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

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

So

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

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

Thus

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

Q.E.D.

1.7

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

$$J\le T(J) \qquad \Rightarrow \qquad T(J)\le F(J), \tag{1}$$

$$J \geq T(J) \qquad \Rightarrow \qquad T(J) \geq F(J). \tag{2}$$

For any μ, 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

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

Fix $\epsilon > 0$. If $J \leq T(J)$, let μ 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 μ 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

$$F(J + re) \leq F(J) + \alpha re, \qquad F(J) - \alpha re \leq F(J - re). \tag{4}$$

Furthermore, F is monotone, that is

$$J \leq J' \qquad \Rightarrow \qquad F(J) \leq F(J'). \tag{5}$$

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

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

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

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 $J : S \to R$, $J' : S \to R$, we have

$$\max_j |F(J)(j) - F(J')(j)| \leq \alpha \max_j |J(j) - J'(j)|, \tag{6}$$

so F is a contraction mapping with modulus α. Indeed, we have

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

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

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

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

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

Combining the last two inequalities, we see that

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

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

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

and

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

Thus

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

or

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

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

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

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

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

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 $\lim_{k \to \infty} T^k(J)(x) = J^*(x)$.

(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 $c \in \Re$, $(c < \infty)$, 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).

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

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

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

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

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

Finally we note that

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

Thus

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

Let $b_{k+1} = \inf_{x \in S} \left(T^{k+1}(J)(x) - T^k(J)(x)\right)$. 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

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

Thus

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

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

$$F(J+re)(2) = \min_{u\in U(2)}\left\{g(2,u)+\alpha p_{21}F(J+re)(1)+\alpha\sum_{J=2}^{n}p_{2j}(J+re)(j)\right\}$$
$$= \min_{u\in U(2)}\left\{g(2,u)+\alpha p_{21}F(J)(1)+\alpha^2 rp_{21}+\alpha\sum_{J=2}^{n}p_2J(j)+\alpha\sum_{J=2}^{n}p_{ij}re(j)\right\}$$

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

Thus we conclude

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

which yields

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

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 α^2, α, it is easy to see that

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

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

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

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 $x = i + j$

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

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

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

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

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

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

Since $0 < \alpha^{i+1} \leq \alpha^2 \leq \alpha < 1$ and $0 \leq p \leq i$ we conclude that $\alpha^{i+1} \leq a^{i+1} p + \alpha(1-p)$ and $a^2 p + (1-p)\alpha \leq \alpha$. 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 $J(x) \leq J'9x)(=)J'(x) - J(x) \geq 0$ Since all the elements m_{ij} are non-negative we conclude that

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

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

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

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 μ 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 μ^k is the unique μ 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 μ^{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 $J_1^*, \ldots, J_n^*$ satisfy

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

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

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

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

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

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

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

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

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

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 $M = m(x_0)$.

Solutions Vol. II, Chapter 2

2.1

(a) (i) First, we need to define a state space for the problem. The obvious choice for a state variable is our location. However, this does not encapsulate all of the necessary information. We also need to include the value of c if it is known. Thus, let the state space consist of the following $2m + 2$ states: $\{S, S_1, \ldots, S_m, I_1, \ldots I_m, D\}$, where S is associated with being at the starting point with no information, S_i and I_i are associated with being at S and I, respectively, and knowing that $c = c_i$, and D is the termination state.

At state S, there are two possible controls: go directly to D (*direct*) or go to an intermediate point (*indirect*). If control *direct* is selected, we go to state D with probability 1, and the cost is $g(S, direct, D) = a$. If control *indirect* is selected, we go to state I_i with probability p_i, and the cost is $g(S, indirect, I_i) = b$.

At state S_i, for $i \in \{1, \ldots, m\}$, we have the same controls as at state S. Again, if control *direct* is selected, we go to state D with probability 1, and the cost is $g(S_i, direct, D) = a$. If, on the other hand, control *indirect* is selected, we go to state I_i with probability 1, and the cost is $g(S, indirect, I_i) = b$.

At state I_i, for $i \in \{1, \ldots, m\}$, there are also two possible controls: go back to the start (*start*) or go to the destination (*dest*). If control *start* is selected, we go to state S_i with probability 1, and the cost is $g(I_i, start, S_i) = b$. If control *dest* is selected, we go to state D with probability 1, and the cost is $g(I_i, dest, D) = c_i$.

