Solved Examples for Chapter 19

Example for Section 19.2

Lisa often finds that she is up to 1 hour late for work. If she is from 1 to 30 minutes late, \$4 is deducted from her pay check; if she is from 31 to 60 minutes late for work, \$8 is deducted from her paycheck. If she drives to work at her normal speed (which is well under the speed limit), she can arrive in 20 minutes. However, if she exceeds the speed limit a little here and there on her way to work, she can get there in 10 minutes, but she runs the risk of getting a speeding ticket. With a probability of 1/8 she will get caught speeding and will be fined \$20 and delayed 10 minutes, so that it takes 20 minutes to reach work.

As Lisa leaves home, let s be the time she has to reach work before being late; that is, s = 10 means she has 10 minutes to get to work, and s = -10 means she is already 10 minutes late for work. For simplicity, she considers s to be in one of four intervals: (20, ∞), (10, 19), (-10, 9), and (-20, -11).

The transition probabilities for s tomorrow if Lisa does not speed today are given by

	(20, ∞)	(10, 19)	(-10, 9)	(-20, -11)
(20, ∞)	3/8	1/4	1/4	1/8
(10, 19)	1/2	1/4	1/8	1/8
(-10, 9)	5/8	1/4	1/8	0
(-20, -11)	3/4	1/4	0	0

The transition probabilities for s tomorrow if she speeds to work today are given by

	(20, ∞)	(10, 19)	(-10, 9)	(-20, -11)
(20, ∞)				
(10, 19)	3/8	1/4	1/4	1/8
(-10, 9)				
(-20, -11)	5/8	1/4	1/8	0

Note that there are no transition probabilities for (20, ∞) and (-10, 9), because Lisa will get to work on time and from 1 to 30 minutes late, respectively, regardless of whether she speeds. Hence, speeding when in these states would not be a logical choice.

Also note that the transition probabilities imply that the later she is for work and the more she has to rush to get there, the more likely she is to leave for work earlier the next day.

Lisa wishes to determine when she should speed and when she should take her time getting to work in order to minimize her (long-run) expected average cost per day.

(a) Formulate this problem as a Markov decision process by identifying the states and decisions and then finding the C_{ik}.

We define the states and decisions as follows.

State	Condition
0	s in $(20, \infty)$
1	s in (10, 19)
2	s in (-10,9)
3	s in (-20, -11)

Decision	Action
1	Do not speed
2	Speed

Using the data given in the problem statement, the C_{ik} then are

$$C_{01} = C_{02} = 0, \quad C_{11} = 4, \quad C_{12} = \frac{1}{8}(20+4) = 3,$$
$$C_{21} = C_{22} = 4, \quad C_{31} = 8, \quad C_{32} = \frac{1}{8}(20+8)+\frac{7}{8}(4) = 7.$$

(b) Identify all these (stationary deterministic) policies. For each one, find the transition matrix and write an expression for the (long-run) expected average cost per period in terms of the unknown steady-state probabilities ($\pi_0, \pi_1, \ldots, \pi_M$).

She does not speed in either state 0 or state 2, since doing so would not change any penalty for being late. Hence, there are four stationary deterministic policies.

Policy	Verbal Description	$d_0(R)$	$d_1(R)$	$d_2(R)$	$d_3(R)$
R_1	Do not speed	1	1	1	1
R_2	Speed in state 3	1	1	1	2
R_3	Speed in state 1	1	2	1	1
R_4	Speed in states 1 and 3	1	2	1	2

The transition probabilities and expected costs for each policy are given in the following tables.

	Policy: $\mathbf{R_1}$			
State	**0**	**1**	**2**	**3**
0	3/8	1/4	1/4	1/8
1	1/2	1/4	1/8	1/8
2	5/8	1/4	1/8	0
3	3/4	1/4	0	0
Expected Cost	$E(C) = 4\pi_1 + 4\pi_2 + 8\pi_3$			

	Policy: $\mathbf{R_2}$			
State	**0**	**1**	**2**	**3**
0	3/8	1/4	1/4	1/8
1	1/2	1/4	1/8	1/8
2	5/8	1/4	1/8	0
3	5/8	1/4	1/8	0
Expected Cost	$E(C) = 4\pi_1 + 4\pi_2 + 7\pi_3$			

	Policy: $\mathbf{R_3}$			
State	**0**	**1**	**2**	**3**
0	3/8	1/4	1/4	1/8
1	3/8	1/4	1/4	1/8
2	5/8	1/4	1/8	0
3	3/4	1/4	0	0
Expected Cost	$E(C) = 3\pi_1 + 4\pi_2 + 8\pi_3$			

	Policy: $\mathbf{R_4}$			
State	**0**	**1**	**2**	**3**
0	3/8	1/4	1/4	1/8
1	3/8	1/4	1/4	1/8
2	5/8	1/4	1/8	0
3	5/8	1/4	1/8	0
Expected Cost	$E(C) = 3\pi_1 + 4\pi_2 + 7\pi_3$			

(c) Use your IOR Tutorial to find these steady-state probabilities for each policy. Then evaluate the expression obtained in part (b) to find the optimal policy by exhaustive enumeration.

