

# *Dynamic Programming and Optimal Control*

THIRD EDITION

Dimitri P. Bertsekas

Massachusetts Institute of Technology

## *Selected Theoretical Problem Solutions*


Last Updated 10/1/2008

Athena Scientific, Belmont, Mass.

WWW site for book information and orders

<http://www.athenasc.com/>

## NOTE

This solution set is meant to be a significant extension of the scope and coverage of the book. It includes solutions to all of the book's exercises marked with the symbol  .

The solutions are continuously updated and improved, and additional material, including new problems and their solutions are being added. Please send comments, and suggestions for additions and improvements to the author at **dimitrib@mit.edu**

The solutions may be reproduced and distributed for personal or educational uses.

# Solutions Vol. I, Chapter 1

## 1.16 www

(a) Given a sequence of matrix multiplications

$$M_1 M_2 \cdots M_k M_{k+1} \cdots M_N$$

we represent it by a sequence of numbers  $\{n_1, \dots, n_{N+1}\}$ , where  $n_k \times n_{k+1}$  is the dimension of  $M_k$ . Let the initial state be  $x_0 = \{n_1, \dots, n_{N+1}\}$ . Then choosing the first multiplication to be carried out corresponds to choosing an element from the set  $x_0 - \{n_1, n_{N+1}\}$ . For instance, choosing  $n_2$  corresponds to multiplying  $M_1$  and  $M_2$ , which results in a matrix of dimension  $n_1 \times n_3$ , and the initial state must be updated to discard  $n_2$ , the control applied at that stage. Hence at each stage the state represents the dimensions of the matrices resulting from the multiplications done so far. The allowable controls at stage  $k$  are  $u_k \in x_k - \{n_1, n_{N+1}\}$ . The system equation evolves according to

$$x_{k+1} = x_k - \{u_k\}.$$

Note that the control will be applied  $N - 1$  times, therefore the horizon of this problem is  $N - 1$ . The terminal state is  $x_{N-1} = \{n_1, n_{N+1}\}$  and the terminal cost is 0. The cost at stage  $k$  is given by the number of multiplications,

$$g_k(x_k, u_k) = n_a n_{\overline{u}_k} n_b,$$

where  $n_{\overline{u}_k} = u_k$  and

$$\begin{aligned} a &= \max\{i \in \{1, \dots, N+1\} \mid i < \overline{u}_k, i \in x_k\}, \\ b &= \min\{i \in \{1, \dots, N+1\} \mid i > \overline{u}_k, i \in x_k\}. \end{aligned}$$

The DP algorithm for this problem is given by

$$\begin{aligned} J_{N-1}(x_{N-1}) &= 0, \\ J_k(x_k) &= \min_{u_k \in x_k - \{n_1, n_{N+1}\}} \{n_a n_{\overline{u}_k} n_b + J_{k+1}(x_k - \{u_k\})\}, \quad k = 0, \dots, N-2. \end{aligned}$$

Now consider the given problem, where  $N = 3$  and

$$\begin{aligned} M_1 &\text{ is } 2 \times 10, \\ M_2 &\text{ is } 10 \times 5, \\ M_3 &\text{ is } 5 \times 1. \end{aligned}$$

The optimal order is  $M_1(M_2M_3)$ , requiring 70 multiplications.

(b) In this part we can choose a much simpler state space. Let the state at stage  $k$  be given by  $\{a, b\}$ , where  $a, b \in \{1, \dots, N\}$  and give the indices of the first and the last matrix in the current partial product.

There are two possible controls at each stage, which we denote by  $L$  and  $R$ . Note that  $L$  can be applied only when  $a \neq 1$  and  $R$  can be applied only when  $b \neq N$ . The system equation evolves according to

$$x_{k+1} = \begin{cases} \{a-1, b\}, & \text{if } u_k = L, \\ \{a, b+1\}, & \text{if } u_k = R, \end{cases} \quad k = 1, \dots, N-1.$$

The terminal state is  $x_N = \{1, N\}$  with cost 0. The cost at stage  $k$  is given by

$$g_k(x_k, u_k) = \begin{cases} n_{a-1}n_a n_{b+1}, & \text{if } u_k = L, \\ n_a n_{b+1} n_{b+2}, & \text{if } u_k = R, \end{cases} \quad k = 1, \dots, N-1.$$

For the initial stage, we can take  $x_0$  to be the empty set and  $u_0 \in \{1, \dots, N\}$ . The next state will be given by  $x_1 = \{u_0, u_0\}$ , and the cost incurred at the initial stage will be 0 for all possible controls.

### 1.18 www

Let  $t_1 < t_2 < \dots < t_{N-1}$  denote the times where  $g_1(t) = g_2(t)$ . Clearly, it is never optimal to switch functions at any other times. We can therefore divide the problem into  $N-1$  stages, where we want to determine for each stage  $k$  whether or not to switch activities at time  $t_k$ .

Define

$$x_k = \begin{cases} 0 & \text{if on activity } g_1 \text{ just before time } t_k, \\ 1 & \text{if on activity } g_2 \text{ just before time } t_k, \end{cases}$$

$$u_k = \begin{cases} 0 & \text{to continue current activity,} \\ 1 & \text{to switch between activities.} \end{cases}$$

Then the state at time  $t_{k+1}$  is simply  $x_{k+1} = (x_k + u_k) \bmod 2$ , and the profit for stage  $k$  is

$$g_k(x_k, u_k) = \int_{t_k}^{t_{k+1}} g_{1+x_{k+1}}(t) dt - u_k c.$$

The DP algorithm is then

$$J_N(x_N) = 0$$

$$J_k(x_k) = \min_{u_k} [g_k(x_k, u_k) + J_{k+1}((x_k + u_k) \bmod 2)].$$

### 1.21 www

We consider part (b), since part (a) is essentially a special case. We will consider the problem of placing  $N-2$  points between the endpoints  $A$  and  $B$  of the given subarc. We will show that the polygon of maximal area is obtained when the  $N-2$  points are equally spaced on the subarc between  $A$  and  $B$ . Based on geometric considerations, we impose the restriction that the angle between any two successive points is no more than  $\pi$ .

As the subarc is traversed in the clockwise direction, we number sequentially the encountered points as  $x_1, x_2, \dots, x_N$ , where  $x_1$  and  $x_N$  are the two endpoints  $A$  and  $B$  of the arc, respectively. For any point  $x$  on the subarc, we denote by  $\phi$  the angle between  $x$  and  $x_N$  (measured clockwise), and we denote by  $A_k(\phi)$  the maximal area of a polygon with vertices the center of the circle, the points  $x$  and  $x_N$ , and  $N - k - 1$  additional points on the subarc that lie between  $x$  and  $x_N$ .

Without loss of generality, we assume that the radius of the circle is 1, so that the area of the triangle that has as vertices two points on the circle and the center of the circle is  $(1/2) \sin u$ , where  $u$  is the angle corresponding to the center.

By viewing as state the angle  $\phi_k$  between  $x_k$  and  $x_N$ , and as control the angle  $u_k$  between  $x_k$  and  $x_{k+1}$ , we obtain the following DP algorithm

$$A_k(\phi_k) = \max_{0 \leq u_k \leq \min\{\phi_k, \pi\}} \left[ \frac{1}{2} \sin u_k + A_{k+1}(\phi_k - u_k) \right], \quad k = 1, \dots, N-2. \quad (1)$$

Once  $x_{N-1}$  is chosen, there is no issue of further choice of a point lying between  $x_{N-1}$  and  $x_N$ , so we have

$$A_{N-1}(\phi) = \frac{1}{2} \sin \phi, \quad (2)$$

using the formula for the area of the triangle formed by  $x_{N-1}$ ,  $x_N$ , and the center of the circle.

It can be verified by induction that the above algorithm admits the closed form solution

$$A_k(\phi_k) = \frac{1}{2} (N - k) \sin \left( \frac{\phi_k}{N - k} \right), \quad k = 1, \dots, N-1, \quad (3)$$

and that the optimal choice for  $u_k$  is given by

$$u_k^* = \frac{\phi_k}{N - k}.$$

Indeed, the formula (3) holds for  $k = N - 1$ , by Eq. (2). Assuming that Eq. (3) holds for  $k + 1$ , we have from the DP algorithm (1)

$$A_k(\phi_k) = \max_{0 \leq u_k \leq \min\{\phi_k, \pi\}} H_k(u_k, \phi_k), \quad (4)$$

where

$$H_k(u_k, \phi_k) = \frac{1}{2} \sin u_k + \frac{1}{2} (N - k - 1) \sin \left( \frac{\phi_k - u_k}{N - k - 1} \right). \quad (5)$$

It can be verified that for a fixed  $\phi_k$  and in the range  $0 \leq u_k \leq \min\{\phi_k, \pi\}$ , the function  $H_k(\cdot, \phi_k)$  is concave (its second derivative is negative) and its derivative is 0 only at the point  $u_k^* = \phi_k / (N - k)$  which must therefore be its unique maximum. Substituting this value of  $u_k^*$  in Eqs. (4) and (5), we obtain

$$A_k(\phi_k) = \frac{1}{2} \sin \left( \frac{\phi_k}{N - k} \right) + \frac{1}{2} (N - k - 1) \sin \left( \frac{\phi_k - \phi_k / (N - k)}{N - k - 1} \right) = \frac{1}{2} (N - k) \sin \left( \frac{\phi_k}{N - k} \right),$$

and the induction is complete.

Thus, given an optimally placed point  $x_k$  on the subarc with corresponding angle  $\phi_k$ , the next point  $x_{k+1}$  is obtained by advancing clockwise by  $\phi_k / (N - k)$ . This process, when started at  $x_1$  with  $\phi_1$  equal to the angle between  $x_1$  and  $x_N$ , yields as the optimal solution an equally spaced placement of the points on the subarc.

### 1.25 www

(a) Consider the problem with the state equal to the number of free rooms. At state  $x \geq 1$  with  $y$  customers remaining, if the inkeeper quotes a rate  $r_i$ , the transition probability is  $p_i$  to state  $x - 1$  (with a reward of  $r_i$ ) and  $1 - p_i$  to state  $x$  (with a reward of 0). The DP algorithm for this problem starts with the terminal conditions

$$J(x, 0) = J(0, y) = 0, \quad \forall x \geq 0, y \geq 0,$$

and is given by

$$J(x, y) = \max_{i=1, \dots, m} [p_i(r_i + J(x - 1, y - 1)) + (1 - p_i)J(x, y - 1)], \quad \forall x \geq 0.$$

From this equation and the terminal conditions, we can compute sequentially  $J(1, 1), J(1, 2), \dots, J(1, \bar{y})$  up to any desired integer  $\bar{y}$ . Then, we can calculate  $J(2, 1), J(2, 2), \dots, J(2, \bar{y})$ , etc.

We first prove by induction on  $y$  that for all  $y$ , we have

$$J(x, y) \geq J(x - 1, y), \quad \forall x \geq 1.$$

Indeed this is true for  $y = 0$ . Assuming this is true for a given  $y$ , we will prove that

$$J(x, y + 1) \geq J(x - 1, y + 1), \quad \forall x \geq 1.$$

This relation holds for  $x = 1$  since  $r_i > 0$ . For  $x \geq 2$ , by using the DP recursion, this relation is written as

$$\max_{i=1, \dots, m} [p_i(r_i + J(x - 1, y)) + (1 - p_i)J(x, y)] \geq \max_{i=1, \dots, m} [p_i(r_i + J(x - 2, y)) + (1 - p_i)J(x - 1, y)].$$

By the induction hypothesis, each of the terms on the left-hand side is no less than the corresponding term on the right-hand side, so the above relation holds.

The optimal rate is the one that maximizes in the DP algorithm, or equivalently, the one that maximizes

$$p_i r_i + p_i(J(x - 1, y - 1) - J(x, y - 1)).$$

The highest rate  $r_m$  simultaneously maximizes  $p_i r_i$  and minimizes  $p_i$ . Since

$$J(x - 1, y - 1) - J(x, y - 1) \leq 0,$$

as proved above, we see that the highest rate simultaneously maximizes  $p_i r_i$  and  $p_i(J(x - 1, y - 1) - J(x, y - 1))$ , and so it maximizes their sum.

(b) The algorithm given is the algorithm of Exercise 1.22 applied to the problem of part (a). Clearly, it is optimal to accept an offer of  $r_i$  if  $r_i$  is larger than the threshold

$$\bar{r}(x, y) = J(x, y - 1) - J(x - 1, y - 1).$$

1.26 www

(a) The total net expected profit from the (buy/sell) investment decisions after transaction costs are deducted is

$$E \left\{ \sum_{k=0}^{N-1} (u_k P_k(x_k) - c |u_k|) \right\},$$

where

$$u_k = \begin{cases} 1 & \text{if a unit of stock is bought at the } k\text{th period,} \\ -1 & \text{if a unit of stock is sold at the } k\text{th period,} \\ 0 & \text{otherwise.} \end{cases}$$

With a policy that maximizes this expression, we simultaneously maximize the expected total worth of the stock held at time  $N$  minus the investment costs (including sale revenues).

The DP algorithm is given by

$$J_k(x_k) = \max_{u_k \in \{-1, 0, 1\}} \left[ u_k P_k(x_k) - c |u_k| + E\{J_{k+1}(x_{k+1}) \mid x_k\} \right],$$

with

$$J_N(x_N) = 0,$$

where  $J_{k+1}(x_{k+1})$  is the optimal expected profit when the stock price is  $x_{k+1}$  at time  $k+1$ . Since  $u_k$  does not influence  $x_{k+1}$  and  $E\{J_{k+1}(x_{k+1}) \mid x_k\}$ , a decision  $u_k \in \{-1, 0, 1\}$  that maximizes  $u_k P_k(x_k) - c |u_k|$  at time  $k$  is optimal. Since  $P_k(x_k)$  is monotonically nonincreasing in  $x_k$ , it follows that it is optimal to set

$$u_k = \begin{cases} 1 & \text{if } x_k \leq \underline{x}_k, \\ -1 & \text{if } x_k \geq \bar{x}_k, \\ 0 & \text{otherwise,} \end{cases}$$

where  $\underline{x}_k$  and  $\bar{x}_k$  are as in the problem statement. Note that the optimal expected profit  $J_k(x_k)$  is given by

$$J_k(x_k) = E \left\{ \sum_{i=k}^{N-1} \max_{u_i \in \{-1, 0, 1\}} [u_i P_i(x_i) - c |u_i|] \right\}.$$

(b) Let  $n_k$  be the number of units of stock held at time  $k$ . If  $n_k$  is less than  $N - k$  (the number of remaining decisions), then the value  $n_k$  should influence the decision at time  $k$ . We thus take as state the pair  $(x_k, n_k)$ , and the corresponding DP algorithm takes the form

$$V_k(x_k, n_k) = \begin{cases} \max_{u_k \in \{-1, 0, 1\}} [u_k P_k(x_k) - c |u_k| + E\{V_{k+1}(x_{k+1}, n_k + u_k) \mid x_k\}] & \text{if } n_k \geq 1, \\ \max_{u_k \in \{0, 1\}} [u_k P_k(x_k) - c |u_k| + E\{V_{k+1}(x_{k+1}, n_k + u_k) \mid x_k\}] & \text{if } n_k = 0, \end{cases}$$

with

$$V_N(x_N, n_N) = 0.$$

Note that we have

$$V_k(x_k, n_k) = J_k(x_k), \quad \text{if } n_k \geq N - k,$$

where  $J_k(x_k)$  is given by the formula derived in part (a). Using the above DP algorithm, we can calculate  $V_{N-1}(x_{N-1}, n_{N-1})$  for all values of  $n_{N-1}$ , then calculate  $V_{N-2}(x_{N-2}, n_{N-2})$  for all values of  $n_{N-2}$ , etc.

To show the stated property of the optimal policy, we note that  $V_k(x_k, n_k)$  is monotonically nondecreasing with  $n_k$ , since as  $n_k$  decreases, the remaining decisions become more constrained. An optimal policy at time  $k$  is to buy if

$$P_k(x_k) - c + E\{V_{k+1}(x_{k+1}, n_k + 1) - V_{k+1}(x_{k+1}, n_k) \mid x_k\} \geq 0, \quad (1)$$

and to sell if

$$-P_k(x_k) - c + E\{V_{k+1}(x_{k+1}, n_k - 1) - V_{k+1}(x_{k+1}, n_k) \mid x_k\} \geq 0. \quad (2)$$

The expected value in Eq. (1) is nonnegative, which implies that if  $x_k \leq \underline{x}_k$ , implying that  $P_k(x_k) - c \geq 0$ , then the buying decision is optimal. Similarly, the expected value in Eq. (2) is nonpositive, which implies that if  $x_k < \bar{x}_k$ , implying that  $-P_k(x_k) - c < 0$ , then the selling decision cannot be optimal. It is possible that buying at a price greater than  $\underline{x}_k$  is optimal depending on the size of the expected value term in Eq. (1).

(c) Let  $m_k$  be the number of allowed purchase decisions at time  $k$ , i.e.,  $m$  plus the number of sale decisions up to  $k$ , minus the number of purchase decisions up to  $k$ . If  $m_k$  is less than  $N - k$  (the number of remaining decisions), then the value  $m_k$  should influence the decision at time  $k$ . We thus take as state the pair  $(x_k, m_k)$ , and the corresponding DP algorithm takes the form

$$W_k(x_k, m_k) = \begin{cases} \max_{u_k \in \{-1, 0, 1\}} \left[ u_k P_k(x_k) - c |u_k| + E\{W_{k+1}(x_{k+1}, m_k - u_k) \mid x_k\} \right] & \text{if } m_k \geq 1, \\ \max_{u_k \in \{-1, 0\}} \left[ u_k P_k(x_k) - c |u_k| + E\{W_{k+1}(x_{k+1}, m_k - u_k) \mid x_k\} \right] & \text{if } m_k = 0, \end{cases}$$

with

$$W_N(x_N, m_N) = 0.$$

From this point the analysis is similar to the one of part (b).