We have thus formulated the problem as a stochastic shortest path problem. Bellman's equation for this problem is

$$J^*(S) = \min[a, b + \sum_{i=1}^{m} p_i J^*(I_i)]$$

$$J^*(S_i) = \min[a, b + J^*(I_i)]$$

$$J^*(I_i) = \min[c_i, b + J^*(S_i)].$$

We assume that $b > 0$. Then, Assumptions 5.1 and 5.2 hold since all improper policies have infinite cost. As a result, if $\mu^*(I_i) = start$, then $\mu^*(S_i) = direct$. If $\mu^*(I_i) \neq start$, then we never reach state S_i and so it doesn't matter what the control is in this case. Thus, $J^*(S_i) = a$, and $\mu^*(S_i) = direct$. From this, it is easy to derive the optimal costs and controls for the other states

$$J^*(I_i) = \min[c_i, b + a] \qquad \mu^*(I_i) = \begin{cases} \text{dest}, & \text{if } c_i < b + a \\ \text{start}, & \text{otherwise,} \end{cases}$$

$$J^*(S) = \min[a, b + \sum_{i=1}^{m} p_i \min(c_i, b + a)]$$

$$\mu^*(S) = \begin{cases} \text{direct}, & \text{if } a < b + \sum_{i=1}^{m} p_i \min(c_i, b + a) \\ \text{indirect}, & \text{otherwise.} \end{cases}$$

For the numerical case given, we see that $a < b + \sum_{i=1}^{m} p_i \min(c_i, b + a)$ since $a = 2$ and $b + \sum_{i=1}^{m} p_i \min(c_i, b + a) = 2.5$. Hence $\mu(S) = direct$. We need not consider the other states since they will never be reached.

(ii) In this case, every time we are at the starting location, our available information is the same. We thus no longer need the states S_i from part (i). Our state space for this part is then $S, I_1, \ldots, I_m, D$.

At state S, the possible controls are $\{direct, indirect\}$. If control *direct* is selected, we go to state D with probability 1, and the cost is $g(S, direct, D) = a$. If control *indirect* is selected, we go to state I_i with probability p_i, and the cost is $g(S, indirect, I_i) = b$ [same as in part (ii)].

At state I_i, for $i \in \{1, \ldots, m\}$, the possible controls are $\{start, dest\}$. If control *start* is selected, we go to state S with probability 1, and the cost is $g(I_i, start, S) = b$. If control *dest* is selected, we go to state D with probability 1, and the cost is $g(I_i, dest, D) = c_i$.

Bellman's equation for this stochastic shortest path problem is

$$J^*(S) = \min[a, b + \sum_{i=1}^{m} p_i J^*(I_i)]$$

$$J^*(I_i) = \min[c_i, b + J^*(S)].$$

The optimal policy can be described by

$$\mu^*(S) = \begin{cases} \text{direct}, & \text{if } a < b + \sum_{i=1}^{m} p_i J^*(I_i) \\ \text{indirect}, & \text{otherwise,} \end{cases}$$

$$\mu^*(I_i) = \begin{cases} \text{dest}, & \text{if } c_i < b + J^*(S) \\ \text{start}, & \text{otherwise.} \end{cases}$$

We will solve the problem for the numerical case by "guessing" an optimal policy and then showing that the resulting cost J_μ satisfies $J = TJ$. Since J^* is the unique solution to this equation, our policy is optimal. So let's guess the initial policy to be

$$\mu^*(S) = direct \qquad \mu^*(I_1) = dest \qquad \mu^*(I_2) = start.$$

Then

$$J(S) = a = 2 \qquad J(I_1) = c_1 = 0 \qquad J(I_2) = b + J^*(S) = 1 + 2 = 3.$$

From Bellman's equation, we have

$$J(S) = \min(2, 1 + 0.5(3 + 0)) = 2$$

$$J(I_1) = \min(0, 1 + 2)) = 0$$

$$J(I_2) = \min(5, 1 + 2)) = 3.$$

Thus, our policy is optimal.

(b) The state space for this problem is the same as for part a(ii): $\{S, I_1, \ldots, I_m, D\}$.