Using IOR Tutorial, we find the steady-state probabilities and the expected average cost for each policy, as summarized in the following table:

Policy	$(\pi_0, \pi_1, \pi_2, \pi_3)$	E(C)
R_1	(0.484, 0.25, 0.174, 0.092)	2.432
R_2	(0.475, 0.25, 0.184, 0.091)	2.373
R_3	(0.459, 0.25, 0.203, 0.089)	2.274
R_4	(0.45, 0.25, 0.213, 0.087)	2.225 ← Minimum

Example for Section 19.3

Consider the scenario involving Lisa and her strategy for driving to work that is described in the preceding example.

(a) Formulate a linear programming model for finding an optimal policy.

Filling the values of the C_{ik} and the $p_{ij}(k)$ found in the preceding example into the form of the linear programming formulation given in Sec. 19.3, the model for finding an optimal policy is

Minimize $Z = 4y_{11} + 4y_{21} + 8y_{31} + 3y_{12} + 7y_{32}$,

subject to

$$\begin{aligned}
y_{01} + y_{11} + y_{12} + y_{21} + y_{31} + y_{32} &= 1 \\
y_{01} - (3/8\, y_{01} + 1/2\, y_{11} + 5/8\, y_{21} + 3/4\, y_{31} + 3/8\, y_{12} + 5/8\, y_{32}) &= 0 \\
y_{11} + y_{12} - 1/4\, (y_{01} + y_{11} + y_{12} + y_{21} + y_{31} + y_{32}) &= 0 \\
y_{21} - (1/4\, y_{01} + 1/8\, y_{11} + 1/8\, y_{21} + 1/4\, y_{12} + 1/8\, y_{32}) &= 0 \\
y_{31} + y_{32} - 1/8\, (y_{01} + y_{11} + y_{12}) &= 0
\end{aligned}$$

and

$$y_{ik} \geq 0, \quad \text{for } i = 0, 1, 2, 3 \text{ and } k = 1, 2.$$

(b) Use the simplex method to solve this model. Use the resulting optimal solution to identify an optimal policy.

The simplex method finds the optimal solution

$$y_{01} = 0.45,\ y_{12} = 0.25,\ y_{21} = 0.213,\ y_{32} = 0.087,\ \text{and } y_{11} = y_{31} = 0.$$

The optimal policy associated with this optimal solution is R_4: Do not speed when in state 0 and state 2; speed when in state 1 and state 3.

Example for Supplement 1 to Chapter 19

Consider the scenario involving Lisa and her strategy for driving to work that is analyzed in the preceding two examples. Use the policy improvement algorithm to find an optimal policy.