(d) The DP algorithm takes the form

$$H_k(x_k, m_k, n_k) = \max_{u_k \in \{-1, 0, 1\}} \left[ u_k P_k(x_k) - c |u_k| + E\{H_{k+1}(x_{k+1}, m_k - u_k, n_k + u_k) \mid x_k\} \right]$$

if  $m_k \geq 1$  and  $n_k \geq 1$ , and similar formulas apply for the cases where  $m_k = 0$  and/or  $n_k = 0$  [compare with the DP algorithms of parts (b) and (c)].

(e) Let  $r$  be the interest rate, so that  $x$  invested dollars at time  $k$  will become  $(1 + r)^{N-k}x$  dollars at time  $N$ . Once we redefine the expected profit  $P_k(x_k)$  to be

$$P_k(x) = E\{x_N \mid x_k = x\} - (1 + r)^{N-k}x,$$

the preceding analysis applies.



# Solutions Vol. I, Chapter 2

## 2.4 www

(a) We denote by  $P_k$  the OPEN list *after* having removed  $k$  nodes from OPEN, (i.e., after having performed  $k$  iterations of the algorithm). We also denote  $d_j^k$  the value of  $d_j$  at this time. Let  $b_k = \min_{j \in P_k} \{d_j^k\}$ . First, we show by induction that  $b_0 \leq b_1 \leq \dots \leq b_k$ . Indeed,  $b_0 = 0$  and  $b_1 = \min_j \{a_{sj}\} \geq 0$ , which implies that  $b_0 \leq b_1$ . Next, we assume that  $b_0 \leq \dots \leq b_k$  for some  $k \geq 1$ ; we shall prove that  $b_k \leq b_{k+1}$ . Let  $j_{k+1}$  be the node removed from OPEN during the  $(k+1)$ th iteration. By assumption  $d_{j_{k+1}}^k = \min_{j \in P_k} \{d_j^k\} = b_k$ , and we also have

$$d_i^{k+1} = \min\{d_i^k, d_{j_{k+1}}^k + a_{j_{k+1}i}\}.$$

We have  $P_{k+1} = (P_k - \{j_{k+1}\}) \cup N_{k+1}$ , where  $N_{k+1}$  is the set of nodes  $i$  satisfying  $d_i^{k+1} = d_{j_{k+1}}^k + a_{j_{k+1}i}$  and  $i \notin P_k$ . Therefore,

$$\min_{i \in P_{k+1}} \{d_i^{k+1}\} = \min_{i \in (P_k - \{j_{k+1}\}) \cup N_{k+1}} \{d_i^{k+1}\} = \min \left[ \min_{i \in P_k - \{j_{k+1}\}} \{d_i^{k+1}\}, \min_{i \in N_{k+1}} \{d_i^{k+1}\} \right].$$

Clearly,

$$\min_{i \in N_{k+1}} \{d_i^{k+1}\} = \min_{i \in N_{k+1}} \{d_{j_{k+1}}^k + a_{j_{k+1}i}\} \geq d_{j_{k+1}}^k.$$

Moreover,

$$\begin{aligned} \min_{i \in P_k - \{j_{k+1}\}} \{d_i^{k+1}\} &= \min_{i \in P_k - \{j_{k+1}\}} \left[ \min\{d_i^k, d_{j_{k+1}}^k + a_{j_{k+1}i}\} \right] \\ &\geq \min \left[ \min_{i \in P_k - \{j_{k+1}\}} \{d_i^k\}, d_{j_{k+1}}^k \right] = \min_{i \in P_k} \{d_i^k\} = d_{j_{k+1}}^k, \end{aligned}$$

because we remove from OPEN this node with the *minimum*  $d_i^k$ . It follows that  $b_{k+1} = \min_{i \in P_{k+1}} \{d_i^{k+1}\} \geq d_{j_{k+1}}^k = b_k$ .

Now, we may prove that once a node exits OPEN, it never re-enters. Indeed, suppose that some node  $i$  exits OPEN after the  $k^*$ th iteration of the algorithm; then,  $d_i^{k^*-1} = b_{k^*-1}$ . If node  $i$  re-enters OPEN after the  $\ell^*$ th iteration (with  $\ell^* > k^*$ ), then we have  $d_i^{\ell^*-1} > d_i^{\ell^*} = d_{j_{\ell^*}}^{\ell^*-1} + a_{j_{\ell^*}i} \geq d_{j_{\ell^*}}^{\ell^*-1} = b_{\ell^*-1}$ . On the other hand, since  $d_i$  is *non-increasing*, we have  $b_{k^*-1} = d_i^{k^*-1} \geq d_i^{\ell^*-1}$ . Thus, we obtain  $b_{k^*-1} > b_{\ell^*-1}$ , which contradicts the fact that  $b_k$  is non-decreasing.

Next, we claim the following: after the  $k$ th iteration,  $d_i^k$  equals the length of the shortest possible path from  $s$  to node  $i \in P_k$  under the restriction that all *intermediate nodes belong to*  $C_k$ . The proof will be done by induction on  $k$ . For  $k = 1$ , we have  $C_1 = \{s\}$  and  $d_i^1 = a_{si}$ , and the claim is obviously true. Next, we assume that the claim is true after iterations  $1, \dots, k$ ; we shall show that it is also true after iteration  $k+1$ . The node  $j_{k+1}$  removed from OPEN at the  $(k+1)$ -st iteration satisfies  $\min_{i \in P_k} \{d_i^k\} = d_{j_{k+1}}^k$ . Notice now that all neighbors of the nodes in  $C_k$  belong either to  $C_k$  or to  $P_k$ .

It follows that the shortest path from  $s$  to  $j_{k+1}$  either goes through  $C_k$  or it exits  $C_k$ , then it passes through a node  $j^* \in P_k$ , and eventually reaches  $j_{k+1}$ . In the latter case, the length of this path is at least equal to the length of the shortest path from  $s$  to  $j^*$  through  $C_k$ ; by the induction hypothesis, this equals  $d_{j^*}^k$ , which is at least  $d_{j_{k+1}}^k$ . It follows that, for node  $j_{k+1}$  exiting the OPEN list,  $d_{j_{k+1}}^k$  equals the length of the shortest path from  $s$  to  $j_{k+1}$ . Similarly, all nodes that have exited previously have their current estimate of  $d_i$  equal to the corresponding shortest distance from  $s$ .

Notice now that

$$d_i^{k+1} = \min\{d_i^k, d_{j_{k+1}}^k + a_{j_{k+1}i}\}.$$

For  $i \notin P_k$  and  $i \in P_{k+1}$  it follows that the only neighbor of  $i$  in  $C_{k+1} = C_k \cup \{j_{k+1}\}$  is node  $j_{k+1}$ ; for such a node  $i$ ,  $d_i^k = \infty$ , which leads to  $d_i^{k+1} = d_{j_{k+1}}^k + a_{j_{k+1}i}$ . For  $i \neq j_{k+1}$  and  $i \in P_k$ , the augmentation of  $C_k$  by including  $j_{k+1}$  offers one more path from  $s$  to  $i$  through  $C_{k+1}$ , namely that through  $j_{k+1}$ . Recall that the shortest path from  $s$  to  $i$  through  $C_k$  has length  $d_i^k$  (by the induction hypothesis). Thus,  $d_i^{k+1} = \min\{d_i^k, d_{j_{k+1}}^k + a_{j_{k+1}i}\}$  is the length of the shortest path from  $s$  to  $i$  through  $C_{k+1}$ .

The fact that each node exits OPEN with its current estimate of  $d_i$  being equal to its shortest distance from  $s$  has been proved in the course of the previous inductive argument.

(b) Since each node enters the OPEN list at most once, the algorithm will terminate in at most  $N - 1$  iterations. Updating the  $d_i$ 's during an iteration and selecting the node to exit OPEN requires  $O(N)$  arithmetic operations (i.e., a constant number of operations per node). Thus, the total number of operations is  $O(N^2)$ .

## 2.6 www

**Proposition:** If there exists a path from the origin to each node in  $T$ , the modified version of the label correcting algorithm terminates with  $\text{UPPER} < \infty$  and yields a shortest path from the origin to each node in  $T$ . Otherwise the algorithm terminates with  $\text{UPPER} = \infty$ .

**Proof:** The proof is analogous to the proof of Proposition 3.1. To show that this algorithm terminates, we can use the identical argument in the proof of Proposition 3.1.

Now suppose that for some node  $t \in T$ , there is no path from  $s$  to  $t$ . Then a node  $i$  such that  $(i, t)$  is an arc cannot enter the OPEN list because this would establish that there is a path from  $s$  to  $i$ , and therefore also a path from  $s$  to  $t$ . Thus,  $d_t$  is never changed and  $\text{UPPER}$  is never reduced from its initial value of  $\infty$ .

Suppose now that there is a path from  $s$  to each node  $t \in T$ . Then, since there is a finite number of distinct lengths of paths from  $s$  to each  $t \in T$  that do not contain any cycles, and each cycle has nonnegative length, there is also a shortest path. For some arbitrary  $t$ , let  $(s, j_1, j_2, \dots, j_k, t)$  be a shortest path and let  $d_t^*$  be the corresponding shortest distance. We will show that the value of  $\text{UPPER}$  upon termination must be equal to  $d^* = \max_{t \in T} d_t^*$ . Indeed, each subpath  $(s, j_1, \dots, j_m), m = 1, \dots, k$ , of the shortest path  $(s, j_1, \dots, j_k, t)$  must be a shortest path from  $s$  to  $j_m$ . If the value of  $\text{UPPER}$  is larger than  $d^*$  at termination, the same must be true throughout the algorithm, and therefore  $\text{UPPER}$  will also be larger than the length of all the paths  $(s, j_1, \dots, j_m), m = 1, \dots, k$ , throughout the algorithm, in view of the nonnegative arc length assumption. If, for each  $t \in T$ , the parent node  $j_k$  enters the OPEN list with  $d_{j_k}$  equal to the shortest distance from  $s$  to  $j_k$ ,  $\text{UPPER}$  will be set to  $d^*$  in step 2 immediately following the next time the last of the nodes  $j_k$  is examined by the algorithm in step 2. It follows that, for

some  $\bar{t} \in T$ , the associated parent node  $\bar{j}_k$  will never enter the OPEN list with  $d_{\bar{j}_k}$  equal to the shortest distance from  $s$  to  $\bar{j}_k$ . Similarly, and using also the nonnegative length assumption, this means that node  $\bar{j}_{k-1}$  will never enter the OPEN list with  $d_{\bar{j}_{k-1}}$  equal to the shortest distance from  $s$  to  $\bar{j}_{k-1}$ . Proceeding backwards, we conclude that  $\bar{j}_1$  never enters the OPEN list with  $d_{\bar{j}_1}$  equal to the shortest distance from  $s$  to  $\bar{j}_1$  [which is equal to the length of the arc  $(s, j_1)$ ]. This happens, however, at the first iteration of the algorithm, obtaining a contradiction. It follows that at termination, UPPER will be equal to  $d^*$ .

Finally, it can be seen that, upon termination of the algorithm, the path constructed by tracing the parent nodes backward from  $d$  to  $s$  has length equal to  $d_t^*$  for each  $t \in T$ . Thus the path is a shortest path from  $s$  to  $t$ .

## 2.13 www

(a) We first need to show that  $d_i^k$  is the length of the shortest  $k$ -arc path originating at  $i$ , for  $i \neq t$ . For  $k = 1$ ,

$$d_i^1 = \min_j c_{ij}$$

which is the length of shortest arc out of  $i$ . Assume that  $d_i^{k-1}$  is the length of the shortest  $(k-1)$ -arc path out of  $i$ . Then

$$d_i^k = \min_j \{c_{ij} + d_j^{k-1}\}$$

If  $d_i^k$  is not the length of the shortest  $k$ -arc path, the initial arc of the shortest path must pass through a node other than  $j$ . This is true since  $d_j^{k-1} \leq \text{length of any } (k-1)\text{-step arc out of } j$ . Let  $\ell$  be the alternative node. From the optimality principle

$$\text{distance of path through } \ell = c_{i\ell} + d_\ell^{k-1} \leq d_i^k$$

But this contradicts the choice of  $d_i^k$  in the DP algorithm. Thus,  $d_i^k$  is the length of the shortest  $k$ -arc path out of  $i$ .

Since  $d_t^k = 0$  for all  $k$ , once a  $k$ -arc path out of  $i$  reaches  $t$  we have  $d_i^\kappa = d_i^k$  for all  $\kappa \geq k$ . But with all arc lengths positive,  $d_i^k$  is just the shortest path from  $i$  to  $t$ . Clearly, there is some finite  $k$  such that the shortest  $k$ -path out of  $i$  reaches  $t$ . If this were not true, the assumption of positive arc lengths implies that the distance from  $i$  to  $t$  is infinite. Thus, the algorithm will yield the shortest distances in a finite number of steps. We can estimate the number of steps,  $N_i$  as

$$N_i \leq \frac{\min_j d_{jt}}{\min_{j,k} d_{jk}}$$

(b) Let  $\bar{d}_i^k$  be the distance estimate generated using the initial condition  $d_i^0 = \infty$  and  $\underline{d}_i^k$  be the estimate generated using the initial condition  $d_i^0 = 0$ . In addition, let  $d_i$  be the shortest distance from  $i$  to  $t$ .

**Lemma:**

$$\underline{d}_i^k \leq \underline{d}_i^{k+1} \leq d_i \leq \bar{d}_i^{k+1} \leq \bar{d}_i^k \quad (1)$$

$$\underline{d}_i^k = d_i = \bar{d}_i^k \quad \text{for } k \text{ sufficiently large} \quad (2)$$

**Proof:** Relation (1) follows from the monotonicity property of DP. Note that  $\underline{d}_i^1 \geq \underline{d}_i^0$  and that  $\bar{d}_i^1 \leq \bar{d}_i^0$ . Equation (2) follows immediately from the convergence of DP (given  $d_i^0 = \infty$ ) and from part a).

**Proposition:** For every  $k$  there exists a time  $T_k$  such that for all  $T \geq T_k$

$$\underline{d}_i^k \leq d_i^T \leq \bar{d}_i^k, \quad i = 1, 2, \dots, N$$

**Proof:** The proof follows by induction. For  $k = 0$  the proposition is true, given the positive arc length assumption. Assume it is true for a given  $k$ . Let  $N(i)$  be a set containing all nodes adjacent to  $i$ . For every  $j \in N(i)$  there exists a time,  $T_k^j$  such that

$$\underline{d}_j^k \leq d_j^T \leq \bar{d}_j^k \quad \forall T \geq T_k^j$$

Let  $T'$  be the first time  $i$  updates its distance estimate given that all  $d_j^{T_k^j}$ ,  $j \in N(i)$ , estimates have arrived. Let  $d_{ij}^{T'}$  be the estimate of  $d_j$  that  $i$  has at time  $T'$ . Note that this may differ from  $d_j^{T_k^j}$  since the later estimates from  $j$  may have arrived before  $T'$ . From the Lemma

$$\underline{d}_j^k \leq d_{ij}^{T'} \leq \bar{d}_j^k$$

which, coupled with the monotonicity of DP, implies

$$\underline{d}_i^{k+1} \leq d_i^T \leq \bar{d}_i^{k+1} \quad \forall T \geq T'$$

Since each node never stops transmitting,  $T'$  is finite and the proposition is proved. Using the Lemma, we see that there is a finite  $k$  such that  $\underline{d}_i^\kappa = d_i = \bar{d}_i^\kappa$ ,  $\forall \kappa \geq k$ . Thus, from the proposition, there exists a finite time  $T^*$  such that  $d_i^T = d_i^*$  for all  $T \geq T^*$  and  $i$ .

# *Solutions Vol. I, Chapter 3*

## 3.6 www

This problem is similar to the Brachistochrone Problem (Example 4.2) described in the text. As in that problem, we introduce the system

$$\dot{x} = u$$

and have a fixed terminal state problem  $[x(0) = a \text{ and } x(T) = b]$ . Letting

$$g(x, u) = \frac{\sqrt{1+u^2}}{Cx},$$

the Hamiltonian is

$$H(x, u, p) = g(x, u) + pu.$$

Minimization of the Hamiltonian with respect to  $u$  yields

$$p(t) = -\nabla_u g(x(t), u(t)).$$

Since the Hamiltonian is constant along an optimal trajectory, we have

$$g(x(t), u(t)) - \nabla_u g(x(t), u(t)) u(t) = \text{constant}.$$

Substituting in the expression for  $g$ , we have

$$\frac{\sqrt{1+u^2}}{Cx} - \frac{u^2}{\sqrt{1+u^2} Cx} = \frac{1}{\sqrt{1+u^2} Cx} = \text{constant},$$

which simplifies to

$$(x(t))^2 \left(1 + (\dot{x}(t))^2\right) = \text{constant}.$$

Thus an optimal trajectory satisfies the differential equation

$$\dot{x}(t) = \frac{\sqrt{D - (x(t))^2}}{(x(t))^2}.$$

It can be seen through straightforward calculation that the curve

$$(x(t))^2 + (t - d)^2 = D$$

satisfies this differential equation, and thus the curve of minimum travel time from  $A$  to  $B$  is an arc of a circle.