At state S, the possible controls are $\{direct, indirect\}$. If control *direct* is selected, we go to state D with probability 1, and the cost is $g(S, direct, D) = a$. If control *indirect* is selected, we go to state I_i with probability p_i, and the cost is $g(S, indirect, I_i) = b$ [same as in part a,(i) and (ii)].

At state I_i, for $i \in \{1, \ldots, m\}$, we have an additional option of waiting. So the possible controls are $\{start, dest, wait\}$. If control *start* is selected, we go to state S with probability 1, and the cost is $g(I_i, start, S) = b$. If control *dest* is selected, we go to state D with probability 1, and the cost is $g(I_i, dest, D) = c_i$. If control *wait* is selected, we go to state I_j with probability p_j, and the cost is $g(I_i, wait, I_j) = d$.

Bellman's equation is

$$J^*(S) = \min[a, b + \sum_{i=1}^{m} p_i J^*(I_i)]$$

$$J^*(I_i) = \min[c_i, b + J^*(S), d + \sum_{j=1}^{m} p_j J^*(I_j)].$$

We can describe the optimal policy as follows:

$$\mu^*(S) = \begin{cases} \text{direct}, & \text{if } a < b + \sum_{i=1}^{m} p_i J^*(I_i) \\ \text{indirect}, & \text{otherwise.} \end{cases}$$

If *direct* was selected, we do not need to consider the other states (other than D) since they will never be reached. If *indirect* was selected, then defining $k = \min(2b, d)$, we see that

$$\mu^*(I_i) = \begin{cases} \text{dest}, & \text{if } c_i < k + \sum_{i=1}^{m} J^*(I_i) \\ \text{start}, & \text{if } c_i > k + \sum_{i=1}^{m} J^*(I_i) \text{ and } 2b < d \\ \text{wait}, & \text{if } c_i > k + \sum_{i=1}^{m} J^*(I_i) \text{ and } 2b > d. \end{cases}$$

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 π_1, this is inevitable. Thus π_1 is an improper policy, and π_2, π_3, and π_4 are proper policies.

(b) Let $J_{\pi_1}(H)$ and $J_{\pi_2}(T)$ be the costs corresponding policy π_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 π_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 π^* from the second stage on. Consider a policy $\tilde{\pi}$ which plays fair at the first stage, and then follows π^* from the second stage on if the outcome of the first stage is H or cheats at the second stage and follows π^* from the third stage on if the outcome of the first stage is T. We have

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

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

where the inequality follows from the fact that $J_{\pi^*}(H) \geq J_{\tilde{\pi}}(H)$ since π^* 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: π_1 and π_3. Under π_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 π_3, we have

$$J_{\pi_3}(T) = (1-p)J_{\pi_3}(H),$$

$$J_{\pi_3}(H) = \frac{1}{2}[1 + J_{\pi_3}(H)] + \frac{1}{2}J_{\pi_3}(T).$$

Solving these two equations yields

$$J_{\pi_3}(T) = \frac{1-p}{p},$$

$$J_{\pi_3}(H) = \frac{1}{p}.$$

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,

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

In the above equation, we can think of the union of $S_{m-1},\ldots,S_1$, 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,

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

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 $m=1,\ldots,M$.

The original stochastic shortest path problem can be solved as M stochastic shortest path subproblems. 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,

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

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

for $m=1,2,\ldots,M$. (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

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

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

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

Then

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

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

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

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

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

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

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

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

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

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

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.

2.17

Consider a modified stochastic shortest path problem where the state space is denoted by S', the control space by U', the transition costs by g', and the transition probabilities by p'. Let the state space $S' = S'_S \cup S'_{S-U}$, where

$$S'_S = \{1, \ldots, n, t\} \text{ where each } i \in S'_S \text{ corresponds to } i \in S$$

$$S'_{S-U} = \{(i,u) | i \in S, u \in U(i)\} \text{ where each } (i,u) \in S'_{S-U} \text{ corresponds to } i \in S \text{ and } u \in U(i).$$

For $i, j \in S'_S, u \in U'(i)$, we define $U'(i) = U(i), g'(i,u,j) = g(i,u,j)$, and $p'_{ij}(u) = p_{ij}(u)$. For $(i,u) \in S'_S - U$ and $j \in S'_S$, the only possible control is $u' = u$ (*i.e.*, $U'(i,u) = \{u\}$), and we have $g'((i,u),u',j) = g(i,u,j)$ and $p'_{(i,u)j}(u') = p_{ij}(u)$.

