# DISCRETE DYNAMIC PROGRAMMING[1]

By David Blackwell

*University of California, Berkeley*

**1. Introduction and summary.** We consider a system with a finite number $S$ of states $s$, labeled by the integers $1, 2, \cdots, S$. Periodically, say once a day, we observe the current state of the system, and then choose an action $a$ from a finite set $A$ of possible actions. As a joint result of the current state $s$ and the chosen action $a$, two things happen: (1) we receive an immediate income $i(s, a)$ and (2) the system moves to a new state $s'$ with the probability of a particular new state $s'$ given by a function $q = q(s' \mid s, a)$. Finally there is specified a discount factor $\beta$, $0 \leqq \beta < 1$, so that the value of unit income $n$ days in the future is $\beta^n$. Our problem is to choose a policy which maximizes our total expected income. This problem, which is an interesting special case of the general dynamic programming problem, has been solved by Howard in his excellent book [3]. The case $\beta = 1$, also studied by Howard, is substantially more difficult. We shall obtain in this case results slightly beyond those of Howard, though still not complete. Our method, which treats $\beta = 1$ as a limiting case of $\beta < 1$, seems rather simpler than Howard's.

**2. Definitions and notation.** Denote by $F$ the (finite) set of functions $f$ from $S$ to $A$. By a *policy* $\pi$, we mean a sequence $\{f_n, n = 1, 2, \cdots\}$ of functions $f_n \varepsilon F$. Using policy $\pi$ means that, if we find the system in state $s$ on the $n$th day, the action chosen that day is $f_n(s)$. For any sequence $g_1, \cdots, g_N$, $g_n \varepsilon F$, and any policy $\pi = \{f_n\}$, we denote by $g_1, \cdots, g_N, \pi$ the policy $\{h_n\}$ with $h_n = g_n$, $1 \leqq n \leqq N$, $h_n = f_{n-N}$, $n > N$. For any $g \varepsilon F$, we denote by $g^{(N)}, \pi$ the policy $\{h_n\}$ with $h_n = g$, $1 \leqq n \leqq N$, $h_n = f_{n-N}$, $n > N$, and by $g^{(\infty)}$ the policy $\{h_n\}$ with $h_n = g$ for all $n$. Finally, we denote by $T\pi$ the policy $\{h_n\}$ with $h_n = f_{n+1}$ for all $n$.

We associate with each $f \varepsilon F$ (1) the $S \times 1$ column vector $r(f)$ whose $s$th element is $i(s, f(s))$, and (2) the $S \times S$ Markov matrix $Q(f)$ whose $(s, s')$ element is $q(s' \mid s, f(s))$. Thus $r(f)$ and $Q(f)$ specify the income and the law of motion, as a function of the current state, on a day when our rule of action is $f$. If we use policy $\pi = \{f_n\}$ and the system is initially in state $s$, the probability that the system will be in state $s'$ at the end of the $n$th day is the $(s, s')$ element of the matrix $Q_n(\pi) = Q(f_1)Q(f_2)\cdots Q(f_n)$. Thus the total expected return from $\pi$ is the column vector

$$V(\pi) = \sum_{n=0}^{\infty} \beta^n Q_n(\pi) r(f_{n+1}),$$

Received September 22, 1961.

[1] This research was supported by the Information Systems Branch of the Office of Naval Research under Contract Nonr 222(53).

where $Q_0(\pi) = I$, the $S \times S$ identity matrix. We have

$$V(\pi) = r(f_1) + \beta Q(f_1) \sum_{n=1}^{\infty} Q_{n-1}(T\pi) r(f_{n+1})$$

$$= r(f_1) + \beta Q(f_1) V(T\pi).$$

We associate with each $f \varepsilon F$ the transformation $L(f)$ which maps the $S \times 1$ column vector $w$ into $L(f)w = r(f) + \beta Q(f)w$. Thus $V(f, \pi) = L(f) V(\pi)$, and $V(f_1, \cdots, f_N, \pi) = L(f_1) \cdots L(f_N) V(\pi)$. For any two column vectors $w_1, w_2$, we write $w_1 \geqq w_2$ if every coordinate of $w_1$ is at least as large as the corresponding coordinate of $w_2$, and $w_1 > w_2$ if $w_1 \geqq w_2$ and $w_1 \neq w_2$. Note that $L(f)$ is *monotone*, i.e., $w_1 \geqq w_2$ implies $L(f)w_1 \geqq L(f)w_2$.