### 3.9 www

We have the system  $\dot{x}(t) = Ax(t) + Bu(t)$ , for which we want to minimize the quadratic cost

$$x(T)'Q_Tx(T) + \int_0^T (x(t)'Qx(t) + u(t)'Ru(t))dt.$$

The Hamiltonian here is

$$H(x, u, p) = x'Qx + u'Ru + p'(Ax + Bu),$$

and the adjoint equation is

$$\dot{p}(t) = -A'p(t) - 2Qx(t),$$

with the terminal condition

$$p(T) = 2Qx(T).$$

Minimizing the Hamiltonian with respect to  $u$  yields the optimal control

$$\begin{aligned} u^*(t) &= \arg \min_u [x^*(t)'Qx^*(t) + u'Ru + p'(Ax^*(t) + Bu)] \\ &= -\frac{1}{2}R^{-1}B'p(t). \end{aligned}$$

We now hypothesize a linear relation between  $x^*(t)$  and  $p(t)$

$$2K(t)x^*(t) = p(t), \quad \forall t \in [0, T],$$

and show that  $K(t)$  can be obtained by solving the Riccati equation. Substituting this value of  $p(t)$  into the previous equation, we have

$$u^*(t) = -R^{-1}B'K(t)x^*(t).$$

By combining this result with the system equation, we have

$$\dot{x}(t) = (A - BR^{-1}B'K(t))x^*(t). \tag{1}$$

Differentiating  $2K(t)x^*(t) = p(t)$  and using the adjoint equation yields

$$2\dot{K}(t)x^*(t) + 2K(t)\dot{x}^*(t) = -A'2K(t)x^*(t) - 2Qx^*(t).$$

Combining with Eq. (1), we have

$$\dot{K}(t)x^*(t) + K(t)(A - BR^{-1}B'K(t))x^*(t) = -A'K(t)x^*(t) - Qx^*(t),$$

and we thus see that  $K(t)$  should satisfy the Riccati equation

$$\dot{K}(t) = -K(t)A - A'K(t) + K(t)BR^{-1}B'K(t) - Q.$$

From the terminal condition  $p(T) = 2Qx(T)$ , we have  $K(T) = Q$ , from which we can solve for  $K(t)$  using the Riccati equation. Once we have  $K(t)$ , we have the optimal control  $u^*(t) = -R^{-1}B'K(t)x^*(t)$ . By reversing the previous arguments, this control can then be shown to satisfy all the conditions of the Pontryagin Minimum Principle.

# Solutions Vol. I, Chapter 4

## 4.10 www

(a) Clearly, the function  $J_N$  is continuous. Assume that  $J_{k+1}$  is continuous. We have

$$J_k(x) = \min_{u \in \{0,1,\dots\}} \{cu + L(x+u) + G(x+u)\}$$

where

$$\begin{aligned} G(y) &= E_{w_k} \{J_{k+1}(y - w_k)\} \\ L(y) &= E_{w_k} \{p \max(0, w_k - y) + h \max(0, y - w_k)\} \end{aligned}$$

Thus,  $L$  is continuous. Since  $J_{k+1}$  is continuous,  $G$  is continuous for bounded  $w_k$ . Assume that  $J_k$  is not continuous. Then there exists a  $\hat{x}$  such that as  $y \rightarrow \hat{x}$ ,  $J_k(y)$  does not approach  $J_k(\hat{x})$ . Let

$$u^y = \arg \min_{u \in \{0,1,\dots\}} \{cu + L(y+u) + G(y+u)\}$$

Since  $L$  and  $G$  are continuous, the discontinuity of  $J_k$  at  $\hat{x}$  implies

$$\lim_{y \rightarrow \hat{x}} u^y \neq u^{\hat{x}}$$

But since  $u^y$  is optimal for  $y$ ,

$$\lim_{y \rightarrow \hat{x}} \{cu^y + L(y+u^y) + G(y+u^y)\} < \lim_{y \rightarrow \hat{x}} \{cu^{\hat{x}} + L(y+u^{\hat{x}}) + G(y+u^{\hat{x}})\} = J_k(\hat{x})$$

This contradicts the optimality of  $J_k(\hat{x})$  for  $\hat{x}$ . Thus,  $J_k$  is continuous.

(b) Let

$$Y_k(x) = J_k(x+1) - J_k(x)$$

Clearly  $Y_N(x)$  is a non-decreasing function. Assume that  $Y_{k+1}(x)$  is non-decreasing. Then

$$\begin{aligned} Y_k(x+\delta) - Y_k(x) &= c(u^{x+\delta+1} - u^{x+\delta}) - c(u^{x+1} - u^x) \\ &\quad + L(x+\delta+1+u^{x+\delta+1}) - L(x+\delta+u^{x+\delta}) \\ &\quad - [L(x+1+u^{x+1}) - L(x+u^x)] \\ &\quad + G(x+\delta+1+u^{x+\delta+1}) - G(x+\delta+u^{x+\delta}) \\ &\quad - [G(x+1+u^{x+1}) - G(x+u^x)] \end{aligned}$$

Since  $J_k$  is continuous,  $u^{y+\delta} = u^y$  for  $\delta$  sufficiently small. Thus, with  $\delta$  small,

$$\begin{aligned} Y_k(x + \delta) - Y_k(x) &= L(x + \delta + 1 + u^{x+1}) - L(x + \delta + u^x) - [L(x + 1 + u^{x+1}) - L(x + u^x)] \\ &\quad + G(x + \delta + 1 + u^{x+1}) - G(x + \delta + u^x) - [G(x + 1 + u^{x+1}) - G(x + u^x)] \end{aligned}$$

Now, since the control and penalty costs are linear, the optimal order given a stock of  $x$  is less than the optimal order given  $x + 1$  stock plus one unit. Thus

$$u^{x+1} \leq u^x \leq u^{x+1} + 1$$

If  $u^x = u^{x+1} + 1$ ,  $Y(x + \delta) - Y(x) = 0$  and we have the desired result. Assume that  $u^x = u^{x+1}$ . Since  $L(x)$  is convex,  $L(x + 1) - L(x)$  is non-decreasing. Using the assumption that  $Y_{k+1}(x)$  is non-decreasing, we have

$$\begin{aligned} Y_k(x + \delta) - Y_k(x) &= \underbrace{L(x + \delta + 1 + u^x) - L(x + \delta + u^x) - [L(x + 1 + u^x) - L(x + u^x)]}_{\geq 0} \\ &\quad + \underbrace{E_{w_k} \{ J_{k+1}(x + \delta + 1 + u^x - w_k) - J_{k+1}(x + \delta + u^x - w_k) \\ &\quad - [J_{k+1}(x + 1 + u^x - w_k) - J_{k+1}(x + u^x - w_k)] \}}_{\geq 0} \\ &\geq 0 \end{aligned}$$

Thus,  $Y_k(x)$  is a non-decreasing function in  $x$ .

(c) From their definition and a straightforward induction it can be shown that  $J_k^*(x)$  and  $J_k(x, u)$  are bounded below. Furthermore, since  $\lim_{x \rightarrow \infty} L_k(x, u) = \infty$ , we obtain  $\lim_{x \rightarrow \infty} J_k(x, 0) = \infty$ .

From the definition of  $J_k(x, u)$ , we have

$$J_k(x, u) = J_k(x + 1, u - 1) + c, \quad \forall u \in \{1, 2, \dots\}. \quad (1)$$

Let  $S_k$  be the smallest real number satisfying

$$J_k(S_k, 0) = J_k(S_k + 1, 0) + c \quad (2)$$

We show that  $S_k$  is well defined. If no  $S_k$  satisfying Eq. (2) exists, we must have either  $J_k(x, 0) - J_k(x + 1, 0) > c, \forall x \in \mathcal{R}$  or  $J_k(x, 0) - J_k(x + 1, 0) < 0, \forall x \in \mathcal{R}$ , because  $J_k$  is continuous. The first possibility contradicts the fact that  $\lim_{x \rightarrow \infty} J_k(x, 0) = \infty$ . The second possibility implies that  $\lim_{x \rightarrow -\infty} J_k(x, 0) + cx$  is finite. However, using the boundedness of  $J_{k+1}^*(x)$  from below, we obtain  $\lim_{x \rightarrow -\infty} J_k(x, 0) + cx = \infty$ . The contradiction shows that  $S_k$  is well defined.

We now derive the form of an optimal policy  $u_k^*(x)$ . Fix some  $x$  and consider first the case  $x \geq S_k$ . Using the fact that  $J_k(x, u) - J_k(x + 1, u)$  is nondecreasing function of  $x$  we have for any  $u \in \{0, 1, 2, \dots\}$

$$J_k(x + 1, u) - J_k(x, u) \geq J_k(S_k + 1, u) - J_k(S_k, u) = J_k(S_k + 1, 0) - J_k(S_k, 0) = -c$$

Therefore,

$$J_k(x, u + 1) = J_k(x + 1, u) + c \geq J_k(x, u) \quad \forall u \in \{0, 1, \dots\}, \forall x \geq S_k.$$



This shows that  $u = 0$  minimizes  $J_k(x, u)$ , for all  $x \geq S_k$ . Now let  $x \in [S_k - n, S_k - n + 1]$ ,  $n \in \{1, 2, \dots\}$ . Using Eq. (1), we have

$$J_k(x, n + m) - J_k(x, n) = J_k(x + n, m) - J_k(x + n, 0) \geq 0 \quad \forall m \text{ in } \{0, 1, \dots\}. \quad (3)$$

However, if  $u < n$  then  $x + u < S_k$  and

$$J_k(x + u + 1, 0) - J_k(x + u, 0) < J_k(S_k + 1, 0) - J_k(S_k, 0) = -c.$$

Therefore,

$$J_k(x, u + 1) = J_k(x + u + 1, 0) + (u + 1)c < J_k(x + u, 0) + uc = J_k(x, u) \quad \forall u \in \{0, 1, \dots\}, \quad n < n. \quad (4)$$

Inequalities (3) and (4) show that  $u = n$  minimizes  $J_k(x, u)$  whenever  $x \in [S_k - n, S_k - n + 1]$ .

#### 4.18 www

Let the state  $x_k$  be defined as

$$x_k = \begin{cases} T, & \text{if the selection has already terminated} \\ 1, & \text{if the } k^{\text{th}} \text{ object observed has rank 1} \\ 0, & \text{if the } k^{\text{th}} \text{ object observed has rank } < 1 \end{cases}$$

The system evolves according to

$$x_{k+1} = \begin{cases} T, & \text{if } u_k = \text{stop or } x_k = T \\ w_k, & \text{if } u_k = \text{continue} \end{cases}$$

The cost function is given by

$$g_k(x_k, u_k, w_k) = \begin{cases} \frac{k}{N}, & \text{if } x_k = 1 \text{ and } u_k = \text{stop} \\ 0, & \text{otherwise} \end{cases}$$

$$g_N(x_N) = \begin{cases} 1, & \text{if } x_N = 1 \\ 0, & \text{otherwise} \end{cases}$$

Note that if termination is selected at stage  $k$  and  $x_k \neq 1$  then the probability of success is 0. Thus, if  $x_k = 0$  it is always optimal to continue.

To complete the model we have to determine  $P(w_k | x_k, u_k) \triangleq P(w_k)$  when the control  $u_k = \text{continue}$ . At stage  $k$ , we have already selected  $k$  objects from a sorted set. Since we know nothing else about these objects the new element can, with equal probability, be in any relation with the already observed objects  $a_j$

$$\underbrace{\dots < a_{i_1} < \dots < a_{i_2} < \dots \quad \dots < a_{i_k} \dots}_{k+1 \text{ possible positions for } a_{k+1}}$$

Thus,

$$P(w_k = 1) = \frac{1}{k+1}, \quad P(w_k = 0) = \frac{k}{k+1}$$

**Proposition:** If  $k \in S_N \triangleq \{i \mid \left(\frac{1}{N-1} + \dots + \frac{1}{i}\right) \leq 1\}$ , then

$$J_k(0) = \frac{k}{N} \left( \frac{1}{N-1} + \dots + \frac{1}{k} \right), \quad J_k(1) = \frac{k}{N}.$$

**Proof:** For  $k = N - 1$

$$J_{N-1}(0) = \max \left[ \underbrace{0}_{\text{stop}}, \underbrace{E\{w_{N-1}\}}_{\text{continue}} \right] = \frac{1}{N}$$

and  $\mu_{N-1}^*(0) = \text{continue}$ . Also,

$$J_{N-1}(1) = \max \left[ \underbrace{\frac{N-1}{N}}_{\text{stop}}, \underbrace{E\{w_{N-1}\}}_{\text{continue}} \right] = \frac{N-1}{N}$$

and  $\mu_{N-1}^*(1) = \text{stop}$ . Note that  $N - 1 \in S_N$  for all  $S_N$ .

Assume the proposition is true for  $J_{k+1}(x_{k+1})$ . Then

$$J_k(0) = \max \left[ \underbrace{0}_{\text{stop}}, \underbrace{E\{J_{k+1}(w_k)\}}_{\text{continue}} \right]$$

$$J_k(1) = \max \left[ \underbrace{\frac{k}{N}}_{\text{stop}}, \underbrace{E\{J_{k+1}(w_k)\}}_{\text{continue}} \right]$$

Now,

$$\begin{aligned} E\{J_{k+1}(w_k)\} &= \frac{1}{k+1} \frac{k+1}{N} + \frac{k}{k+1} \frac{k+1}{N} \left( \frac{1}{N-1} + \dots + \frac{1}{k+1} \right) \\ &= \frac{k}{N} \left( \frac{1}{N-1} + \dots + \frac{1}{k} \right) \end{aligned}$$

Clearly

$$J_k(0) = \frac{k}{N} \left( \frac{1}{N-1} + \dots + \frac{1}{k} \right)$$

and  $\mu_k^*(0) = \text{continue}$ . If  $k \in S_N$ ,

$$J_k(1) = \frac{k}{N}$$

and  $\mu_k^*(1) = \text{stop}$ . **Q.E.D.**

**Proposition:** If  $k \notin S_N$

$$J_k(0) = J_k(1) = \frac{\delta-1}{N} \left( \frac{1}{N-1} + \cdots + \frac{1}{\delta-1} \right)$$

where  $\delta$  is the minimum element of  $S_N$ .

**Proof:** For  $k = \delta - 1$

$$\begin{aligned} J_k(0) &= \frac{1}{\delta} \frac{\delta}{N} + \frac{\delta-1}{\delta} \frac{\delta}{N} \left( \frac{1}{N-1} + \cdots + \frac{1}{\delta} \right) \\ &= \frac{\delta-1}{N} \left( \frac{1}{N-1} + \cdots + \frac{1}{\delta-1} \right) \end{aligned}$$

$$\begin{aligned} J_k(1) &= \max \left[ \frac{\delta-1}{N}, \frac{\delta-1}{N} \left( \frac{1}{N-1} + \cdots + \frac{1}{\delta-1} \right) \right] \\ &= \frac{\delta-1}{N} \left( \frac{1}{N-1} + \cdots + \frac{1}{\delta-1} \right) \end{aligned}$$

and  $\mu_{\delta-1}^*(0) = \mu_{\delta-1}^*(1) = \text{continue}$ .

Assume the proposition is true for  $J_k(x_k)$ . Then

$$J_{k-1}(0) = \frac{1}{k} J_k(1) + \frac{k-1}{k} J_k(0) = J_k(0)$$

and  $\mu_{k-1}^*(0) = \text{continue}$ .

$$\begin{aligned} J_{k-1}(1) &= \max \left[ \frac{1}{k} J_k(1) + \frac{k-1}{k} J_k(0), \frac{k-1}{N} \right] \\ &= \max \left[ \frac{\delta-1}{N} \left( \frac{1}{N-1} + \cdots + \frac{1}{\delta-1} \right), \frac{k-1}{N} \right] \\ &= J_k(0) \end{aligned}$$

and  $\mu_{k-1}^*(1) = \text{continue}$ . **Q.E.D.**

Thus the optimum policy is to continue until the  $\delta^{\text{th}}$  object, where  $\delta$  is the minimum integer such that  $\left( \frac{1}{N-1} + \cdots + \frac{1}{\delta} \right) \leq 1$ , and then stop at the first time an element is observed with largest rank.

#### 4.31 www

(a) In order that  $A_k x + B_k u + w \in X$  for all  $w \in W_k$ , it is sufficient that  $A_k x + B_k u$  belong to some ellipsoid  $\tilde{X}$  such that the vector sum of  $\tilde{X}$  and  $W_k$  is contained in  $X$ . The ellipsoid

$$\tilde{X} = \{z \mid z' F z \leq 1\},$$

where for some scalar  $\beta \in (0, 1)$ ,

$$F^{-1} = (1 - \beta)(\Psi^{-1} - \beta^{-1}D_k^{-1})$$

has this property (based on the hint and assuming that  $F^{-1}$  is well-defined as a positive definite matrix). Thus, it is sufficient that  $x$  and  $u$  are such that

$$(A_k x + B_k u)' F (A_k x + B_k u) \leq 1. \quad (1)$$

In order that for a given  $x$ , there exists  $u$  with  $u' R_k u \leq 1$  such that Eq. (1) is satisfied as well as

$$x' \Xi x \leq 1,$$

it is sufficient that  $x$  is such that

$$\min_{u \in \mathbb{R}^m} [x' \Xi x + u' R_k u + (A_k x + B_k u)' F (A_k x + B_k u)] \leq 1, \quad (2)$$

or by carrying out explicitly the quadratic minimization above,

$$x' K x \leq 1,$$

where

$$K = A'_k (F^{-1} + B_k R_k^{-1} B'_k)^{-1} + \Xi.$$

The control law

$$\mu(x) = -(R_k + B'_k F B_k)^{-1} B'_k F A_k x$$

attains the minimum in Eq. (2) for all  $x$ , so it achieves reachability.

(b) Follows by iterative application of the results of part (a), starting with  $k = N - 1$  and proceeding backwards.

(c) Follows from the arguments of part (a).