Since trajectories originating from a state $i \in S'_S$ are equivalent to trajectories in the original problem, the optimal cost-to-go value for state i in the modified problem is $J^*(i)$, the optimal cost-to-go value from the original problem. Let us denote the optimal cost-to-go value for $(i,u) \in S'_{S-U}$ by $J^*(i,u)$. Then $J^*(i)$ and $J^*(i,u)$ solve uniquely Bellman's equation of the modified problem, which is

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

$$J^*(i,u) = \sum_{j=1}^{n} p_{ij}(u)(g(i,u,j) + J^*(j)). \tag{2}$$

The Q-factors for the original problem are defined as

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

so from Eq. (2), we have

$$Q(i,u) = J^*(i,u), \quad \forall\ (i,u). \tag{3}$$

Also from Eqs. (1) and (2), we have

$$J^*(i) = \min_{u \in U(i)} J^*(i,u), \quad \forall\ i. \tag{4}$$

Thus from Eqs. (1)-(4), we obtain

$$Q(i,u) = \sum_{j=1}^{n} p_{ij}(u) \left(g(i,u,j) + \min_{u' \in U(j)} Q(j,u') \right). \tag{5}$$

There remains to show that there is no other solution to Eq. (5). Indeed, if $\hat{Q}(i,u)$ were such that

$$\hat{Q}(i,u) = \sum_{j=1}^{n} p_{ij}(u) \left(g(i,u,j) + \min_{u' \in U(j)} \hat{Q}(j,u') \right), \quad \forall\ (i,u), \tag{6}$$

then by defining

$$\hat{J}(i) = \min_{u \in U(i)} \hat{Q}(i,u) \tag{7}$$

we obtain from Eq. (6)

$$\hat{Q}(i,u) = \sum_{j=1}^{n} p_{ij}(u)(g(i,u,j) + \hat{J}(j)), \quad \forall\ (i,u). \tag{8}$$

By combining Eqs. (7) and (8), we have

$$\hat{J}(i) = \min_{u \in U(i)} \sum_{j=1}^{n} p_{ij}(u)(g(i,u,j) + \hat{J}(j)), \quad \forall\ i. \tag{9}$$

Thus $\hat{J}(i)$ and $\hat{Q}(i,u)$ satisfy Bellman's Eq. (1)-(2) for the modified problem. Since this Bellman equation is solved uniquely by $J^*(i)$ and $J^*(i,u)$, we see that

$$\hat{Q}(i,u) = J^*(i,u) = Q(i,u), \quad \forall\ (i,u).$$

Thus the Q-factors $Q(i,u)$ solve uniquely Eq. (5).

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_μ, 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 $k \to \infty$, 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 $k \to \infty$, 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 π 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,

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

If $\alpha < 1$, choose

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

which is independent of i. In this case, π_ϵ is stationary. If $\alpha = 1$, we may not have a stationary policy π_ϵ. 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 $\pi^* = \{\mu_0^*, \mu_1^*, \ldots\}$ be an optimal policy. Then we know that

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

From monotone convergence we know that

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

Thus $T_{\mu_0^*}(J^*)(x) = J^*(x)$. Hence by Prop. 1.3, the stationary policy $\{\mu_0^*, \mu_0^*, \ldots\}$ 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 $K_k \to K$ with K being the solution to the algebraic Ricatti equation. Thus $J_\infty(x) = x'Kx = \lim_{N \to \infty} T^N(J_0)(x)$. 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 $\{u_0, \ldots, u_r\}$ such that $x_{r+1} = 0$. Then the optimal control sequence is periodic

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

where

$$\mu_i^* = -(R_i + B_i' K_{i+1} B_i)^{-1} b_i' K_{k+1} A_i x$$

$$\mu_{p-1}^* = -(R_{p-1} + B_{p-1}' K_0 B_{p-1})^{-1} B_{p-1}' K_0 A_{p-1} x$$

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

$$K_i = A_i'[K_{i+1} - K_{i+1} B_i (R_i + B_i' K_{i+1} B_i)^{-1} B_i' K_{i+1}] A_i + Q_i, \quad i = 0, \ldots, p-2,$$

$$K_{p-1} = A_{p-1}'[K_0 - K_0 B_{p-1} (R_{p-1} + B_{p-1}' K_0 B_{p-1})^{-1} B_{p-1}' K_0] A_{p-1} + Q_{p-1}.$$

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

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

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

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

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

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

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, $\lim_{k \to \infty} \tilde{J}_k(x_i, i) = \tilde{J}^*(x_i, i)$ and the form of the solution obtained from successive approximation is as described above.