For any two policies $\pi_1, \pi_2$, we write $\pi_1 \geqq \pi_2$ if $V(\pi_1) \geqq V(\pi_2)$, and $\pi_1 > \pi_2$ if $V(\pi_1) > V(\pi_2)$. A policy $\pi^*$ is called *optimal* if $\pi^* \geqq \pi$ for all $\pi$.

**3. Optimal policies for $\beta < 1$.** The methods of this section are familiar to workers in dynamic programming, from the work of Dvoretzky, Kiefer, and Wolfowitz [2], Karlin [4], and Bellman [1].

THEOREM 1. *If $\pi^* \geqq (f, \pi^*)$ for all $f \varepsilon F$, then $\pi^*$ is optimal.*

PROOF. Our hypothesis is that

$$L(f) V(\pi^*) \leqq V(\pi^*) \qquad \text{for all } f \varepsilon F.$$

Then for any policy $\pi = \{f_n\}$, we have $L(f_N) V(\pi^*) \leqq V(\pi^*)$, so that, using the monotoneity of $L(f_1) \cdots L(f_{N-1})$, $L(f_1) \cdots L(f_N) V(\pi^*) \leqq L(f_1) \cdots L(f_{N-1}) V(\pi^*)$, i.e., $(f_1, \cdots, f_N, \pi^*) \leqq (f_1, \cdots, f_{N-1}, \pi^*)$. Thus

$$\pi^* \geqq (f_1, \cdots, f_N, \pi^*)$$

for all $N$, i.e., $V(\pi^*) \geqq V(f_1, \cdots, f_N, \pi^*)$ for all $N$. Letting $N \to \infty$ we obtain $(\beta < 1)$,

$$V(\pi^*) \geqq V(\pi),$$

and the proof is complete.

THEOREM 2. *If $(f, \pi) > \pi$, then $f^{(\infty)} > \pi$.*

PROOF. Our hypothesis is $L(f) V(\pi) > V(\pi)$. Applying the monotone operator $L^{N-1}(f)$ yields

$$L^N(f) V(\pi) \geqq L^{N-1}(f) V(\pi),$$

so that $(f^{(N)}, \pi) \geqq (f, \pi)$ for all $N \geqq 1$. Letting $N \to \infty$ yields $f^{(\infty)} \geqq (f, \pi)$, so that $f^{(\infty)} > \pi$.

Our principal result, describing the Howard policy improvement routine for $\beta < 1$, is

THEOREM 3. *Take any $f \varepsilon F$. For each $s \varepsilon S$ denote by $G(s, f)$ the set of all $a$ for which*

$$i(s, a) + \beta p(s, a) V(f^{(\infty)}) > V_s(f^{(\infty)}),$$

where $p(s, a)$ *is the* $1 \times S$ *row vector whose* $s'$th *coordinate is* $q(s' \mid s, a)$ *and* $V_s(f^{(\infty)})$ *denotes the* $s$th *coordinate of* $V(f^{(\infty)})$. *If* $G(s, f)$ *is empty for all* $s$, *then* $f^{(\infty)}$ *is optimal. For any* $g$ *such that*

(a) $g(s) \varepsilon G(s,f)$ *for some* $s$ *and*

(b) $g(s) = f(s)$ *whenever* $g(s) \not\varepsilon G(s,f)$, *we have* $g^{(\infty)} > f^{(\infty)}$.

PROOF. The $s$th coordinate of $V(g, f^{(\infty)})$ is $i(s, g(s)) + \beta p(s, g(s)) V(f^{(\infty)})$. This will exceed $V_s(f^{(\infty)})$ if and only if $g(s) \varepsilon G(s, f)$, and will equal $V_s(f^{(\infty)})$ if $g(s) = f(s)$. Thus if $G(s,f)$ is empty for all $s, f^{(\infty)} \geqq (g, f^{(\infty)})$, for all $g$ so that, from Theorem 1, $f^{(\infty)}$ is optimal. On the other hand, for any $g$ satisfying (a) and (b), we have $(g, f^{(\infty)}) > f^{(\infty)}$ so that, from Theorem 2, $g^{(\infty)} > f^{(\infty)}$.

Call a policy $\pi = \{f_n\}$ *stationary* if $f_n$ is independent of $n$, i.e., if $\pi = f^{(\infty)}$ for some $f \varepsilon F$. As a consequence of Theorem 3, we have the

COROLLARY. *There is an optimal policy which is stationary.*

PROOF. According to Theorem 3, if we take any stationary policy $f^{(\infty)}$, either it is optimal (case $G(s, f)$ empty for all $s$) or it has a stationary improvement $g^{(\infty)}$ (case $G(s,f)$ nonempty for some $s$). Since there are only finitely many stationary policies, there is one which has no stationary improvement, so that it must be optimal.