# Solutions Vol. I, Chapter 5

## 5.1 www

Define

$$y_N = x_N, \\ y_k = x_k + A_k^{-1}w_k + A_k^{-1}A_{k+1}^{-1}w_{k+1} + \dots + A_k^{-1}\dots A_{N-1}^{-1}w_{N-1}.$$

Then

$$\begin{aligned} y_k &= x_k + A_k^{-1}(w_k - x_{k+1}) + A_k^{-1}y_{k+1} \\ &= x_k + A_k^{-1}(-A_kx_k - B_ku_k) + A_k^{-1}y_{k+1} \\ &= -A_k^{-1}B_ku_k + A_k^{-1}y_{k+1} \end{aligned}$$

and

$$y_{k+1} = A_k y_k + B_k u_k.$$

Now, the cost function is the expected value of

$$x_N' Q x_N + \sum_{k=0}^{N-1} u_k' R_k u_k = y_0' K_0 y_0 + \sum_{k=0}^{N-1} (y_{k+1}' K_{k+1} y_{k+1} - y_k' K_k y_k + u_k' R_k u_k).$$

We have

$$\begin{aligned} y_{k+1}' K_{k+1} y_{k+1} - y_k' K_k y_k + u_k' R_k u_k &= (A_k y_k + B_k u_k)' K_{k+1} (A_k y_k + B_k u_k) + u_k' R_k u_k \\ &\quad - y_k' A_k' [K_{k+1} - K_{k+1} B_k (B_k' K_{k+1} B_k)^{-1} B_k' K_{k+1}] A_k y_k \\ &= y_k' A_k' K_{k+1} A_k y_k + 2y_k' A_k' K_{k+1} B_k u_k + u_k' B_k' K_{k+1} B_k u_k \\ &\quad - y_k' A_k' K_{k+1} A_k y_k + y_k' A_k' K_{k+1} B_k P_k^{-1} B_k' K_{k+1} A_k y_k \\ &\quad + u_k' R_k u_k \\ &= -2y_k' L_k' P_k u_k + u_k' P_k u_k + y_k' L_k' P_k L_k y_k \\ &= (u_k - L_k y_k)' P_k (u_k - L_k y_k). \end{aligned}$$

Thus, the cost function can be written as

$$E \left\{ y_0' K_0 y_0 + \sum_{k=0}^{N-1} (u_k - L_k y_k)' P_k (u_k - L_k y_k) \right\}.$$

The problem now is to find  $\mu_k^*(I_k)$ ,  $k = 0, 1, \dots, N-1$ , that minimize over admissible control laws  $\mu_k(I_k)$ ,  $k = 0, 1, \dots, N-1$ , the cost function

$$E \left\{ y_0' K_0 y_0 + \sum_{k=0}^{N-1} (\mu_k(I_k) - L_k y_k)' P_k (\mu_k(I_k) - L_k y_k) \right\}.$$

We do this minimization by first minimizing over  $\mu_{N-1}$ , then over  $\mu_{N-2}$ , etc. The minimization over  $\mu_{N-1}$  involves just the last term in the sum and can be written as

$$\begin{aligned} \min_{\mu_{N-1}} E \left\{ (\mu_{N-1}(I_{N-1}) - L_{N-1}y_{N-1})' P_{N-1} (\mu_{N-1}(I_{N-1}) - L_{N-1}y_{N-1}) \right\} \\ = E \left\{ \min_{u_{N-1}} E \left\{ (u_{N-1} - L_{N-1}y_{N-1})' P_{N-1} (u_{N-1} - L_{N-1}y_{N-1}) \middle| I_{N-1} \right\} \right\}. \end{aligned}$$

Thus this minimization yields the optimal control law for the last stage:

$$\mu_{N-1}^*(I_{N-1}) = L_{N-1} E \left\{ y_{N-1} \middle| I_{N-1} \right\}.$$

[Recall here that, generically,  $E\{z|I\}$  minimizes over  $u$  the expression  $E_z\{(u-z)'P(u-z)|I\}$  for any random variable  $z$ , any conditioning variable  $I$ , and any positive semidefinite matrix  $P$ .] The minimization over  $\mu_{N-2}$  involves

$$\begin{aligned} E \left\{ (\mu_{N-2}(I_{N-2}) - L_{N-2}y_{N-2})' P_{N-2} (\mu_{N-2}(I_{N-2}) - L_{N-2}y_{N-2}) \right\} \\ + E \left\{ (E\{y_{N-1}|I_{N-1}\} - y_{N-1})' L_{N-1}' P_{N-1} L_{N-1} (E\{y_{N-1}|I_{N-1}\} - y_{N-1}) \right\}. \end{aligned}$$

However, as in Lemma 5.2.1, the term  $E\{y_{N-1}|I_{N-1}\} - y_{N-1}$  does not depend on any of the controls (it is a function of  $x_0, w_0, \dots, w_{N-2}, v_0, \dots, v_{N-1}$ ). Thus the minimization over  $\mu_{N-2}$  involves just the first term above and yields similarly as before

$$\mu_{N-2}^*(I_{N-2}) = L_{N-2} E \left\{ y_{N-2} \middle| I_{N-2} \right\}.$$

Proceeding similarly, we prove that for all  $k$

$$\mu_k^*(I_k) = L_k E \left\{ y_k \middle| I_k \right\}.$$

*Note:* The preceding proof can be used to provide a quick proof of the separation theorem for linear-quadratic problems in the case where  $x_0, w_0, \dots, w_{N-1}, v_0, \dots, v_{N-1}$  are independent. If the cost function is

$$E \left\{ x_N' Q_N x_N + \sum_{k=0}^{N-1} (x_k' Q_k x_k + u_k' R_k u_k) \right\}$$

the preceding calculation can be modified to show that the cost function can be written as

$$E \left\{ x_0' K_0 x_0 + \sum_{k=0}^{N-1} ((u_k - L_k x_k)' P_k (u_k - L_k x_k) + w_k' K_{k+1} w_k) \right\}.$$

By repeating the preceding proof we then obtain the optimal control law as

$$\mu_k^*(I_k) = L_k E \left\{ x_k \middle| I_k \right\}$$

### 5.3 www

The control at time  $k$  is  $(u_k, \alpha_k)$ , where  $\alpha_k$  is a variable taking values 1 (if the next measurement at time  $k + 1$  is of type 1) or 2 (if the next measurement is of type 2). The cost functional is

$$E \left\{ x_N' Q x_N + \sum_{k=0}^{N-1} (x_k' Q x_k + u_k' R u_k) + \sum_{k=0}^{N-1} g_{\alpha_k} \right\}.$$

We apply the DP algorithm for  $N = 2$ . We have from the Riccati equation

$$\begin{aligned} J_1(I_1) &= J_1(z_0, z_1, u_0, \alpha_0) \\ &= E_{x_1} \{ x_1' (A' Q A + Q) x_1 \mid I_1 \} + E_{w_1} \{ w' Q w \} \\ &\quad + \min_{u_1} \{ u_1' (B' Q B + R) u_1 + 2 E \{ x_1 \mid I_1 \}' A' Q B u_1 \} \\ &\quad + \min[g_1, g_2]. \end{aligned}$$

So

$$\begin{aligned} \mu_1^*(I_1) &= -(B' Q B + R)^{-1} B' Q A E \{ x_1 \mid I_1 \}, \\ \alpha_1^*(I_1) &= \begin{cases} 1, & \text{if } g_1 \leq g_2, \\ 2, & \text{otherwise.} \end{cases} \end{aligned}$$

Note that the measurement selected at  $k = 1$  does not depend on  $I_1$ . This is intuitively clear since the measurement  $z_2$  will not be used by the controller, so its selection should be based on measurement cost alone and not on the basis of the quality of estimate. The situation is different once more than one stage is considered.

Using a simple modification of the analysis in Section 5.2 of the text, we have

$$\begin{aligned} J_0(I_0) &= J_0(z_0) \\ &= \min_{u_0} \left\{ E_{x_0, w_0} \{ x_0' Q x_0 + u_0' R u_0 + A x_0 + B u_0 + w_0' K_0 A x_0 + B u_0 + w_0 \mid z_0 \} \right\} \\ &\quad + \min_{\alpha_0} \left[ E_{z_1} \left\{ E_{x_1} \{ [x_1 - E \{ x_1 \mid I_1 \}]' P_1 [x_1 - E \{ x_1 \mid I_1 \}] \mid I_1 \} \mid z_0, u_0, \alpha_0 \right\} + g_{\alpha_0} \right] \\ &\quad + E_{w_1} \{ w_1' Q w_1 \} + \min[g_1, g_2]. \end{aligned}$$

Note that the minimization of the second bracket is indicated only with respect to  $\alpha_0$  and not  $u_0$ . The reason is that quantity in the second bracket is the error covariance of the estimation error (weighted by  $P_1$ ) and, as shown in the text, it does not depend on  $u_0$ . Because all stochastic variables are Gaussian, the quantity in the second bracket does not depend on  $z_0$ . (The weighted error covariance produced by the Kalman filter is precomputable and depends only on the system and measurement matrices and noise covariances but not on the measurements received.) In fact

$$\begin{aligned} &E_{z_1} \left\{ E_{x_1} \{ [x_1 - E \{ x_1 \mid I_1 \}]' P_1 [x_1 - E \{ x_1 \mid I_1 \}] \mid I_1 \} \mid z_0, u_0, \alpha_0 \right\} \\ &= \begin{cases} \text{Tr} \left( P_1^{\frac{1}{2}} \sum_{1|1}^1 P_1^{\frac{1}{2}} \right), & \text{if } \alpha_0 = 1, \\ \text{Tr} \left( P_1^{\frac{1}{2}} \sum_{1|1}^2 P_1^{\frac{1}{2}} \right), & \text{if } \alpha_0 = 2, \end{cases} \end{aligned}$$

where  $Tr(\cdot)$  denotes the trace of a matrix, and  $\sum_{1|1}^1 (\sum_{1|1}^2)$  denotes the error covariance of the Kalman filter estimate if a measurement of type 1 (type 2) is taken at  $k = 0$ . Thus at time  $k = 0$ , we have that the optimal measurement chosen does not depend on  $z_0$  and is of type 1 if

$$Tr \left( P_1^{\frac{1}{2}} \Sigma_{1|1}^1 P_1^{\frac{1}{2}} \right) + g_1 \leq Tr \left( P_1^{\frac{1}{2}} \Sigma_{1|1}^2 P_1^{\frac{1}{2}} \right) + g_2$$

and is of type 2 otherwise.

## 5.7 www

a) We have

$$\begin{aligned} p_{k+1}^j &= P(x_{k+1} = j \mid z_0, \dots, z_{k+1}, u_0, \dots, u_k) \\ &= P(x_{k+1} = j \mid I_{k+1}) \\ &= \frac{P(x_{k+1} = j, z_{k+1} \mid I_k, u_k)}{P(z_{k+1} \mid I_k, u_k)} \\ &= \frac{\sum_{i=1}^n P(x_k = i) P(x_{k+1} = j \mid x_k = i, u_k) P(z_{k+1} \mid u_k, x_{k+1} = j)}{\sum_{s=1}^n \sum_{i=1}^n P(x_k = i) P(x_{k+1} = s \mid x_k = i, u_k) P(z_{k+1} \mid u_k, x_{k+1} = s)} \\ &= \frac{\sum_{i=1}^n p_k^i p_{ij}(u_k) r_j(u_k, z_{k+1})}{\sum_{s=1}^n \sum_{i=1}^n p_k^i p_{is}(u_k) r_s(u_k, z_{k+1})}. \end{aligned}$$

Rewriting  $p_{k+1}^j$  in vector form, we have

$$p_{k+1}^j = \frac{r_j(u_k, z_{k+1}) [P(u_k)' P_k]_j}{\sum_{s=1}^n r_s(u_k, z_{k+1}) [P(u_k)' P_k]_s}, \quad j = 1, \dots, n.$$

Therefore,

$$P_{k+1} = \frac{[r(u_k, z_{k+1})] * [P(u_k)' P_k]}{r(u_k, z_{k+1})' P(u_k)' P_k}.$$

b) The DP algorithm for this system is:

$$\begin{aligned} \bar{J}_{N-1}(P_{N-1}) &= \min_u \left\{ \sum_{i=1}^n p_{N-1}^i \sum_{j=1}^n p_{ij}(u) g_{N-1}(i, u, j) \right\} \\ &= \min_u \left\{ \sum_{i=1}^n p_{N-1}^i [G_{N-1}(u)]_i \right\} \\ &= \min_u \{ P'_{N-1} G_{N-1}(u) \} \end{aligned}$$



$$\begin{aligned}\bar{J}_k(P_k) &= \min_u \left\{ \sum_{i=1}^n p_k^i \sum_{j=1}^n p_{ij}(u) g_k(i, u, j) + \sum_{i=1}^n p_k^i \sum_{j=1}^n p_{ij}(u) \sum_{\theta=1}^q r_j(u, \theta) \bar{J}_{k+1}(P_{k+1} | P_k, u, \theta) \right\} \\ &= \min_u \left\{ P'_k G_k(u) + \sum_{\theta=1}^q r(u, \theta)' P(u)' P_k \bar{J}_{k+1} \left[ \frac{[r(u, \theta)] * [P(u)' P_k]}{r(u, \theta)' P(u)' P_k} \right] \right\}.\end{aligned}$$

c) For  $k = N - 1$ ,

$$\begin{aligned}\bar{J}_{N-1}(\lambda P'_{N-1}) &= \min_u \{ \lambda P'_{N-1} G_{N-1}(u) \} \\ &= \min_u \left\{ \sum_{i=1}^n \lambda p_{N-1}^i [G_{N-1}(u)]_i \right\} \\ &= \min_u \left\{ \lambda \sum_{i=1}^n p_{N-1}^i [G_{N-1}(u)]_i \right\} \\ &= \lambda \min_u \left\{ \sum_{i=1}^n p_{N-1}^i [G_{N-1}(u)]_i \right\} \\ &= \lambda \min_u \left\{ \sum_{i=1}^n p_{N-1}^i [G_{N-1}(u)]_i \right\} \\ &= \lambda \bar{J}_{N-1}(P_{N-1}).\end{aligned}$$

Now assume  $\bar{J}_k(\lambda P_k) = \lambda \bar{J}_k(P_k)$ . Then,

$$\begin{aligned}\bar{J}_{k-1}(\lambda P'_{k-1}) &= \min_u \left\{ \lambda P'_{k-1} G_{k-1}(u) + \sum_{\theta=1}^q r(u, \theta)' P(u)' \lambda P_{k-1} \bar{J}_k(P_k | P_{k-1}, u, \theta) \right\} \\ &= \min_u \left\{ \lambda P'_{k-1} G_{k-1}(u) + \lambda \sum_{\theta=1}^q r(u, \theta)' P(u)' P_{k-1} \bar{J}_k(P_k | P_{k-1}, u, \theta) \right\} \\ &= \lambda \min_u \left\{ P'_{k-1} G_{k-1}(u) + \sum_{\theta=1}^q r(u, \theta)' P(u)' P_{k-1} \bar{J}_k(P_k | P_{k-1}, u, \theta) \right\} \\ &= \lambda \bar{J}_{k-1}(P_{k-1}).\end{aligned}\quad \text{Q.E.D.}$$

For any  $u$ ,  $r(u, \theta)' P(u)' P_k$  is a scalar. Therefore, letting  $\lambda = r(u, \theta)' P(u)' P_k$ , we have

$$\begin{aligned}\bar{J}_k(P_k) &= \min_u \left\{ P'_k G_k(u) + \sum_{\theta=1}^q r(u, \theta)' P(u)' P_k \bar{J}_{k+1} \left[ \frac{[r(u, \theta)] * [P(u)' P_k]}{r(u, \theta)' P(u)' P_k} \right] \right\} \\ &= \min_u \left[ P'_k G_k(u) + \sum_{\theta=1}^q \bar{J}_{k+1}([r(u, \theta)] * [P(u)' P_k]) \right].\end{aligned}$$

d) For  $k = N - 1$ , we have  $\bar{J}_{N-1}(P_{N-1}) = \min_u [P'_{N-1} G_{N-1}(u)]$ , and so  $\bar{J}_{N-1}(P_{N-1})$  has the desired form

$$\bar{J}_{N-1}(P_{N-1}) = \min [P'_{N-1} \alpha_{N-1}^1, \dots, P'_{N-1} \alpha_{N-1}^m],$$

where  $\alpha_{N-1}^j = G_{N-1}(u^j)$  and  $u^j$  is the  $j$ th element of the control constraint set. Assume that

$$\bar{J}_{k+1}(P_{k+1}) = \min [P'_{k+1}\alpha_{k+1}^1, \dots, P'_{k+1}\alpha_{k+1}^{m_{k+1}}].$$

Then, using the expression from part (c) for  $\bar{J}_k(P_k)$ ,

$$\begin{aligned} \bar{J}_k(P_k) &= \min_u \left[ P'_k G_k(u) + \sum_{\theta=1}^q \bar{J}_{k+1}([r(u, \theta)] * [P(u)'P_k]) \right] \\ &= \min_u \left[ P'_k G_k(u) + \sum_{\theta=1}^q \min_{m=1, \dots, m_{k+1}} \left[ \{[r(u, \theta)] * [P(u)'P_k]\}' \alpha_{k+1}^m \right] \right] \\ &= \min_u \left[ P'_k G_k(u) + \sum_{\theta=1}^q \min_{m=1, \dots, m_{k+1}} [P'_k P(u) r(u, \theta)' \alpha_{k+1}^m] \right] \\ &= \min_u \left[ P'_k \left\{ G_k(u) + \sum_{\theta=1}^q \min_{m=1, \dots, m_{k+1}} [P(u) r(u, \theta)' \alpha_{k+1}^m] \right\} \right] \\ &= \min [P'_k \alpha_k^1, \dots, P'_k \alpha_k^{m_k}], \end{aligned}$$

where  $\alpha_k^1, \dots, \alpha_k^{m_k}$  are all possible vectors of the form

$$G_k(u) + \sum_{\theta=1}^q P(u) r(u, \theta)' \alpha_{k+1}^{m_{u, \theta}},$$

as  $u$  ranges over the finite set of controls,  $\theta$  ranges over the set of observation vector indexes  $\{1, \dots, q\}$ , and  $m_{u, \theta}$  ranges over the set of indexes  $\{1, \dots, m_{k+1}\}$ . The induction is thus complete.

For a quick way to understand the preceding proof, based on polyhedral concavity notions, note that the conclusion is equivalent to asserting that  $\bar{J}_k(P_k)$  is a positively homogeneous, concave polyhedral function. The preceding induction argument amounts to showing that the DP formula of part (c) preserves the positively homogeneous, concave polyhedral property of  $\bar{J}_{k+1}(P_{k+1})$ . This is indeed evident from the formula, since taking minima and nonnegative weighted sums of positively homogeneous, concave polyhedral functions results in a positively homogeneous, concave polyhedral function.