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

$$\begin{aligned}
\tilde{J}^*(x,i) &= \min_{u_i \in U(x_i)} E_{w_i}\{x'Q_ix + u_i(x)'R_1u_i(x) + \alpha\tilde{J}^*(A_1x + B_1u_i + w_i, i+1)\} \\
&= \min_{u_i \in U(x_i)} E_{w_i}\{x'Q_ix + u_i(x)'R_1u_i(x) + \alpha\left[(A_ix + B_iu_i + w_i)'k_{i+1}(A_ix + B_iu_i + w_i) + c_{i+1}\right]\} \\
&= \min_{u_i \in U(x_i)} E_{w_i}\{x'(Q_i + \alpha A_i'K_{i+1}A_i)x_i + u_i'(r_i + \alpha B_i'K_{i+1}B_i)u_i + 2\alpha x'A_i'K_{i+1}B_iu_i + \\
&\qquad + 2\alpha w_i'K_{i+1}B_iu_i + 2\alpha x'A_i'K_{i+1}w_i + w_i'K_{i+1}w_i + \alpha c_{i+1}\} \\
&= \min_{u_i \in U(x_i)} \{x'(Q_i + \alpha A_i'K_{i+1}A_i)x_i + u_i'(R_i + \alpha B_i'K_{i+1}B_i)u_i + 2\alpha x'A_i'K_{i+1}B_iu_i + \\
&\qquad + w_i'K_{i+1}w_i + \alpha c_1\}
\end{aligned}$$

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

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

Thus

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

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

$$\begin{aligned}
K_0 &= Q_0 + A_0'(\alpha K_1 - \alpha^2 K_1(R_0 + \alpha B_0'K_1B_0)^{-1}B_0'K_1)A_0 \\
&\cdots \\
K_i &= Q_i + A_i'(\alpha K_{i+1} - \alpha^2 K_{i+1}(R_i + \alpha B_i'K_{i+1}B_i)^{-1}B_i'K_{i+1})A_i \\
&\cdots \\
K_{p-1} &= Q_{p-1} + A_{p-1}'(\alpha K_0 - \alpha^2 K_0(R_{p-1} + \alpha B_{p-1}'K_0B_{p-1})^{-1}B_{p-1}'K_0)A_{p-1}
\end{aligned} \tag{3}$$

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 $\{u_0, \ldots, u_r\}$ such that $x_{r+1} = 0$. Thus the result follows.

3.16

(a) Consider the stationary policy, $\{\mu_0, \mu_0, \ldots, \}$, where $\mu_0 = L_0x$. We have

$$J_0(x) = 0$$

$$T_{\mu_0}(J_0)(x) = x'Qx + x'L_0'RL_0x$$

$$T^2_{\mu_0}(J_0)(x) = x'Qx + x'L_0'RL_0x + \alpha(Ax + BL_0x + w)'Q(Ax + BL_0x + w)$$

$$= x'M_1x + \text{constant}$$

where $M_1 = Q + L_0'RL_0 + \alpha(A + BL_0)'Q(A + BL_0)$,

$$T^3_{\mu_0}(J_0)(x) = x'Qx + x'L_0'RL_0x + \alpha(Ax + BL_0x + w)'M_1(Ax + BL_0 + w) + \alpha \cdot (\text{constant})$$

$$= x'M_2x + \text{constant}$$

Continuing similarly, we get

$$M_{k+1} = Q + L_0'RL_0 + \alpha(A + BL_0)'M_k(A + BL_0).$$

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

$$M_k \to K_0$$

where

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

(b)