**4. Optimal policies for $\beta = 1$.** For the case $\beta = 1$, the total income from a given policy is typically infinite. We may attempt instead to maximize the average rate of income or to find policies which are optimal for all $\beta$ sufficiently near 1. We shall adopt the second approach. Since $\beta$ is now variable, it will sometimes be desirable to exhibit the dependence of $V(\pi)$ and other quantities on $\beta$; thus we shall write $V_\beta(\pi)$ and speak of $\beta$-optimal policies. Denote by $U(\beta)$ the expected total return from a $\beta$-optimal policy. We shall say that a policy $\pi$ is *optimal* if it is $\beta$-optimal for all $\beta$ sufficiently near 1, i.e., if $V_\beta(\pi) = U(\beta)$ for all $\beta$ sufficiently near 1, and shall say that $\pi$ is *nearly optimal* if

$$U(\beta) - V_\beta(\pi) \to 0 \qquad \text{as } \beta \to 1.$$

Our problem is then to find optimal and nearly optimal policies.

We shall need certain known facts about Markov matrices, summarized as

LEMMA 1. *Let* $Q$ *be any* $S \times S$ *Markov matrix.*

(a) *The sequence* $I + Q + \cdots + Q^N/N + 1$ *converges as* $N \to \infty$ *to a Markov matrix* $Q^*$ *such that*

$$QQ^* = Q^*Q = Q^*Q^* = Q^*,$$

(b) *rank* $(I - Q) +$ *rank* $Q^* = S$.

(c) *For every* $S \times 1$ *column vector* $c$, *the system*

$$Qx = x, \qquad Q^*x = Q^*c$$

*has a unique solution.*

(d) $I - (Q - Q^*)$ *is nonsingular, and*

$$H(\beta) = \sum_0^\infty \beta^n(Q^n - Q^*) \to H = (I - Q + Q^*)^{-1} - Q^*$$

*as* $\beta \to 1$.

$$H(\beta)Q^* = Q^*H(\beta) = HQ^* = Q^*H = 0$$

*and*

$$(I - Q)H = H(I - Q) = I - Q^*.$$

These facts may all be found in Kemeny and Snell [5]; we indicate the proof of (d) only.

PROOF OF (d). From (a) we have, for $n > 0$, $Q^n - Q^* = (Q - Q^*)^n$, so that $H(\beta) = \sum_0^\infty \beta^n(Q - Q^*)^n - Q^* = [I - \beta(Q - Q^*)]^{-1} - Q^*$, i.e.,

$$(H(\beta) + Q^*)(I - \beta(Q - Q^*)) = I,$$

i.e.,

$$(H(\beta) + Q^*)(I - Q + Q^*) = I - (1 - \beta)H(\beta)(Q - Q^*). \tag{1}$$

Now $C - 1$ summability of $\{Q^n\}$ to $Q^*$ implies Abel summability of $\{Q^n - Q^*\}$ to $Q$:

$$(1 - \beta)\sum_0^\infty \beta^n(Q^n - Q^*) = (1 - \beta)H(\beta) \to 0 \qquad \text{as } \beta \to 1.$$

Thus the matrix on the right of (1) goes to $I$ as $\beta \to 1$, and $I - Q + Q^*$ is nonsingular. Multiplying (1) by $(I - Q + Q^*)^{-1}$ and letting $\beta \to 1$ yields $H(\beta) + Q^* \to (I - Q + Q^*)^{-1}$ as $\beta \to 1$. Verification of the equalities asserted in (d) is straightforward.

Our results for $\beta = 1$ are summarized as Theorem 4 below. We shall sometimes, to simplify statements, speak of "the policy $f$" when we mean the policy $f^{(\infty)}$. For example, we write $V_\beta(f)$ instead of $V_\beta(f^{(\infty)})$.