# Solutions Vol. I, Chapter 6

## 6.8 www

First, we notice that  $\alpha - \beta$  pruning is applicable only for arcs that point to *right* children, so that at least one sequence of moves (starting from the current position and ending at a terminal position, that is, one with no children) has been considered. Furthermore, due to depth-first search the score at the ancestor positions has been derived without taking into account the positions that can be reached from the current point. Suppose now that  $\alpha$ -pruning applies at a position with Black to play. Then, if the current position is reached (due to a move by White), Black can respond in such a way that the final position will be worse (for White) than it would have been if the current position were not reached. What  $\alpha$ -pruning saves is searching for even worse positions (emanating from the current position). The reason for this is that White will never play so that Black reaches the current position, because he certainly has a better alternative. A similar argument applies for  $\beta$  pruning.

*A second approach:* Let us suppose that it is the WHITE's turn to move. We shall prove that a  $\beta$ -cutoff occurring at the  $n$ th position will not affect the backed up score. We have from the definition of  $\beta$   $\beta = \min\{TBS \text{ of all ancestors of } n \text{ (white) where BLACK has the move}\}$ . For a cutoff to occur:  $TBS(n) > \beta$ . Observe first of all that  $\beta = TBS(n_1)$  for some ancestor  $n_1$  where BLACK has the move. Then there exists a path  $n_1, n_2, \dots, n_k, n$ . Since it is WHITE's move at  $n$  we have that  $TBS(n) = \max\{TBS(n), BS(n_i)\} > \beta$ , where  $n_i$  are the descendants of  $n$ . Consider now a position  $n_k$ . Then  $TBS(n_k)$  will either remain unchanged or will increase to a value greater than  $\beta$  as a result of the exploration of node  $n$ . Proceeding similarly, we conclude that  $TBS(n_2)$  will either remain the same or change to a value greater than  $\beta$ . Finally at node  $n_1$  we have that  $TBS(n_1)$  will not change since it is BLACK's turn to move and he will choose the move with minimum score. Thus the backed up score and the choice of the next move are unaffected from  $\beta$ -pruning. A similar argument holds for  $\alpha$ -pruning.

# Solutions Vol. I, Chapter 7

## 7.8 www

A threshold policy is specified by a threshold integer  $m$  and has the form

Process the orders if and only if their number exceeds  $m$ .

The cost function corresponding to a threshold policy specified by  $m$  will be denoted by  $J_m$ . By Prop. 3.1(c), this cost function is the unique solution of system of equations

$$J_m(i) = \begin{cases} K + \alpha(1-p)J_m(0) + \alpha p J_m(1) & \text{if } i > m, \\ ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1) & \text{if } i \leq m. \end{cases} \quad (1)$$

Thus for all  $i \leq m$ , we have

$$J_m(i) = \frac{ci + \alpha p J_m(i+1)}{1 - \alpha(1-p)},$$

$$J_m(i-1) = \frac{c(i-1) + \alpha p J_m(i)}{1 - \alpha(1-p)}.$$

From these two equations it follows that for all  $i \leq m$ , we have

$$J_m(i) \leq J_m(i+1) \quad \Rightarrow \quad J_m(i-1) < J_m(i). \quad (2)$$

Denote now

$$\gamma = K + \alpha(1-p)J_m(0) + \alpha p J_m(1).$$

Consider the policy iteration algorithm, and a policy  $\bar{\mu}$  that is the successor policy to the threshold policy corresponding to  $m$ . This policy has the form

Process the orders if and only if

$$K + \alpha(1-p)J_m(0) + \alpha p J_m(1) \leq ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1)$$

or equivalently

$$\gamma \leq ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1).$$

In order for this policy to be a threshold policy, we must have for all  $i$

$$\gamma \leq c(i-1) + \alpha(1-p)J_m(i-1) + \alpha p J_m(i) \quad \Rightarrow \quad \gamma \leq ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1). \quad (3)$$

This relation holds if the function  $J_m$  is monotonically nondecreasing, which from Eqs. (1) and (2) will be true if  $J_m(m) \leq J_m(m+1) = \gamma$ .

Let us assume that the opposite case holds, where  $\gamma < J_m(m)$ . For  $i > m$ , we have  $J_m(i) = \gamma$ , so that

$$ci + \alpha(1-p)J_m(i) + \alpha p J_m(i+1) = ci + \alpha\gamma. \quad (4)$$

We also have

$$J_m(m) = \frac{cm + \alpha p \gamma}{1 - \alpha(1 - p)},$$

from which, together with the hypothesis  $J_m(m) > \gamma$ , we obtain

$$cm + \alpha \gamma > \gamma. \quad (5)$$

Thus, from Eqs. (4) and (5) we have

$$ci + \alpha(1 - p)J_m(i) + \alpha p J_m(i + 1) > \gamma, \quad \text{for all } i > m, \quad (6)$$

so that Eq. (3) is satisfied for all  $i > m$ .

For  $i \leq m$ , we have  $ci + \alpha(1 - p)J_m(i) + \alpha p J_m(i + 1) = J_m(i)$ , so that the desired relation (3) takes the form

$$\gamma \leq J_m(i - 1) \Rightarrow \gamma \leq J_m(i). \quad (7)$$

To show that this relation holds for all  $i \leq m$ , we argue by contradiction. Suppose that for some  $i \leq m$  we have  $J_m(i) < \gamma \leq J_m(i - 1)$ . Then since  $J_m(m) > \gamma$ , there must exist some  $\bar{i} > i$  such that  $J_m(\bar{i} - 1) < J_m(\bar{i})$ . But then Eq. (2) would imply that  $J_m(j - 1) < J_m(j)$  for all  $j \leq \bar{i}$ , contradicting the relation  $J_m(i) < \gamma \leq J_m(i - 1)$  assumed earlier. Thus, Eq. (7) holds for all  $i \leq m$  so that Eq. (3) holds for all  $i$ . The proof is complete.

## 7.12 WWW

Let Assumption 2.1 hold and let  $\pi = \{\mu_0, \mu_1, \dots\}$  be an admissible policy. Consider also the sets  $S_k(i)$  given in the hint with  $S_0(i) = \{i\}$ . If  $t \in S_n(i)$  for all  $\pi$  and  $i$ , we are done. Otherwise, we must have for some  $\pi$  and  $i$ , and some  $k < n$ ,  $S_k(i) = S_{k+1}(i)$  while  $t \notin S_k(i)$ . For  $j \in S_k(i)$ , let  $m(j)$  be the smallest integer  $m$  such that  $j \in S_m$ . Consider a stationary policy  $\mu$  with  $\mu(j) = \mu_{m(j)}(j)$  for all  $j \in S_k(i)$ . For this policy we have for all  $j \in S_k(i)$ ,

$$p_{jl}(\mu(j)) > 0 \Rightarrow l \in S_k(i).$$

This implies that the termination state  $t$  is not reachable from all states in  $S_k(i)$  under the stationary policy  $\mu$ , and contradicts Assumption 2.1.

# Solutions Vol. II, Chapter 1

1.5

(a) We have

$$\begin{aligned}\sum_{j=1}^n \tilde{p}_{ij}(u) &= \sum_{j=1}^n \left\{ \frac{p_{ij}(u) - m_j}{1 - \sum_{k=1}^n m_k} \right\} \\ &= \frac{\sum_{j=1}^n p_{ij}(u) - \sum_{j=1}^n m_j}{1 - \sum_{k=1}^n m_k} \\ &= 1.\end{aligned}$$

Therefore,  $\tilde{p}_{ij}(u)$  are transition probabilities.

(b) We have for the modified problem

$$\begin{aligned}J'(i) &= \min_{u \in U(i)} \left[ g(i, u) + \alpha \left( 1 - \sum_{j=1}^n m_j \right) \sum_{j=1}^n \frac{p_{ij}(u) - m_j}{1 - \sum_{k=1}^n m_k} J'(j) \right] \\ &= \min_{u \in U(i)} \left[ g(i, u) + \alpha \sum_{j=1}^n p_{ij}(u) J'(j) - \alpha \sum_{k=1}^n m_k J'(k) \right].\end{aligned}$$

So

$$\begin{aligned}J'(i) + \frac{\alpha \sum_{k=1}^n m_k J'(k)}{1 - \alpha} &= \min_{u \in U(i)} \left[ g(i, u) + \alpha \sum_{j=1}^n p_{ij}(u) J'(j) - \alpha \sum_{k=1}^n m_k \underbrace{\left( 1 - \frac{1}{1 - \alpha} \right)}_{\frac{\alpha}{1 - \alpha}} J'(k) \right] \\ \Rightarrow J'(i) + \frac{\alpha \sum_{k=1}^n m_k J'(k)}{1 - \alpha} &= \min_{u \in U(i)} \left[ g(i, u) + \alpha \sum_{j=1}^n p_{ij}(u) \left( J'(j) + \frac{\alpha \sum_{k=1}^n m_k J'(k)}{1 - \alpha} \right) \right].\end{aligned}$$

Thus

$$J'(i) + \frac{\alpha \sum_{k=1}^n m_k J'(k)}{1 - \alpha} = J^*(i), \quad \forall i.$$

**Q.E.D.**

## 1.7

We show that for any bounded function  $J : S \rightarrow R$ , we have

$$J \leq T(J) \quad \Rightarrow \quad T(J) \leq F(J), \quad (1)$$

$$J \geq T(J) \quad \Rightarrow \quad T(J) \geq F(J). \quad (2)$$

For any  $\mu$ , define

$$F_\mu(J)(i) = \frac{g(i, \mu(i)) + \alpha \sum_{j \neq i} p_{ij}(\mu(i)) J(j)}{1 - \alpha p_{ii}(\mu(i))}$$

and note that

$$F_\mu(J)(i) = \frac{T_\mu(J)(i) - \alpha p_{ii}(\mu(i)) J(i)}{1 - \alpha p_{ii}(\mu(i))}. \quad (3)$$

Fix  $\epsilon > 0$ . If  $J \leq T(J)$ , let  $\mu$  be such that  $F_\mu(J) \leq F(J) + \epsilon e$ . Then, using Eq. (3),

$$F(J)(i) + \epsilon \geq F_\mu(J)(i) = \frac{T_\mu(J)(i) - \alpha p_{ii}(\mu(i)) J(i)}{1 - \alpha p_{ii}(\mu(i))} \geq \frac{T(J)(i) - \alpha p_{ii}(\mu(i)) T(J)(i)}{1 - \alpha p_{ii}(\mu(i))} = T(J)(i).$$

Since  $\epsilon > 0$  is arbitrary, we obtain  $F(J)(i) \geq T(J)(i)$ . Similarly, if  $J \geq T(J)$ , let  $\mu$  be such that  $T_\mu(J) \leq T(J) + \epsilon e$ . Then, using Eq. (3),

$$F(J)(i) \leq F_\mu(J)(i) = \frac{T_\mu(J)(i) - \alpha p_{ii}(\mu(i)) J(i)}{1 - \alpha p_{ii}(\mu(i))} \leq \frac{T(J)(i) + \epsilon - \alpha p_{ii}(\mu(i)) T(J)(i)}{1 - \alpha p_{ii}(\mu(i))} \leq T(J)(i) + \frac{\epsilon}{1 - \alpha}.$$

Since  $\epsilon > 0$  is arbitrary, we obtain  $F(J)(i) \leq T(J)(i)$ .

From (1) and (2) we see that  $F$  and  $T$  have the same fixed points, so  $J^*$  is the unique fixed point of  $F$ . Using the definition of  $F$ , it can be seen that for any scalar  $r > 0$  we have

$$F(J + re) \leq F(J) + \alpha re, \quad F(J) - \alpha re \leq F(J - re). \quad (4)$$

Furthermore,  $F$  is monotone, that is

$$J \leq J' \quad \Rightarrow \quad F(J) \leq F(J'). \quad (5)$$

For any bounded function  $J$ , let  $r > 0$  be such that

$$J - re \leq J^* \leq J + re.$$

Applying  $F$  repeatedly to this equation and using Eqs. (4) and (5), we obtain

$$F^k(J) - \alpha^k re \leq J^* \leq F^k(J) + \alpha^k re.$$

Therefore  $F^k(J)$  converges to  $J^*$ . From Eqs. (1), (2), and (5) we see that

$$J \leq T(J) \quad \Rightarrow \quad T^k(J) \leq F^k(J) \leq J^*,$$

$$J \geq T(J) \quad \Rightarrow \quad T^k(J) \geq F^k(J) \geq J^*.$$

These equations demonstrate the faster convergence property of  $F$  over  $T$ .

As a final result (not explicitly required in the problem statement), we show that for any two bounded functions  $J : S \rightarrow R$ ,  $J' : S \rightarrow R$ , we have

$$\max_j |F(J)(j) - F(J')(j)| \leq \alpha \max_j |J(j) - J'(j)|, \quad (6)$$

so  $F$  is a contraction mapping with modulus  $\alpha$ . Indeed, we have

$$\begin{aligned} F(J)(i) &= \min_{u \in U(i)} \left\{ \frac{g(i, u) + \alpha \sum_{j \neq i} p_{ij}(u) J(j)}{1 - \alpha p_{ii}(u)} \right\} \\ &= \min_{u \in U(i)} \left\{ \frac{g(i, u) + \alpha \sum_{j \neq i} p_{ij}(u) J'(j)}{1 - \alpha p_{ii}(u)} + \frac{\alpha \sum_{j \neq i} p_{ij}(u) [J(j) - J'(j)]}{1 - \alpha p_{ii}(u)} \right\} \\ &\leq F(J')(i) + \max_j |J(j) - J'(j)|, \quad \forall i, \end{aligned}$$

where we have used the fact

$$1 - \alpha p_{ii}(u) \geq 1 - p_{ii}(u) = \sum_{j \neq i} p_{ij}(u).$$

Thus, we have

$$F(J)(i) - F(J')(i) \leq \alpha \max_j |J(j) - J'(j)|, \quad \forall i.$$

The roles of  $J$  and  $J'$  may be reversed, so we can also obtain

$$F(J')(i) - F(J)(i) \leq \alpha \max_j |J(j) - J'(j)|, \quad \forall i.$$

Combining the last two inequalities, we see that

$$|F(J)(i) - F(J')(i)| \leq \alpha \max_j |J(j) - J'(j)|, \quad \forall i.$$

By taking the maximum over  $i$ , Eq. (6) follows.

## 1.9

(a) Since  $J, J' \in B(S)$ , i.e., are real-valued, bounded functions on  $S$ , we know that the infimum and the supremum of their difference is finite. We shall denote

$$m = \min_{x \in S} (J(x) - J'(x))$$

and

$$M = \max_{x \in S} (J(x) - J'(x)).$$



Thus

$$m \leq J(x) - J'(x) \leq M, \quad \forall x \in S,$$

or

$$J'(x) + m \leq J(x) \leq J'(x) + M, \quad \forall x \in S.$$

Now we apply the mapping  $T$  on the above inequalities. By property (1) we know that  $T$  will preserve the inequalities. Thus

$$T(J' + me)(x) \leq T(J)(x) \leq T(J' + Me)(x), \quad \forall x \in S.$$

By property (2) we know that

$$T(J)(x) + \min[a_1 r, a_2 r] \leq T(J + re)(x) \leq T(J)(x) + \max[a_1 r, a_2 r].$$

If we replace  $r$  by  $m$  or  $M$ , we get the inequalities

$$T(J')(x) + \min[a_1 m, a_2 m] \leq T(J' + me)(x) \leq T(J')(x) + \max[a_1 m, a_2 m]$$

and

$$T(J')(x) + \min[a_1 M, a_2 M] \leq T(J' + Me)(x) \leq T(J')(x) + \max[a_1 M, a_2 M].$$

Thus

$$T(J')(x) + \min[a_1 m, a_2 m] \leq T(J)(x) \leq T(J')(x) + \max[a_1 M, a_2 M],$$

so that

$$|T(J)(x) - T(J')(x)| \leq \max[a_1 |M|, a_2 |M|, a_1 |m|, a_2 |m|].$$

We also have

$$\max[a_1 |M|, a_2 |M|, a_1 |m|, a_2 |m|] \leq a_2 \max[|M|, |m|] \leq a_2 \sup_{x \in S} |J(x) - J'(x)|.$$

Thus

$$|T(J)(x) - T(J')(x)| \leq a_2 \max_{x \in S} |J(x) - J'(x)|$$

from which

$$\max_{x \in S} |T(J)(x) - T(J')(x)| \leq a_2 \max_{x \in S} |J(x) - J'(x)|.$$

Thus  $T$  is a contraction mapping since we know by the statement of the problem that  $0 \leq a_1 \leq a_2 < 1$ .

Since the set  $B(S)$  of bounded real valued functions is a complete linear space, we conclude that the contraction mapping  $T$  has a unique fixed point,  $J^*$ , and  $\lim_{k \rightarrow \infty} T^k(J)(x) = J^*(x)$ .