$$J_{\mu_1}(x) = \lim_{N\to\infty} E_{\substack{w_k \\ k=0,\cdots,N-1}} \left\{ \sum_{k=0}^{N-1} \alpha^k \left[x_k'Qx_k + \mu_1(x_k)'R\mu_1(x_k) \right] \right\}$$

$$= \lim_{N\to\infty} T^N_{\mu_1}(J_{\mu_0})(x)$$

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

$$J_{\mu_0}(x) \geq T_{\mu_1}(J_{\mu_0})(x) \geq \ldots \geq T^k_{\mu_1}(J_{\mu_0})(x) \geq \ldots$$

implying,

$$J_{\mu_0}(x) \geq \lim_{k\to\infty} T^k_{\mu_1}(J_{\mu_0})(x) = J_{\mu_1}(x).$$

(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

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

we have

$$K_k \to K.$$

The form of K is,

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

$$L = -\alpha(\alpha B'KB + R)^{-1}B'KA$$

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

$$J_{\mu_0}(x) = \lim_{n\to\infty} T^n_{\mu_0}(J_0)(x).$$

Following the analysis at §8.1 we have

$$J_0(x) = 0$$

$$\begin{aligned} T_{\mu_0}(J)(x) &= E\{x'Qx + \mu_0(x)'R\mu_0(x)\} = x'Qx + \mu_0(x)'R\mu_0(x) = x'(Q + L_0'RL_0)x \\ T^2_{\mu_0}(J)(x) &= E\{x'Qx + \mu_0'(x)R\mu_0(x) + \alpha(Ax + B\mu_0(x) + w)'Q(Ax + B\mu_0(x) + w)\} \\ &= x'\left(Q + L_0'RL_0 + \alpha(A + BL_0)'Q(A + BL_0)\right)x + \alpha E\{w'Qw\}. \end{aligned}$$

Define

$$K_0^0 = Q$$

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

Then

$$T^{k+1}_{\mu_0}(J)(x) = x'K_0^{k+1}x + \sum_{m=0}^{k-1} \alpha^{k-m}E\{w'K_0^m w\}.$$

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

$$J_{\mu_1}(x) = \sum_{k=0}^{\infty} \alpha^k \left(x_k'Qx_k + \mu_1(x_k)'Rm_1(x_k)\right) = \sum_{k=0}^{\infty} \alpha^k x_k'(Q + L_1'RL_1)x_k.$$

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

$$J_{\mu_1}(x) = x_0' \sum_{k=0}^{\infty} \alpha^k (A + BL_1)'^k(Q + L_1RL_1)(A + BL_1)^k x_0.$$

Let us call

$$K_1 = \sum_{k=0}^{\infty} \alpha^k (A + BL_1)'^k(Q + L_1'RL_1)(A + BL_1)^k x_0. \tag{1}$$

We know that

$$K - 0 - \alpha(A + BL_0)'K_0(A + BL_0) - L_0'RL_0 = Q.$$

Substituting in (1) we have

$$\begin{aligned} K_1 = &\sum_{k=0}^{\infty} \alpha^k (A + BL_1)'^k(K_0 + \alpha(A + BL_1)'K_0(A + BL_1))(A + BL_1) + \\ &+ \sum_{k=0}^{\infty}\{\alpha^k (A + BL_1)'^k[\alpha(A + BL_1)'K_0(A + BL_1) - \alpha(A + BL_0)'K_0(A + BL_0) + \\ &+ L_1'RL_1 - L_0'RL_0](A + BL_1)^k\}. \end{aligned}$$

However, we know that

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

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

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

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

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

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 $K_1 \to \infty$). Thus, the sum converges and K_1 is the solution of $K_1 = \alpha(A + BL_1)' K_1 (A + L_1) + Q + L_1' R L_1$. Now returning to the case with the disturbances w we conclude as in case (a) that

$$J_{\mu_1}(x) = x' K_1 x + \frac{\alpha}{1 - \alpha} E\{w' K_1 w\}.$$

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 $\mu_k(x) = L_k x$ and

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

where K_k is obtained as the solution of

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

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

$$K \leq K_k \leq \ldots \leq K_1 \leq K_0. \tag{2}$$

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

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

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

$$K_\infty = \alpha(A + BL_\infty)' K_\infty (A + L_\infty) + Q + L_\infty' R L_\infty$$