THEOREM 4. *Take any $f \in F$ and denote by $Q^*(f)$ the matrix $Q^*$ associated with $Q(f)$. Then*

$$V_\beta(f) = [x(f)/(1 - \beta)] + y(f) + \epsilon(\beta, f), \tag{a}$$

*where $x(f)$ is the unique solution of*

$$(I - Q(f))x = 0, \qquad Q^*(f)x = Q^*(f)r(f),$$

*$y(f)$ is the unique solution of*

$$(I - Q(f))y = r(f) - x(f), \qquad Q^*(f)y = 0,$$

*and $\epsilon(\beta, f) \to 0$ as $\beta \to 1$.*

(b) *For each $s$, denote by $G(s, f)$ the set of $a$ for which either*

$$p(s, a)x(f) > x_s(f)$$

*or*

$$p(s, a)x(f) = x_s(f)$$

*and*

$$i(s, a) + p(s, a)y(f) > x_s(f) + y_s(f),$$

*where* $x_s(f)$, $y_s(f)$ *denote the* sth *coordinates of* $x(f)$, $y(f)$. *For any* $g$ *such that* $g(s) \varepsilon G(s, f)$ *for some* $s$ *and* $g(s) = f(s)$ *whenever* $g(s) \not\varepsilon G(s, f)$, $g > f$ *for all* $\beta$ *sufficiently near* 1.

(c) *For each* $s$, *denote by* $E(s, f)$ *the set of* $a$ *for which*

$$p(s, a)x(f) = x_s(f)$$

*and*

$$i(s, a) + p(s, a)y(f) = x_s(f) + y_s(f)$$

(*always* $f(s) \varepsilon E(s, f)$). *If, for each* $s$, $G(s, f)$ *is empty and* $E(s, f)$ *contains only the point* $f(s)$, *then* $f$ *is optimal.*

(d) *If for each* $s$, $G(s, f)$ *is empty and* $g(s) \varepsilon E(s, f)$ *for all* $s$ *implies*

$$Q^*(g)Q^*(f) = Q^*(g),$$

*then* $f$ *is nearly optimal.*

(e) *For any* $f_0$ *for which* $G(s, f_0)$ *is empty for all* $s$, $x(f_0) \geqq x(g)$ *for all* $g$. *Denote by* $F^*$ *the set of all* $g$ *such that* $x(g) = x(f_0)$. *There is an* $f^* \varepsilon F^*$ *with* $y(f^*) \geqq y(g)$ *for all* $g \varepsilon F^*$. *The nearly optimal* $g$'s *are exactly those for which* $x(g) = x(f^*)$ *and* $y(g) = y(f^*)$.

PROOF. For (a), we have

$$\begin{aligned} V_\beta(f^{(\infty)}) &= [I - \beta Q(f)]^{-1} r(f) = \sum_0^\infty \beta^n Q^n(f) r(f) \\ &= \left( \sum_0^\infty \beta^n Q^*(f) + \sum_0^\infty \beta^n (Q^n(f) - Q^*(f)) \right) r(f) \\ &= \frac{Q^*(f) r(f)}{1 - \beta} + H(f) r(f) + (H(\beta, f) - H(f)) r(f). \end{aligned}$$

Thus (a) is established, with $x(f) = Q^*(f)r(f)$, $y(f) = H(f)r(f)$, and $\epsilon(\beta, f) = (H(\beta, f) - H(f))r(f)$. For the rest of the theorem, we simply calculate $V_\beta(g, f^{(\infty)})$, using the representation (a), and ask when, for $\beta$ near 1, does this exceed $V_\beta(f^{(\infty)})$. We have

$$\begin{aligned} V_\beta(g, f^{(\infty)}) &= r(g) + \beta Q(g) V_\beta(f^{(\infty)}) \\ &= \frac{Q(g)x(f)}{1 - \beta} + r(g) - Q(g)x(f) + Q(g)y(f) + \epsilon_1(\beta, f, g), \end{aligned} \tag{2}$$

where $\epsilon_1(\beta, f, g) = -(1 - \beta)Q(g)y(f) + \beta Q(g)\epsilon(\beta, f) \to 0$ as $\beta \to 1$.

We see that $g(s) \varepsilon G(s, f)$ implies that, for $\beta$ near 1, the $s$th coordinate of $V_\beta(g, f^{(\infty)})$ exceeds that of $V_\beta(f^{(\infty)})$. Since $g(s) = f(s)$ implies equality of the $s$th coordinates of $V_\beta(g, f^{(\infty)})$ and $V_\beta(f^{(\infty)})$ for all $\beta$, we obtain (b) at once from Theorem 3. Similarly, the hypotheses of (c) imply that, for all $\beta$ near 1,

$$V_\beta(g, f^{(\infty)}) \leqq V_\beta(f^{(\infty)})$$

(with strict inequality unless $g = f$), so that from Theorem 3 $f$ is optimal.