(b) We shall first prove the lower bounds of  $J^*(x)$ . The upper bounds follow by a similar argument. Since  $J, T(J) \in B(S)$ , there exists a  $c \in \mathfrak{R}$ , ( $c < \infty$ ), such that

$$J(x) + c \leq T(J)(x). \tag{1}$$

We apply  $T$  on both sides of (1) and since  $T$  preserves the inequalities (by assumption (1)) we have by applying the relation of assumption (2).

$$J(x) + \min[c + a_1 c, c + a_2 c] \leq T(J)(x) + \min[a_1 c, a_2 c] \leq T(J + ce)(x) \leq T^2(J)(x). \quad (2)$$

Similarly, if we apply  $T$  again we get,

$$\begin{aligned} J(x) + \min_{i \in (1,2)} [c + a_i c, c + a_i^2 c] &\leq T(J) + \min[a_1 c + a_1^2 c, a_2 c + a_2^2 c] \\ &\leq T^2(J) + \min[a_1^2 c, a_2^2 c] \leq T(T(J) + \min[a_1 c, a_2 c]e)(x) \leq T^3(J)(x). \end{aligned}$$

Thus by induction we conclude

$$\begin{aligned} J(x) + \min\left[\sum_{m=0}^k a_1^m c, \sum_{m=0}^k a_2^m c\right] &\leq T(J)(x) + \min\left[\sum_{m=1}^k a_1^m c, \sum_{m=1}^k a_2^m c\right] \leq \dots \\ &\leq T^k(J)(x) + \min[a_1^k c, a_2^k c] \leq T^{k+1}(J)(x). \end{aligned} \quad (3)$$

By taking the limit as  $k \rightarrow \infty$  and noting that the quantities in the minimization are monotone, and either nonnegative or nonpositive, we conclude that

$$\begin{aligned} J(x) + \min\left[\frac{1}{1-a_1}c, \frac{1}{1-a_2}c\right] &\leq T(J)(x) + \min\left[\frac{a_1}{1-a_1}c, \frac{a_2}{1-a_2}c\right] \\ &\leq T^k(J)(x) + \min\left[\frac{a_1^k}{1-a_1}c, \frac{a_2^k}{1-a_2}c\right] \\ &\leq T^{k+1}(J)(x) + \min\left[\frac{a_1^{k+1}}{1-a_1}c, \frac{a_2^{k+1}}{1-a_2}c\right] \\ &\leq J^*(x). \end{aligned} \quad (4)$$

Finally we note that

$$\min[a_1^k c, a_2^k c] \leq T^{k+1}(J)(x) - T^k(J)(x).$$

Thus

$$\min[a_1^k c, a_2^k c] \leq \inf_{x \in S} (T^{k+1}(J)(x) - T^k(J)(x)).$$

Let  $b_{k+1} = \inf_{x \in S} (T^{k+1}(J)(x) - T^k(J)(x))$ . Thus  $\min[a_1^k c, a_2^k c] \leq b_{k+1}$ . From the above relation we infer that

$$\min\left[\frac{a_1^{k+1}c}{1-a_1}, \frac{a_2^{k+1}c}{1-a_2}\right] \leq \min\left[\frac{a_1}{1-a_1}b_{k+1}, \frac{a_2}{1-a_2}b_{k+1}\right] = c_{k+1}$$

Therefore

$$T^k(J)(x) + \min\left[\frac{a_1^k c}{1-a_1}, \frac{a_2^k c}{1-a_2}\right] \leq T^{k+1}(J)(x) + c_{k+1}.$$

This relationship gives for  $k = 1$

$$T(J)(x) + \min\left[\frac{a_1 c}{1-a_1}, \frac{a_2 c}{1-a_2}\right] \leq T^2(J)(x) + c_2$$

Let

$$c = \inf_{x \in S} (T(J)(x) - J(x))$$

Then the above inequality still holds. From the definition of  $c_1$  we have

$$c_1 = \min \left[ \frac{a_1 c}{1 - a_1}, \frac{a_2 c}{1 - a_2} \right].$$

Therefore

$$T(J)(x) + c_1 \leq T^2(J)(x) + c_2$$

and  $T(J)(x) + c_1 \leq J^*(x)$  from Eq. (4). Similarly, let  $J_1(x) = T(J)(x)$ , and let

$$b_2 = \min_{x \in S} (T^2(J)(x) - T(J)(x)) = \min_{x \in S} (T(J_1)(x) - T(J_1)(x)).$$

If we proceed as before, we get

$$\begin{aligned} J_1(x) + \min \left[ \frac{1}{1 - a_3} b_2, \frac{1}{1 - a_2} b_2 \right] &\leq T(J_1)(x) + \min \left[ \frac{a_1 b_2}{1 - a_2}, \frac{a_1 b_2}{1 - a_2} \right] \\ &\leq T^2(J_1)(x) + \min \left[ \frac{a_1^2 b_2}{1 - a_2}, \frac{a_2^2 b_2}{1 - a_2} \right] \leq J^*(x). \end{aligned}$$

Then

$$\min[a_1 b_2, a_2 b_2] \leq \min_{x \in S} [T^2(J_1)(x) - T(J_1)(x)] = \min_{x \in S} [T^3(J)(x) - T^2(J)(x)] = b_3$$

Thus

$$\min \left[ \frac{a_1^2 b_2}{1 - a_1}, \frac{a_2^2 b_2}{1 - a_2} \right] \leq \min \left[ \frac{a_1 b_3}{1 - a_2}, \frac{a_2 b_3}{1 - a_2} \right].$$

Thus

$$T(J_1)(x) + \min \left[ \frac{a_1 b_2}{1 - a_2}, \frac{a_2 b_2}{1 - a_2} \right] \leq T^2(J_1)(x) + \min \left[ \frac{a_1 b_3}{1 - a_2}, \frac{a_2 b_2}{1 - a_2} \right]$$

or

$$T^2(J)(x) + c_2 \leq T^3(J)(x) + c_3$$

and

$$T^2(J)(x) + c_2 \leq J^*(x).$$

Proceeding similarly the result is proved.

The reverse inequalities can be proved by a similar argument.

(c) Let us first consider the state  $x = 1$

$$F(J)(1) = \min_{u \in U(1)} \left\{ g(j, j) + a \sum_{j=1}^n p_{1j} J(j) \right\}$$

Thus

$$\begin{aligned} F(J + re)(1) &= \min_{u \in U(1)} \left\{ g(1, u) + \alpha \sum_{j=1}^n p_{ij}(J + re)(j) \right\} = \min_{u \in U(1)} \left\{ g(1, u) + \alpha \sum_{j=1}^n p_{1j}J(j) + \alpha r \right\} \\ &= F(J)(1) + \alpha r \end{aligned}$$

Thus

$$\frac{F(J + re)(1) - F(J)(1)}{r} = \alpha \quad (1)$$

Since  $0 \leq \alpha \leq 1$  we conclude that  $\alpha^n \leq \alpha$ . Thus

$$\alpha^n \leq \frac{F(J + re)(1) - F(J)(1)}{r} = \alpha$$

For the state  $x = 2$  we proceed similarly and we get

$$F(J)(2) = \min_{u \in U(2)} \left\{ g(2, u) + \alpha p_{21}F(J)(1) + \alpha \sum_{j=2}^n p_{2j}J(j) \right\}$$

and

$$\begin{aligned} F(J + re)(2) &= \min_{u \in U(2)} \left\{ g(2, u) + \alpha p_{21}F(J + re)(1) + \alpha \sum_{j=2}^n p_{2j}(J + re)(j) \right\} \\ &= \min_{u \in U(2)} \left\{ g(2, u) + \alpha p_{21}F(J)(1) + \alpha^2 r p_{21} + \alpha \sum_{j=2}^n p_{2j}J(j) + \alpha \sum_{j=2}^n p_{ij}re(j) \right\} \end{aligned}$$

where, for the last equality, we used relation (1).

Thus we conclude

$$F(J + re)(2) = F(J)(2) + \alpha^2 r p_{21} + \alpha \sum_{j=2}^n p_{2j}r = F(J)(2) + \alpha^2 r p_{21} + \alpha r(1 - p_{21})$$

which yields

$$\frac{F(J + re)(2) - F(J)(2)}{r} = \alpha^2 p_{21} + \alpha(1 - p_{21}) \quad (2)$$

Now let us study the behavior of the right-hand side of Eq. (2). We have  $0 < \alpha < 1$  and  $0 < p_{21} < 1$ , so since  $\alpha^2 \leq \alpha$ , and  $\alpha^2 p_{21} + \alpha(1 - p_{21})$  is a convex combination of  $\alpha^2$ ,  $\alpha$ , it is easy to see that

$$\alpha^2 \leq \alpha^2 p_{21} + (1 - p_{21})\alpha \leq \alpha \quad (3)$$

If we combine Eq. (2) with Eq. (3) we get

$$\alpha^n \leq \alpha^2 \leq \frac{F(J + re)(2) - F(J)(2)}{r} \leq \alpha$$

which is the pursued result.

**Claim:**

$$\alpha^i \leq \frac{F(J+re)(x) - F(J)(x)}{r} \leq \alpha$$

**Proof:** We shall employ an inductive argument. Obviously the result holds for  $x = 1, 2$ . Let us assume that it holds for all  $x \leq i$ . We shall prove it for  $x = i + j$

$$F(J)(i+1) = \min_{u \in U(i+1)} \left\{ g(i+1, u) + \alpha \sum_{j=1}^i p_{1+ij} F(J)(j) + \alpha \sum_{j=i+1}^n p_{i+1j} p_{i+1j} J(j) \right\}$$

$$F(J+re)(i+1) = \min_{u \in U(i+1)} \left\{ g(i+1, u) + \alpha \sum_{j=1}^i p_{i+1j} F(J+re)(j) + \alpha \sum_{j=i+1}^n p_{i+1j} (J+re)(j) \right\}$$

We know  $\alpha^j \leq F(J+re)(j) \leq \alpha, \forall j \leq i$ , thus

$$F(J)(i+1) + r\alpha \sum_{j=1}^i F(J)(i+1) + \alpha^2 rp + \alpha r(1-p)$$

where

$$p = \sum_{j=1}^i p_{1+ij}$$

Obviously

$$\sum_{j=1}^i \alpha^j p_{i+1j} \geq \alpha^i \sum_{j=1}^i p_{i+1j} = \alpha^i p$$

Thus

$$\alpha^{i+1}p + \alpha(1-p) \leq \frac{F(J+re)(j) - F(J)(j)}{r} \leq \alpha^2p + (1-p)\alpha$$

Since  $0 < \alpha^{i+1} \leq \alpha^2 \leq \alpha < 1$  and  $0 \leq p \leq i$  we conclude that  $\alpha^{i+1} \leq \alpha^{i+1}p + \alpha(1-p)$  and  $\alpha^2p + (1-p)\alpha \leq \alpha$ . Thus

$$\alpha^{i+1} \leq \frac{F(J+re)(i+1) - F(J)(i+1)}{r} \leq \alpha$$

which completes the inductive proof.

Since  $0 \leq \alpha^n \leq \alpha^i \leq 1$  for  $i \leq i \leq n$ , the result follows.

(d) Let  $J(x) \leq J'(x) (=) J'(x) - J(x) \geq 0$  Since all the elements  $m_{ij}$  are non-negative we conclude that

$$M(J'(x) - J(x)) \geq 0 (=) MJ'(x) \geq MJ(x)$$

$$g(x) + MJ'(x) \geq g(x) + MJ(x)$$

$$T(J')(x) \geq T(J)(x)$$

thus property (1) holds.

For property (2) we note that

$$T(J + re)(x) = g(x) + M(J + re)(x) = g(x) + MJ(x) + rMe(x) = T(J)(x) + rMe(x)$$

We have

$$\alpha_1 \leq Me(x) \leq \alpha_2$$

so that

$$\frac{T(J + re)(x) - T(J)(x)}{r} = Me(x)$$

and

$$\alpha_1 \leq \frac{T(J + re)(x) - T(J)(x)}{r} \leq \alpha_2$$

Thus property (2) also holds if  $\alpha_2 < 1$ .

### 1.10

(a) If there is a unique  $\mu$  such that  $T_\mu(J) = T(J)$ , then there exists an  $\epsilon > 0$  such that for all  $\Delta \in \mathcal{R}^n$  with  $\max_i |\Delta(i)| \leq \epsilon$  we have

$$F(J + \Delta) = T(J + \Delta) - J - \Delta = g_\mu + \alpha P_\mu(J + \Delta) - J - \Delta = g_\mu + (\alpha P_\mu - I)(J + \Delta).$$

It follows that  $F$  is linear around  $J$  and its Jacobian is  $\alpha P_\mu - I$ .

(b) We first note that the equation defining Newton's method is the first order Taylor series expansion of  $F$  around  $J_k$ . If  $\mu^k$  is the unique  $\mu$  such that  $T_\mu(J_k) = T(J_k)$ , then  $F$  is linear near  $J_k$  and coincides with its first order Taylor series expansion around  $J_k$ . Therefore the vector  $J_{k+1}$  is obtained by the Newton iteration satisfies

$$F(J_{k+1}) = 0$$

or

$$T_{\mu^k}(J_{k+1}) = J_{k+1}.$$

This equation yields  $J_{k+1} = J_{\mu^k}$ , so the next policy  $\mu^{k+1}$  is obtained as

$$\mu^{k+1} = \arg \min_{\mu} T_\mu(J_{\mu^k}).$$

This is precisely the policy iteration of the algorithm.

### 1.12

For simplicity, we consider the case where  $U(i)$  consists of a single control. The calculations are very similar for the more general case. We first show that  $\sum_{j=1}^n \bar{M}_{ij} = \alpha$ . We apply the definition of the quantities  $\bar{M}_{ij}$

$$\begin{aligned} \sum_{j=1}^n \bar{M}_{ij} &= \sum_{j=1}^n \left( \delta_{ij} + \frac{(1-\alpha)(M_{ij} - \delta_{ij})}{1-m_i} \right) = \sum_{j=1}^n \delta_{ij} + \sum_{j=1}^n \frac{(1-\alpha)(M_{ij} - \delta_{ij})}{1-m_i} \\ &= 1 + (1-\alpha) \sum_{j=1}^n \frac{M_{ij}}{1-m_i} - \frac{(1-\alpha)}{1-m_i} \sum_{j=1}^n \delta_{ij} = 1 + (1-\alpha) \frac{m_i}{1-m_i} - \frac{(1-\alpha)}{1-m_i} \\ &= 1 - (1-\alpha) = \alpha. \end{aligned}$$

Let  $J_1^*, \dots, J_n^*$  satisfy

$$J_i^* = g_i + \sum_{j=1}^n M_{ij} J_j^*. \quad (1)$$

We substitute  $J^*$  into the new equation

$$J_i^* = \bar{g}_i + \sum_{j=1}^n \bar{M}_{ij} J_j^*$$

and manipulate the equation until we reach a relation that holds trivially

$$\begin{aligned} J_1^* &= \frac{g_i(1-\alpha)}{1-m_i} + \sum_{j=1}^n \delta_{ij} J_j^* + \frac{1-\alpha}{1-m_i} \sum_{j=1}^n (M_{ij} - \delta_{ij}) J_j^* \\ &= \frac{g_i(1-\alpha)}{1-m_i} + J_i^* + \frac{1-\alpha}{1-m_i} \sum_{j=1}^n M_{ij} J_j^* - \frac{1-\alpha}{1-m_i} J_i^* \\ &= J_i^* + \frac{1-\alpha}{1-m_i} \left( g_i + \sum_{j=1}^n M_{ij} J_j^* - J_i^* \right). \end{aligned}$$

This relation follows trivially from Eq. (1) above. Thus  $J^*$  is a solution of

$$J_i = \bar{g}_i + \sum_{j=1}^n \bar{M}_{ij} J_j.$$

### 1.17

The form of Bellman's Equation for the tax problem is

$$J(x) = \min_i \left[ \sum_{j \neq i} c^j(x^i) + \alpha E_{w^i} \{ J[x^i, x^{i-1}, f^i(x^i, w^i)] \} \right]$$

Let  $\bar{J}(x) = -J(x)$

$$\bar{J}(x) = \max_i \left[ - \sum_{j=1}^n c^j(x^j) + c^i(x^i) + \alpha E_{w^i} \{ \bar{J}[\cdot \cdot] \} \right]$$

Let  $\tilde{J}(x) = (1 - \alpha)\bar{J}(x) + \sum_{j=1}^n C^j(x^j)$  By substitution we obtain

$$\begin{aligned} \tilde{J}(x) &= \max_i \left[ -(1 - \alpha) \sum_{j=1}^n c^j(x^j) + (1 - \alpha)c^i(x^i) + \alpha E_{w^i} \{ (1 - \alpha)\bar{J}[\cdot \cdot] \} \right] \\ &= \max_i [c^i(x^i) - \alpha E_{w^i} \{ c^i(f(x^i, w^i)) \}] + \alpha E_{w^i} \{ \tilde{J}(\cdot \cdot) \}. \end{aligned}$$

Thus  $\tilde{J}$  satisfies Bellman's Equation of a multi-armed Bandit problem with

$$R_i(x^i) = c^i(x^i) - \alpha E_{w^i} \{ c^i(f(x^i, w^i)) \}.$$

### 1.18

Bellman's Equation for the restart problem is

$$J(x) = \max[R(x_0) + \alpha E\{J[f(x_0, w)]\}, R(x) + \alpha E\{J[f(x, w)]\}]. \quad (A)$$

Now, consider the one-armed bandit problem with reward  $R(x)$

$$J(x, M) = \max\{M, R(x) + \alpha E[J(f(x, w), M)]\}. \quad (B)$$

We have

$$J(x_0, M) = R(x_0) + \alpha E[J(f(x_0, w), M)] > M$$

if  $M < m(x_0)$  and  $J(x_0, M) = M$ . This implies that

$$R(x_0) + \alpha E[J(f(x_0, w))] = m(x_0).$$

Therefore the forms of both Bellman's Equations (A) and (B) are the same when  $M = m(x_0)$ .



**7.28** www

- (a) This follows from the nonnegativity of the one-stage cost.
- (b) Follow the hint.
- (c) Take the limit as  $\alpha \rightarrow 1$  in Bellman's equation for a discounted cost.
- (d) Follow the hint.
- (e) Follow the hint.

# Solutions Vol. II, Chapter 2

## 2.2

Let's define the following states:

$H$ : Last flip outcome was heads

$T$ : Last flip outcome was tails

$C$ : Caught (this is the termination state)

(a) We can formulate this problem as a stochastic shortest path problem with state  $C$  being the termination state. There are four possible policies:  $\pi_1 = \{\text{always flip fair coin}\}$ ,  $\pi_2 = \{\text{always flip two-headed coin}\}$ ,  $\pi_3 = \{\text{flip fair coin if last outcome was heads / flip two-headed coin if last outcome was tails}\}$ , and  $\pi_4 = \{\text{flip fair coin if last outcome was tails / flip two-headed coin if last outcome was heads}\}$ . The only way to reach the termination state is to be caught cheating. Under all policies except  $\pi_1$ , this is inevitable. Thus  $\pi_1$  is an improper policy, and  $\pi_2, \pi_3$ , and  $\pi_4$  are proper policies.

