \right), \tag{15}$$

with $x_{ref}(k) = (k-5)^2$. Implement a Matlab script that computes the optimal cost $J_0(x_0)$ for all initial states x_0, $0 \leq x_0 \leq 10$, by applying the DP algorithm.