We use IOR Tutorial to implement the policy improvement algorithm when starting with the initial policy $d_0(R) = d_1(R) = d_2(R) = d_3(R) = 1$.
The printout from IOR Tutorial is given below:

```
 Markov Decision Processes Model:
  Number of states = 4
  Number of decisions = 2
  Cost Matrix, C(ik):
   _             _
  |‾ 0      ----  ‾|
  |  4       3     |
  |  4      ----   |
  |_ 8       7    _|
  Transition Matrix, p(ij)[1]:
   _                          _
  |‾ 0.375  0.25   0.25   0.125 ‾|
  |  0.5    0.25   0.125  0.125  |
  |  0.625  0.25   0.125  0      |
  |_ 0.75   0.25   0      0     _|

  Transition Matrix, p(ij)[2]:
   _                          _
  |‾ 0      0      0      0     ‾|
  |  0.375  0.25   0.25   0.125  |
  |  0      0      0      0      |
  |_ 0.625  0.25   0.125  0     _|

  Initial Policy:
  d0(R1) = 1
  d1(R1) = 1
  d2(R1) = 1
  d3(R1) = 1

  Average Cost Policy Improvement Algorithm:
  ITERATION # 1

  Value Determination:
  g(R1) = 0    +0.375v0(R1) +  0.25v1(R1) +  0.25v2(R1) + 0.125v3(R1) -
v0(R1)
  g(R1) = 4    +  0.5v0(R1) +  0.25v1(R1) + 0.125v2(R1) + 0.125v3(R1) -
v1(R1)
  g(R1) = 4    +0.625v0(R1) +  0.25v1(R1) + 0.125v2(R1) +     0v3(R1) -
v2(R1)
  g(R1) = 8    + 0.75v0(R1) +  0.25v1(R1) +     0v2(R1) +     0v3(R1) -
v3(R1)


  Solution of Value Determination Equations:
  g(R1)  = 2.43
  v0(R1) = -6.48
  v1(R1) = -2.84
  v2(R1) = -3.65
  v3(R1) = 0

  Policy Improvement:
  State 0:
  0  + 0.375(-6.48) + 0.25 (-2.84) + 0.25 (-3.65) + 0.125(0) - (-6.48)
= 2.43
  --- + 0 (-6.48) + 0    (-2.84) + 0    (-3.65) + 0    (0) - (-6.48) =
---

  State 1:
  4  + 0.5  (-6.48) + 0.25 (-2.84) + 0.125(-3.65) + 0.125(0) - (-2.84)
= 2.43
  3   + 0.375(-6.48) + 0.25 (-2.84) + 0.25 (-3.65) + 0.125(0) - (-2.84)
= 1.785

  State 2:
  4   + 0.625(-6.48) + 0.25 (-2.84) + 0.125(-3.65) + 0    (0) - (-3.65)
= 2.43
  --- + 0    (-6.48) + 0    (-2.84) + 0    (-3.65) + 0    (0) - (-3.65)
= ---

  State 3:
  8    + 0.75 (-6.48) + 0.25 (-2.84) + 0    (-3.65) + 0    (0) - (0) =
2.43
  7    + 0.625(-6.48) + 0.25 (-2.84) + 0.125(-3.65) + 0    (0) - (0) =
1.785

  New Policy:
    d0(R2) = 1
    d1(R2) = 2
    d2(R2) = 1
    d3(R2) = 2




 ITERATION # 2
  Value Determination:
  g(R2) = 0    +0.375v0(R2) +  0.25v1(R2) +  0.25v2(R2) + 0.125v3(R2) -
v0(R2)
  g(R2) = 3    +0.375v0(R2) +  0.25v1(R2) +  0.25v2(R2) + 0.125v3(R2) -
v1(R2)
  g(R2) = 4    +0.625v0(R2) +  0.25v1(R2) + 0.125v2(R2) +     0v3(R2) -
v2(R2)
  g(R2) = 7    +0.625v0(R2) +  0.25v1(R2) + 0.125v2(R2) +     0v3(R2) -
v3(R2)

  Solution of Value Determination Equations:
  g(R2)  = 2.212
  v0(R2) = -5.9
  v1(R2) = -2.9
  v2(R2) = -3
  v3(R2) = 0

  Policy Improvement:
```

```
  State 0:
  0   + 0.375( -5.9) + 0.25 ( -2.9) + 0.25 (   -3) + 0.125(0) - (-5.9 )
= 2.212
  --- + 0    ( -5.9) + 0    ( -2.9) + 0    (   -3) + 0    (0) - (-5.9 )
= ---

  State 1:
  4   + 0.5  ( -5.9) + 0.25 ( -2.9) + 0.125(   -3) + 0.125(0) - (-2.9 )
= 2.85
  3   + 0.375( -5.9) + 0.25 ( -2.9) + 0.25 (   -3) + 0.125(0) - (-2.9 )
= 2.212

  State 2:
  4   + 0.625( -5.9) + 0.25 ( -2.9) + 0.125(   -3) + 0    (0) - (-3   )
= 2.212
  --- + 0    ( -5.9) + 0    ( -2.9) + 0    (   -3) + 0    (0) - (-3   )
= ---

  State 3:
  8    + 0.75 ( -5.9) + 0.25 ( -2.9) + 0    (   -3) + 0    (0) - (0) =
2.85
  7    + 0.625( -5.9) + 0.25 ( -2.9) + 0.125(   -3) + 0    (0) - (0) =
2.212

  Optimal Policy:
    d0(R3) = 1
    d1(R3) = 2
    d2(R3) = 1
    d3(R3) = 2


  g(R3)  = 2.212
  v0(R3) = -5.9
  v1(R3) = -2.9
  v2(R3) = -3
  v3(R3) = 0
```