(b) Let  $J_{\pi_1}(H)$  and  $J_{\pi_1}(T)$  be the costs corresponding policy  $\pi_1$  where the starting state is  $H$  and  $T$ , respectively. The expected benefit starting from state  $T$  up to the first return to  $T$  (and always using the fair coin), is

$$\frac{1}{2} \left( 1 + \frac{1}{2} + \frac{1}{2^2} + \cdots \right) - \frac{m}{2} = \frac{1}{2}(2 - m).$$

Therefore

$$J_{\pi_1}(T) = \begin{cases} +\infty & \text{if } m < 2 \\ 0 & \text{if } m = 2 \\ -\infty & \text{if } m > 2. \end{cases}$$

Also we have

$$J_{\pi_1}(H) = \frac{1}{2}(1 + J_n(H)) + \frac{1}{2}J_n(T),$$

so

$$J_{\pi_1}(H) = 1 + J_{\pi}(T).$$

It follows that if  $m > 2$ , then  $\pi_1$  results in infinite cost for any initial state.

(c,d) The expected one-stage rewards at each stage are

Play Fair in State  $H$ :  $\frac{1}{2}$

Cheat in State  $H$ :  $1 - p$

Play Fair in State  $T$ :  $\frac{1-m}{2}$

Cheat in State  $T$ : 0

We show that any policy that cheats at  $H$  at some stage cannot be optimal. As a result we can eliminate cheating from the control constraint set of state  $H$ .

Indeed suppose we are at state  $H$  at some stage and consider a policy  $\hat{\pi}$  which cheats at the first stage and then follows the optimal policy  $\pi^*$  from the second stage on. Consider a policy  $\tilde{\pi}$  which plays fair at the first stage, and then follows  $\pi^*$  from the second stage on if the outcome of the first stage is  $H$  or cheats at the second stage and follows  $\pi^*$  from the third stage on if the outcome of the first stage is  $T$ . We have

$$J_{\hat{\pi}}(H) = (1 - p)[1 + J_{\pi^*}(H)]$$

$$\begin{aligned}
J_{\hat{\pi}}(H) &= \frac{1}{2}(1 + J_{\pi^*}(H)) + \frac{1}{2}\{(1-p)[1 + J_{\pi^*}(H)]\} \\
&= \frac{1}{2} + \frac{1}{2}[J_{\pi^*}(H) + J_{\hat{\pi}}(H)] \geq \frac{1}{2} + J_{\hat{\pi}}(H),
\end{aligned}$$

where the inequality follows from the fact that  $J_{\pi^*}(H) \geq J_{\hat{\pi}}(H)$  since  $\pi^*$  is optimal. Therefore the reward of policy  $\hat{\pi}$  can be improved by at least  $\frac{1}{2}$  by switching to policy  $\tilde{\pi}$ , and therefore  $\hat{\pi}$  cannot be optimal.

We now need only consider policies in which the gambler can only play fair at state  $H$ :  $\pi_1$  and  $\pi_3$ . Under  $\pi_1$ , we saw from part b) that the expected benefits are

$$J_{\pi_1}(T) = \begin{cases} +\infty & \text{if } m < 2 \\ 0 & \text{if } m = 2 \\ -\infty & \text{if } m > 2, \end{cases}$$

and

$$J_{\pi_1}(H) = \begin{cases} +\infty & \text{if } m < 2 \\ 1 & \text{if } m = 2 \\ -\infty & \text{if } m > 2. \end{cases}$$

Under  $\pi_3$ , we have

$$\begin{aligned}
J_{\pi_3}(T) &= (1-p)J_{\pi_3}(H), \\
J_{\pi_3}(H) &= \frac{1}{2}[1 + J_{\pi_3}(H)] + \frac{1}{2}J_{\pi_3}(T).
\end{aligned}$$

Solving these two equations yields

$$\begin{aligned}
J_{\pi_3}(T) &= \frac{1-p}{p}, \\
J_{\pi_3}(H) &= \frac{1}{p}.
\end{aligned}$$

Thus if  $m > 2$ , it is optimal to cheat if the last flip was tails and play fair otherwise, and if  $m < 2$ , it is optimal to always play fair.

## 2.7

(a) Let  $i$  be any state in  $S_m$ . Then,

$$\begin{aligned}
J(i) &= \min_{u \in U(i)} [E\{g(i, u, j) + J(j)\}] \\
&= \min_{u \in U(i)} \left[ \sum_{j \in S_m} p_{ij}(u)[g(i, u, j) + J(j)] + \sum_{j \in S_{m-1} \cup \dots \cup S_1 \cup t} p_{ij}(u)[g(i, u, j) + J(j)] \right] \\
&= \min_{u \in U(i)} \left[ \sum_{j \in S_m} p_{ij}(u)[g(i, u, j) + J(j)] + (1 - \sum_{j \in S_m} p_{ij}(u)) \frac{\sum_{j \in S_{m-1} \cup \dots \cup S_1 \cup t} p_{ij}(u)[g(i, u, j) + J(j)]}{(1 - \sum_{j \in S_m} p_{ij}(u))} \right].
\end{aligned}$$

In the above equation, we can think of the union of  $S_{m-1}, \dots, S_1$ , and  $t$  as an aggregate termination state  $t_m$  associated with  $S_m$ . The probability of a transition from  $i \in S_m$  to  $t_m$  (under  $u$ ) is given by,

$$p_{it_m}(u) = 1 - \sum_{j \in S_m} p_{ij}(u).$$

The corresponding cost of a transition from  $i \in S_m$  to  $t_m$  (under  $u$ ) is given by,

$$\tilde{g}(i, u, t_m) = \frac{\sum_{j=S_{m-1} \cup \dots \cup S_1 \cup t} p_{ij}(u)[g(i, u, j) + J(j)]}{p_{it_m}(u)}.$$

Thus, for  $i \in S_m$ , Bellman's equation can be written as,

$$J(i) = \min_{u \in U(i)} \left[ \sum_{j \in S_m} p_{ij}(u)[g(i, u, j) + J(j)] + p_{it_m}(u)[\tilde{g}(i, u, t_m) + 0] \right].$$

Note that with respect to  $S_m$ , the termination state  $t_m$  is both absorbing and of zero cost. Let  $t_m$  and  $\tilde{g}(i, u, t_m)$  be similarly constructed for  $m = 1, \dots, M$ .

The original stochastic shortest path problem can be solved as  $M$  stochastic shortest path sub-problems. To see how, start with evaluating  $J(i)$  for  $i \in S_1$  (where  $t_1 = \{t\}$ ). With the values of  $J(i)$ , for  $i \in S_1$ , in hand, the  $\tilde{g}$  cost-terms for the  $S_2$  problem can be computed. The solution of the original problem continues in this manner as the solution of  $M$  stochastic shortest path problems in succession.

(b) Suppose that in the finite horizon problem there are  $\tilde{n}$  states. Define a new state space  $S_{new}$  and sets  $S_m$  as follows,

$$S_{new} = \{(k, i) | k \in \{0, 1, \dots, M-1\} \text{ and } i \in \{1, 2, \dots, \tilde{n}\}\}$$

$$S_m = \{(k, i) | k = M - m \text{ and } i \in \{1, 2, \dots, \tilde{n}\}\}$$

for  $m = 1, 2, \dots, M$ . (Note that the  $S_m$ 's do not overlap.) By associating  $S_m$  with the state space of the original finite-horizon problem at stage  $k = M - m$ , we see that if  $i_k \in S_{m-1}$  under all policies. By augmenting a termination state  $t$  which is absorbing and of zero cost, we see that the original finite-horizon problem can be cast as a stochastic shortest path problem with the special structure indicated in the problem statement.

## 2.8

Let  $J^*$  be the optimal cost of the original problem and  $\tilde{J}$  be the optimal cost of the modified problem. Then we have

$$J^*(i) = \min_u \sum_{j=1}^n p_{ij}(u) (g(i, u, j) + J^*(j)),$$

and

$$\tilde{J}(i) = \min_u \sum_{j=1, j \neq i}^n \frac{p_{ij}(u)}{1 - p_{ii}(u)} \left( g(i, u, j) + \frac{g(i, u, i)p_{ii}(u)}{1 - p_{ii}(u)} + \tilde{J}(j) \right).$$

For each  $i$ , let  $\mu^*(i)$  be a control such that

$$J^*(i) = \sum_{j=1}^n p_{ij}(\mu^*(i)) (g(i, \mu^*(i), j) + J^*(j)).$$

Then

$$J^*(i) = \left[ \sum_{j=1, j \neq i}^n p_{ij}(\mu^*(i)) (g(i, \mu^*(i), j) + J^*(j)) \right] + p_{ii}(\mu^*(i)) (g(i, \mu^*(i), i) + J^*(i)).$$

By collecting the terms involving  $J^*(i)$  and then dividing by  $1 - p_{ii}(\mu^*(i))$ ,

$$J^*(i) = \frac{1}{1 - p_{ii}(\mu^*(i))} \left\{ \left[ \sum_{j=1, j \neq i}^n p_{ij}(\mu^*(i)) (g(i, \mu^*(i), j) + J^*(j)) \right] + p_{ii}(\mu^*(i)) g(i, \mu^*(i), i) \right\}.$$

Since  $\sum_{j=1, j \neq i}^n \frac{p_{ij}(\mu^*(i))}{1 - p_{ii}(\mu^*(i))} = 1$ , we have

$$\begin{aligned} J^*(i) &= \frac{1}{1 - p_{ii}(\mu^*(i))} \left\{ \left[ \sum_{j=1, j \neq i}^n p_{ij}(\mu^*(i)) (g(i, \mu^*(i), j) + J^*(j)) \right] + \sum_{j=1, j \neq i}^n \frac{p_{ij}(\mu^*(i))}{1 - p_{ii}(\mu^*(i))} p_{ii}(\mu^*(i)) g(i, \mu^*(i), i) \right\} \\ &= \sum_{j=1, j \neq i}^n \left[ \frac{p_{ij}(\mu^*(i))}{1 - p_{ii}(\mu^*(i))} (g(i, \mu^*(i), j) + J^*(j)) + \frac{p_{ii}(\mu^*(i)) g(i, \mu^*(i), i)}{1 - p_{ii}(\mu^*(i))} \right]. \end{aligned}$$

Therefore  $J^*(i)$  is the cost of stationary policy  $\{\mu^*, \mu^*, \dots\}$  in the modified problem. Thus

$$J^*(i) \geq \tilde{J}(i) \quad \forall i.$$

Similarly, for each  $i$ , let  $\tilde{\mu}(i)$  be a control such that

$$\tilde{J}(i) = \sum_{j=1, j \neq i}^n \frac{p_{ij}(\tilde{\mu}(i))}{1 - p_{ii}(\tilde{\mu}(i))} \left( g(i, \tilde{\mu}(i), j) + \frac{g(i, \tilde{\mu}(i), i) p_{ii}(\tilde{\mu}(i))}{1 - p_{ii}(\tilde{\mu}(i))} + \tilde{J}(j) \right).$$

Then, using a reverse argument from before, we see that  $\tilde{J}(i)$  is the cost of stationary policy  $\{\tilde{\mu}, \tilde{\mu}, \dots\}$  in the original problem. Thus

$$\tilde{J}(i) \geq J^*(i) \quad \forall i.$$

Combining the two results, we have  $\tilde{J}(i) = J^*(i)$ , and thus the two problems have the same optimal costs.

If  $p_{ii}(u) = 1$  for some  $i \neq t$ , we can eliminate  $u$  from  $U(i)$  without increasing  $J^*(i)$  or any other optimal cost  $J^*(j)$ ,  $j \neq i$ . If that were not so, every optimal stationary policy must use  $u$  at state  $i$  and therefore must be improper, which is a contradiction.

## Solutions Vol. II, Chapter 3

### 3.4

By using the relation  $T_\mu(J^*) \leq T(J^*) + \epsilon e = J^* + \epsilon e$  and the monotonicity of  $T_\mu$ , we obtain

$$T_\mu^2(J^*) \leq T_\mu(J^*) + \alpha \epsilon e \leq J^* + \alpha \epsilon e + \epsilon e.$$

Proceeding similarly, we obtain

$$T_\mu^k(J^*) \leq T_\mu(J^*) + \alpha \left( \sum_{i=0}^{k-2} \alpha^i \right) \epsilon e \leq J^* + \sum_{i=0}^{k-1} \alpha^i \epsilon e$$

and by taking limit as  $k \rightarrow \infty$ , the desired result  $J_\mu \leq J^* + (\epsilon/(1-\alpha))e$  follows.

### 3.5

Under assumption P, we have by Prop. 1.2(a),  $J' \geq J^*$ . Let  $r > 0$  be such that

$$J^* \geq J' - re.$$

Then, applying  $T^k$  to this inequality, we have

$$J^* = T^k(J^*) \geq T^k(J') - \alpha^k re.$$

Taking the limit as  $k \rightarrow \infty$ , we obtain  $J^* \geq J'$ , which combined with the earlier shown relation  $J' \geq J^*$ , yields  $J' = J^*$ . Under assumption N, the proof is analogous, using Prop. 1.2(b).

### 3.8

From the proof of Proposition 1.1, we know that there exists a policy  $\pi$  such that, for all  $\epsilon_i > 0$ .

$$J_\pi(x) \leq J^*(x) + \sum_{i=0}^{\infty} \alpha^i \epsilon_i$$

Let

$$\epsilon_i = \frac{\epsilon}{2^{i+1} \alpha^i} > 0.$$

Thus,

$$J_{\pi_\epsilon}(x) \leq J^*(x) + \epsilon \sum_{i=0}^{\infty} \frac{1}{2^{i+1}} = J^*(x) + \epsilon \quad \forall x \in S.$$

If  $\alpha < 1$ , choose

$$\epsilon_i = \frac{\epsilon}{\sum_{i=0}^{\infty} \alpha^i}$$

which is independent of  $i$ . In this case,  $\pi_\epsilon$  is stationary. If  $\alpha = 1$ , we may not have a stationary policy  $\pi_\epsilon$ . In particular, let us consider a system with only one state, i.e.  $S = \{0\}$ ,  $U = (0, \infty)$ ,  $J_0(0) = 0$ , and  $g(0, u) = u$ . Then  $J^*(0) = \inf_{\pi \in \Pi} J_\pi(0) = 0$  but for every stationary policy,  $J_\mu = \sum_{k=0}^{\infty} u = \infty$ .

### 3.9

Let  $\pi^* = \{\mu_0^*, \mu_1^*, \dots\}$  be an optimal policy. Then we know that

$$J^*(x) = J_{\pi^*}(x) = \lim_{k \rightarrow \infty} (T_{\mu_0^*} T_{\mu_1^*} \dots T_{\mu_k^*})(J_0)(x) = \lim_{k \rightarrow \infty} (T_{\mu_0^*} (T_{\mu_1^*} \dots T_{\mu_k^*}))(J_0)(x).$$

From monotone convergence we know that

$$\begin{aligned} J^*(x) &= \lim_{k \rightarrow \infty} T_{\mu_0^*} (T_{\mu_1^*} \dots T_{\mu_k^*})(J_0)(x) = T_{\mu_0^*} (\lim_{k \rightarrow \infty} (T_{\mu_1^*} \dots T_{\mu_k^*})(J_0))(x) \\ &\geq T_{\mu_0^*}(J^*)(x) \geq T(J^*)(x) = J^*(x) \end{aligned}$$

Thus  $T_{\mu_0^*}(J^*)(x) = J^*(x)$ . Hence by Prop. 1.3, the stationary policy  $\{\mu_0^*, \mu_0^*, \dots\}$  is optimal.

### 3.12

We shall make an analysis similar to the one of §3.1. In particular, let

$$J_0(x) = 0$$

$$T(J_0)(x) = \min[x'Qx + u'Ru] = xqx = x'K_0x$$

$$T^2(J_0)(x) = \min[x'Qx + u'Ru + (Ax + Bu)'Q(Ax + Bu)] = x'K_1x,$$

where  $K_1 = Q + R + D_1'K_0D_1$  with  $D_1 = A + BL_1$  and  $L_1 = -(R + B'K_0B)^{-1}B'K_0A$ . Thus

$$T^k(J_0)(x) = x'K_kx$$

where  $K_k = Q + R + D_k'K_{k-1}D_k$  with  $D_k = A + BL_k$  and  $L_k = -(R + B'K_{k-1}B)^{-1}B'K_{k-1}A$ . By the analysis of Chapter 4 we conclude that  $K_k \rightarrow K$  with  $K$  being the solution to the algebraic Ricatti equation. Thus  $J_\infty(x) = x'Kx = \lim_{N \rightarrow \infty} T^N(J_0)(x)$ . Then it is easy to verify that  $J_\infty(x) = T(J_\infty)(x)$  and by Prop. 1.5 in Chapter 1, we have that  $J_\infty(x) = J^*(x)$ .