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, Chapter 4

4.4

(a) We have

$$T^{k+1}h^0 = T\left(T^k h^0\right) = T\left(h_i^k + \left(T^k h^0\right)(i)e\right) = Th_i^k + \left(T^k h^0\right)(i).$$

The ith component of this equation yields

$$\left(T^{k+1}h^0\right)(i) = \left(Th_i^k\right)(i) + \left(T^k h^0\right)(i).$$

Subtracting these two relations, we obtain

$$T^{k+1}h^0 - \left(T^{k+1}h^0\right)(i) = Th_i^k - \left(Th_i^k\right)(i),$$

from which

$$h_i^{k+1} = Th_i^k - \left(Th_i^k\right)(i).$$

Similarly, we have

$$T^{k+1}h^0 = T\left(T^k h^0\right) = T\left(\hat{h}^k + \frac{1}{n}\sum\left(T^k h^0\right)(i)e\right) = T\hat{h}^k + \frac{1}{n}\sum\left(T^k h^0\right)(i)e.$$

From this equation, we obtain

$$\frac{1}{n}\sum\left(T^{k+1}h^0\right)(i) = \frac{1}{n}\sum\left(T\hat{h}^k\right)(i) + \frac{1}{n}\sum\left(T^k h^0\right)(i)e.$$

By subtracting these two relations, we obtain

$$h^{k+1} = T\hat{h}^k - \frac{1}{n}\sum\left(T\hat{h}^k\right)(i).$$

The proof for $\tilde{h}^k$ is similar.

(b) We have

$$\hat{h}^k = T^k h^0 - \left(\frac{1}{n}\sum_i\left(T^k h^0\right)(i)\right)e = \frac{1}{n}\sum_{i=1}^n h_i^k.$$

So since h_i^k converges, the same is true for $\hat{h}^k$. Also,

$$\tilde{h}^k = T^k h^0 - \min_i\left(T^k h^0\right)(i)e$$

and

$$\begin{aligned}\tilde{h}^k(j) &= \left(T^k h^0\right)(j) - \min_i\left(T^k h^0\right)(i)\\ &= \max_i\Big(\left(T^k h^0\right)(j) - \left(T^k h^0\right)(i)\Big)\\ &= \max_i h_i^k(j).\end{aligned}$$

Since h_i^k converges, the same is true for $\tilde{h}^k$.

4.8

Bellman's equation for the auxiliary $(1-\beta)$–discounted problem is as follows:

$$\bar{J}(i) = \min_{u \in U(i)} [g(i,u) + (1-\beta) \sum_j \bar{p}_{ij}(u) \bar{J}(j)]. \tag{1}$$

Using the definition of $\bar{p}_{ij}(u)$, we obtain

$$\sum_j \bar{p}_{ij}(u) \bar{J}(j) = \sum_{j \neq t} (1-\beta)^{-1} p_{ij}(u) \cdot \bar{J}(j) + (1-\beta)^{-1}(p_{it}(u) - \beta) \bar{J}(t),$$

or

$$\sum_j \bar{p}_{ij}(u) \bar{J}(j) = \sum_j (1-\beta)^{-1} p_{ij}(u) \bar{J}(j) - (1-\beta)^{-1} \beta \bar{J}(t).$$

This together with (1) leads to

$$\bar{J}(i) = \min_{u \in U(i)} [g(i,u) + \sum_j p_{ij}(u) \bar{J}(j) - \beta \bar{J}(t)],$$

or, equivalently,

$$\beta \bar{J}(t) + \bar{J}(i) = \min_{u \in U(i)} [g(i,u) + \sum_j p_{ij}(u) \bar{J}(j)]. \tag{2}$$

Returning to the problem of minimizing the average cost per stage, we notice that we have to solve the equation

$$\lambda + h(i) = \min_{u \in U(i)} [g(i,u) + \sum_j p_{ij}(u) h(j)]. \tag{3}$$

Using (2), it follows that (3) is satisfied for $\lambda = \beta \bar{J}(t)$ and $h(i) = \bar{J}(i)$ for all i. Thus, by Proposition 2.1, we conclude that $\beta \bar{J}(t)$ is the optimal average cost and $\bar{J}(i)$ is a corresponding differential cost at state i.