For the periodic problem the controllability assumption is that there exists a finite sequence of controls  $\{u_0, \dots, u_r\}$  such that  $x_{r+1} = 0$ . Then the optimal control sequence is periodic

$$\pi^* = \{\mu_0^*, \mu_1^*, \dots, \mu_{p-1}^*, \mu_0^*, \mu_1^*, \dots, \mu_{p-1}^*, \dots\},$$

where

$$\begin{aligned} \mu_i^* &= -(R_i + B_i'K_{i+1}B_i)^{-1}b_i'K_{i+1}A_i x \\ \mu_{p-1}^* &= -(R_{p-1} + B_{p-1}'K_0B_{p-1})^{-1}B_{p-1}'K_0A_{p-1}x \end{aligned}$$

and  $K_0, \dots, K_{p-1}$  satisfy the coupled set of  $p$  algebraic Ricatti equations

$$\begin{aligned} K_i &= A_i'[K_{i+1} - K_{i+1}B_i(R_i + B_i'K_{i+1}B_i)^{-1}B_i'K_{i+1}]A_i + Q_i, \quad i = 0, \dots, p-2, \\ K_{p-1} &= A_{p-1}'[K_0 - K_0B_{p-1}(R_{p-1} + B_{p-1}'K_0B_{p-1})^{-1}B_{p-1}'K_0]A_{p-1} + Q_{p-1}. \end{aligned}$$

### 3.14

The formulation of the problem falls under assumption P for periodic policies. All the more, the problem is discounted. Since  $w_k$  are independent with zero mean, the optimality equation for the equivalent stationary problem reduces to the following system of equations

$$\begin{aligned}\tilde{J}^*(x_0, 0) &= \min_{u_0 \in U(x_0)} E_{w_0} \{x'_0 Q_0 x_0 + u_0(x_0)' R_0 u_0(x_0) + \alpha \tilde{J}^*(A_0 x_0 + B_0 u_0 + w_0, 1)\} \\ \tilde{J}^*(x_1, 1) &= \min_{u_1 \in U(x_1)} E_{w_1} \{x'_1 Q_1 x_1 + u_1(x_1)' R_1 u_1(x_1) + \alpha \tilde{J}^*(A_1 x_1 + B_1 u_1 + w_1, 2)\} \\ &\dots \\ \tilde{J}^*(x_{p-1}, p-1) &= \min_{u_{p-1} \in U(x_{p-1})} E_{w_{p-1}} \{x'_{p-1} Q_{p-1} x_{p-1} + u_{p-1}(x_{p-1})' R_{p-1} u_{p-1}(x_{p-1}) \\ &\quad + \alpha \tilde{J}^*(A_{p-1} x_{p-1} + B_{p-1} u_{p-1} + w_{p-1}, 0)\}\end{aligned}\tag{1}$$

From the analysis in §7.8 in Ch.7 on periodic problems we see that there exists a periodic policy

$$\{\mu_0^*, \mu_1^*, \dots, \mu_{p-1}^*, \mu_1^*, \mu_2^*, \dots, \mu_{p-1}^*, \dots\}$$

which is optimal. In order to obtain the solution we argue as follows: Let us assume that the solution is of the same form as the one for the general quadratic problem. In particular, assume that

$$\tilde{J}^*(x, i) = x' K_i x + c_i,$$

where  $c_i$  is a constant and  $K_i$  is positive definite. This is justified by applying the successive approximation method and observing that the sets

$$U_k(x_i, \lambda, i) = \{u_i \in \mathcal{R}^m | x' Q x + u'_i R u_i + (Ax + Bu_i)' K_{i+1}^k (Ax + Bu_i) \leq \lambda\}$$

are compact. The latter claim can be seen from the fact that  $R \geq 0$  and  $K_{i+1}^k \geq 0$ . Then by Proposition 7.7,  $\lim_{k \rightarrow \infty} \tilde{J}_k(x_i, i) = \tilde{J}^*(x_i, i)$  and the form of the solution obtained from successive approximation is as described above.

In particular, we have for  $0 \leq i \leq p-1$

$$\begin{aligned}\tilde{J}^*(x, i) &= \min_{u_i \in U(x_i)} E_{w_i} \{x' Q_i x + u_i(x)' R_i u_i(x) + \alpha \tilde{J}^*(A_i x + B_i u_i + w_i, i+1)\} \\ &= \min_{u_i \in U(x_i)} E_{w_i} \{x' Q_i x + u_i(x)' R_i u_i(x) + \alpha [(A_i x + B_i u_i + w_i)' k_{i+1} (A_i x + B_i u_i + w_i) + c_{i+1}]\} \\ &= \min_{u_i \in U(x_i)} E_{w_i} \{x' (Q_i + \alpha A'_i K_{i+1} A_i) x_i + u'_i (R_i + \alpha B'_i K_{i+1} B_i) u_i + 2\alpha x' A'_i K_{i+1} B_i u_i + \\ &\quad + 2\alpha w'_i K_{i+1} B_i u_i + 2\alpha x' A'_i K_{i+1} w_i + w'_i K_{i+1} w_i + \alpha c_{i+1}\} \\ &= \min_{u_i \in U(x_i)} \{x' (Q_i + \alpha A'_i K_{i+1} A_i) x_i + u'_i (R_i + \alpha B'_i K_{i+1} B_i) u_i + 2\alpha x' A'_i K_{i+1} B_i u_i + \\ &\quad + w'_i K_{i+1} w_i + \alpha c_1\}\end{aligned}$$

where we have taken into consideration the fact that  $E(w_i) = 0$ . Minimizing the above quantity will give us

$$u_i^* = -\alpha (R_i + \alpha B'_i K_{i+1} B_i)^{-1} B'_i K_{i+1} A_i x \tag{2}$$



Thus

$$\tilde{J}^*(x, i) = x' [Q_i + A_i'(\alpha K_{i+1} - \alpha^2 K_{i+1}(R_i + \alpha B_i' K_{i+1} B_i)^{-1} B_i' K_{i+1}) A_i] x + c_i = x' K_i x + c_i$$

where  $c_i = E_{w_i} \{w_i' K_{i+1} w_i\} + \alpha c_{i+1}$  and

$$K_i = Q_i + A_i'(\alpha K_{i+1} - \alpha^2 K_{i+1}(R_i + \alpha B_i' K_{i+1} B_i)^{-1} B_i' K_{i+1}) A_i.$$

Now for this solution to be consistent we must have  $K_p = K_0$ . This leads to the following system of equations

$$\begin{aligned} K_0 &= Q_0 + A_0'(\alpha K_1 - \alpha^2 K_1(R_0 + \alpha B_0' K_1 B_0)^{-1} B_0' K_1) A_0 \\ &\dots \\ K_i &= Q_i + A_i'(\alpha K_{i+1} - \alpha^2 K_{i+1}(R_i + \alpha B_i' K_{i+1} B_i)^{-1} B_i' K_{i+1}) A_i \\ &\dots \\ K_{p-1} &= Q_{p-1} + A_{p-1}'(\alpha K_0 - \alpha^2 K_0(R_{p-1} + \alpha B_{p-1}' K_0 B_{p-1})^{-1} B_{p-1}' K_0) A_{p-1} \end{aligned} \quad (3)$$

This system of equations has a positive definite solution since (from the description of the problem) the system is controllable, i.e. there exists a sequence of controls such that  $\{u_0, \dots, u_r\}$  such that  $x_{r+1} = 0$ . Thus the result follows.

### 3.16

(a) Consider the stationary policy,  $\{\mu_0, \mu_0, \dots\}$ , where  $\mu_0 = L_0 x$ . We have

$$J_0(x) = 0$$

$$T_{\mu_0}(J_0)(x) = x' Q x + x' L_0' R L_0 x$$

$$\begin{aligned} T_{\mu_0}^2(J_0)(x) &= x' Q x + x' L_0' R L_0 x + \alpha(Ax + BL_0 x + w)' Q (Ax + BL_0 x + w) \\ &= x' M_1 x + \text{constant} \end{aligned}$$

where  $M_1 = Q + L_0' R L_0 + \alpha(A + BL_0)' Q (A + BL_0)$ ,

$$\begin{aligned} T_{\mu_0}^3(J_0)(x) &= x' Q x + x' L_0' R L_0 x + \alpha(Ax + BL_0 x + w)' M_1 (Ax + BL_0 x + w) + \alpha \cdot (\text{constant}) \\ &= x' M_2 x + \text{constant} \end{aligned}$$

Continuing similarly, we get

$$M_{k+1} = Q + L_0' R L_0 + \alpha(A + BL_0)' M_k (A + BL_0).$$

Using a very similar analysis as in Section 8.2, we get

$$M_k \rightarrow K_0$$

where

$$K_0 = Q + L_0'RL_0 + \alpha(A + BL_0)'K_0(A + BL_0)$$

(b)

$$\begin{aligned} J_{\mu_1}(x) &= \lim_{N \rightarrow \infty} E_{w_k, k=0, \dots, N-1} \left\{ \sum_{k=0}^{N-1} \alpha^k [x_k' Q x_k + \mu_1(x_k)' R \mu_1(x_k)] \right\} \\ &= \lim_{N \rightarrow \infty} T_{\mu_1}^N(J_{\mu_0})(x) \end{aligned}$$

Proceeding as in the proof of the validity of policy iteration (Section 7.3, Chapter 7). We have

$$T_{\mu_1}(J_{\mu_0}) = T(J_{\mu_0})$$

$$J_{\mu_0}(x) = x'K_0x + \text{constant} = T_{\mu_0}(J_{\mu_0})(x) \geq T_{\mu_1}(J_{\mu_0})(x)$$

Hence, we obtain

$$J_{\mu_0}(x) \geq T_{\mu_1}(J_{\mu_0})(x) \geq \dots \geq T_{\mu_1}^k(J_{\mu_0})(x) \geq \dots$$

implying,

$$J_{\mu_0}(x) \geq \lim_{k \rightarrow \infty} T_{\mu_1}^k(J_{\mu_0})(x) = J_{\mu_1}(x).$$

(c) As in part (b), we show that

$$J_{\mu_k}(x) = x'K_kx + \text{constant} \leq J_{\mu_{k-1}}(x).$$

Now since

$$0 \leq x'K_kx \leq x'K_{k-1}x, \quad \forall x$$

we have

$$K_k \rightarrow K.$$

The form of  $K$  is,

$$\begin{aligned} K &= \alpha(A + BL)'K(A + BL) + Q + L'RL \\ L &= -\alpha(\alpha B'KB + R)^{-1}B'KA \end{aligned}$$

To show that  $K$  is indeed the optimal cost matrix, we have to show that it satisfies

$$\begin{aligned} K &= A'[\alpha K - \alpha^2 KB(\alpha B'KB + R)^{-1}B'K]A + Q \\ &= A'[\alpha KA + \alpha KBL] + Q \end{aligned}$$

Let us expand the formula for  $K$ , using the formula for  $L$ ,

$$K = \alpha(A'KA + A'KBL + L'B'KA + L'B'KBL) + Q + L'RL.$$

Substituting, we get

$$\begin{aligned} K &= \alpha(A'KA + A'KBL + L'B'KA) + Q - \alpha L'B'KA \\ &= \alpha A'KA + \alpha A'KBL + Q. \end{aligned}$$

Thus  $K$  is the optimal cost matrix.

*A second approach:* (a) We know that

$$J_{\mu_0}(x) = \lim_{n \rightarrow \infty} T_{\mu_0}^n(J_0)(x).$$

Following the analysis at §8.1 we have

$$\begin{aligned} J_0(x) &= 0 \\ T_{\mu_0}(J)(x) &= E\{x'Qx + \mu_0(x)'R\mu_0(x)\} = x'Qx + \mu_0(x)'R\mu_0(x) = x'(Q + L_0'RL_0)x \\ T_{\mu_0}^2(J)(x) &= E\{x'Qx + \mu_0'(x)R\mu_0(x) + \alpha(Ax + B\mu_0(x) + w)'Q(Ax + B\mu_0(x) + w)\} \\ &= x'(Q + L_0'RL_0 + \alpha(A + BL_0)'Q(A + BL_0))x + \alpha E\{w'Qw\}. \end{aligned}$$

Define

$$\begin{aligned} K_0^0 &= Q \\ K_0^{k+1} &= Q + L_0'RL_0 + \alpha(A + BL_0)'K_0^k(A + BL_0). \end{aligned}$$

Then

$$T_{\mu_0}^{k+1}(J)(x) = x'K_0^{k+1}x + \sum_{m=0}^{k-1} \alpha^{k-m} E\{w'K_0^m w\}.$$

The convergence of  $K_0^{k+1}$  follows from the analysis of §4.1. Thus

$$J_{\mu_0}(x) = x'K_0x + \frac{\alpha}{1-\alpha} E\{w'K_0w\}$$

(as in §8.1) which proves the required relation.

(b) Let  $\mu_1(x)$  be the solution of the following

$$\min_u \{u'Ru + \alpha(Ax + Bu)'K_0(Ax + Bu)\}$$

which yields

$$u_1 = -(R + \alpha B'K_0B)^{-1} \alpha B'K_0Ax = L_1x.$$

Thus

$$L_1 = -(R + \alpha B'K_0B)^{-1} \alpha B'K_0A = -M^{-1}\Pi$$

where  $M = R + \alpha B'K_0B$  and  $\Pi = \alpha B'K_0A$ . Let us consider the cost associated with  $u_1$  if we ignore  $w$

$$J_{\mu_1}(x) = \sum_{k=0}^{\infty} \alpha^k (x_k'Qx_k + \mu_1(x_k)'R\mu_1(x_k)) = \sum_{k=0}^{\infty} \alpha^k x_k'(Q + L_1'RL_1)x_k.$$

However, we know the following

$$x_{k+1} = (A + BL_1)^{k+1}x_0 + \sum_{m=1}^{k+1} (A + BL_1)^{k+1-m}w_m.$$

Thus, if we ignore the disturbance  $w$  we get

$$J_{\mu_1}(x) = x'_0 \sum_{k=0}^{\infty} \alpha^k (A + BL_1)'^k (Q + L_1 RL_1) (A + BL_1)^k x_0.$$

Let us call

$$K_1 = \sum_{k=0}^{\infty} \alpha^k (A + BL_1)'^k (Q + L'_1 RL_1) (A + BL_1)^k x_0. \quad (1)$$

We know that

$$K - 0 - \alpha(A + BL_0)' K_0 (A + BL_0) - L'_0 RL_0 = Q.$$

Substituting in (1) we have

$$\begin{aligned} K_1 &= \sum_{k=0}^{\infty} \alpha^k (A + BL_1)'^k (K_0 + \alpha(A + BL_1)' K_0 (A + BL_1)) (A + BL_1)^k + \\ &+ \sum_{k=0}^{\infty} \{ \alpha^k (A + BL_1)'^k [\alpha(A + BL_1)' K_0 (A + BL_1) - \alpha(A + BL_0)' K_0 (A + BL_0) + \\ &+ L'_1 RL_1 - L'_0 RL_0] (A + BL_1)^k \}. \end{aligned}$$

However, we know that

$$K_0 = \sum_{k=0}^{\infty} \alpha^k (A + BL_1)'^k (K_0 - \alpha(A + BL_1)' K_0 (A + BL_1)) (A + BL_1)^k.$$

Thus we conclude that

$$K_1 - K_0 = \sum_{k=0}^{\infty} \alpha^k (A + BL_1)^k \Psi (A + BL_1)^k$$

where

$$\Psi = \alpha(A + BL_1)' K_0 (A + BL_1) - \alpha(A + BL_0)' K_0 (A + BL_0) + L'_1 K_0 L_1 + L'_0 K_0 L_0.$$

We manipulate the above equation further and we obtain

$$\begin{aligned} \Psi &= L'_1 (R + \alpha B' K_0 B) L_1 - L'_0 (R + \alpha B' K_0 B) L_0 + \alpha L'_1 B' K_0 A + \alpha A' K_0 B L_1 - \\ &- \alpha L'_0 B' K_0 A - \alpha A' K_0 B L_0 \\ &= L'_1 M L_1 - L'_0 M L_0 + L'_1 \Pi + \Pi' L_1 - L'_0 \Pi - \Pi' L_0 \\ &= -(L_0 - L_1)' M (L_0 - L_1) - (\Pi + M L_1)' (L_0 - L_1) - (L_0 - L_1)' (\Pi + M L_1). \end{aligned}$$

However, it is seen that

$$\Pi + M L_1 = 0.$$

Thus

$$\Psi = -(L_0 - L_1)' M (L_0 - L_1).$$

Since  $M \geq 0$  we conclude that

$$K_0 - K_1 = \sum_{k=0}^{\infty} \alpha^k (A + BL_1)'^k (L_0 - L_1) M (L_0 - L_1) (A + BL_1)^k \geq 0.$$

Similarly, the optimal solution for the case where there are no disturbances satisfies the equation

$$K = Q + L'RL + \alpha(A + BL)'K(A + BL)$$

with  $L = -\alpha(R + B'KB)^{-1}B'KA$ . If we follow the same steps as above we will obtain

$$K_1 - K = \sum_{k=0}^{\infty} \alpha^k (A + BL_1)'^k (L_1 - L)' M (L_1 - L) (A + BL_1)^k \geq 0.$$

Thus  $K \leq K_1 \leq K_0$ . Since  $K_1$  is bounded, we conclude that  $A + BL_1$  is stable (otherwise  $K_1 \rightarrow \infty$ ). Thus, the sum converges and  $K_1$  is the solution of  $K_1 = \alpha(A + BL_1)'K_1(A + L_1) + Q + L_1'RL_1$ . Now returning to the case with the disturbances  $w$  we conclude as in case (a) that

$$J_{\mu_1}(x) = x'K_1x + \frac{\alpha}{1-\alpha}E\{w'K_1w\}.$$

Since  $K_1 \leq K_0$  we conclude that  $J_{\mu_1}(x) \leq J_{\mu_0}(x)$  which proves the result.

c) The policy iteration is defined as follows: Let

$$L_k = -\alpha(R + \alpha B'K_{k-1}B)^{-1}B'K_{k-1}A.$$

Then  $\mu_k(x) = L_kx$  and

$$J_{\mu_k}(x) = x'K_kx + \frac{\alpha}{1-\alpha}E\{w'K_kw\}$$

where  $K_k$  is obtained as the solution of

$$K_k = \alpha(A + BL_k)'K_k(A + BL_k) + Q + L_k'RL_k.$$

If we follow the steps of (b) we can prove that

$$K \leq K_k \leq \dots \leq K_1 \leq K_0. \quad (2)$$

Thus by the theorem of monotonic convergence of positive operators (Kantorovich and Akilov p. 189: "Functional Analysis in Normed Spaces") we conclude that

$$K_{\infty} = \lim_{p \rightarrow \infty} K_k$$

exists. Then if we take the limit of both sides of eq. (2) we have

$$K_{\infty} = \alpha(A + BL_{\infty})'K_{\infty}(A + L_{\infty}) + Q + L_{\infty}'RL_{\infty}$$

with

$$L_{\infty} = -\alpha(R + \alpha B'K_{\infty}B)^{-1}B'K_{\infty}A.$$

However, according to §4.1,  $K$  is the unique solution of the above equation. Thus,  $K_{\infty} = K$  and the result follows.

*Solutions Vol. II, Chapters 4, 5, 6*