For (d) we shall need

LEMMA 2. *For any $f, g \varepsilon F$ for which $g(s) \varepsilon E(s, f)$ for all $s$, we have $x(g) = x(f)$. If in addition $Q^*(g)Q^*(f) = Q^*(g)$, then $y(g) = y(f)$.*

PROOF OF LEMMA 2. That $g(s) \varepsilon E(s, f)$ for all $s$ is equivalent to, writing $x$, $y$ for $x(f)$, $y(f)$,

$$Q(g)x = x \tag{3}$$

and

$$r(g) + Q(g)y = x + y. \tag{4}$$

Multiplying (4) by $Q^*(g)$ yields

$$Q^*(g)r(g) = Q^*(g)x. \tag{5}$$

But (3) and (5) have the unique solution $x = x(g)$, so that $x(g) = x(f)$. Also from $Q^*(f)y = 0$ we obtain $Q^*(g)Q^*(f)y = 0$, so that, if $Q^*(g)Q^*(f) = Q^*(g)$, we obtain

$$Q^*(g)y = 0. \tag{6}$$

But, since $x = x(g)$, the unique solution of (4) and (6) is $y = y(g)$, so that $y(g) = y(f)$.

We return to (d). Let $f$ satisfy the hypotheses of (d), and choose $\beta$ so near 1 that, for any pair $f_1, f_2$, we have $V_\beta(f_1, f_2^{(\infty)}) \geqq V_\beta(f_2^{(\infty)})$ implies $f_1(s) \varepsilon G(s, f_1) \cup E(s, f_1)$ for all $s$. If our $f$ is not $\beta$-optimal, let $f_0 = f_1, f_2, \cdots, f_k$ be a sequence of $\beta$-improvements, obtained as in Theorem 3, terminating in a $\beta$-optimal $f_k$. Then

$$f_{i+1}(s) \varepsilon G(s, f_i) \cup E(s, f_i)$$

for all $i$. We show by induction on $i$ that $x(f_i) = x(f_0)$ and $y(f_i) = y(f_0)$. This is true for $i = 0$. If true for a given $i$, then, since $G(s, f)$, $E(s, f)$ depend only on $x(f)$, $y(f)$, we have $G(s, f_i)$ is empty and $E(s, f_i) = E(s, f)$. Then $f$, $f_{i+1}$ satisfy the hypotheses of $f$, $g$ in Lemma 2, so that $x(f_{i+1}) = x(f)$, $y(f_{i+1}) = y(f)$. Thus, writing $f(\beta)$ for the $\beta$-optimal $f_k$, we have

$$U(\beta) = [x(f)/(1 - \beta)] + y(f) + \epsilon(\beta, f_\beta).$$

Since

$$V_\beta(f^{(\infty)}) = [x(f)/(1 - \beta)] + y(f) + \epsilon(\beta, f),$$

we have $U(\beta) - V_\beta(f^{(\infty)}) \rightarrow 0$ as $\beta \rightarrow 1$, and $f^{(\infty)}$ is nearly optimal.

To establish (e), we obtain from (2), if $G(s, f_0)$ is empty for all $s$, the inequality

$$V_\beta(g, f_0^{(\infty)}) \leqq V_\beta(f_0^{(\infty)}) + \tau(\beta)\delta \qquad \text{for } \beta \text{ near } 1, \tag{7}$$

where $\tau(\beta)$ is a scalar function of $\beta$, the maximum coordinate of $\epsilon_1(\beta, f_0, g)$ −

$\epsilon(\beta, f_0)$, and $\delta$ is the $S \times 1$ column vector with all coordinates unity. We have $\tau(\beta) \to 0$ as $\beta \to 1$. Denoting $L_\beta(g)$ by $L$, we rewrite (7) as $LV_\beta(f_0) \leqq V_\beta(f_0) + \tau(\beta)\delta$ for $\beta$ near 1. We show by induction on $n$ that, for all $n$

$$L^n V_\beta(f_0) \leqq V_\beta(f_0) + (1 + \beta + \cdots + \beta^{n-1})\tau(\beta)\delta \quad \text{for } \beta \text{ near } 1. \tag{8}$$

If (8) holds for a given $n$, we obtain, applying $L$,

$$\begin{aligned} L^{n+1}V_\beta(f_0) &\leqq L[\text{r.h.s. of (8)}] \\ &= r(g) + \beta Q(g)V_\beta(f_0) + \beta(1 + \beta + \cdots + \beta^{n-1})\tau(\beta)\delta, \\ &= V_\beta(g, f_0^{(\infty)}) + \beta(1 + \beta + \cdots + \beta^{n-1})\tau(\beta)\delta \\ &\leqq V_\beta(f_0) + [1 + \beta + \cdots + \beta^n]\tau(\beta)\delta, \end{aligned}$$

where the last inequality is obtained by using (7).

Thus, $L^n V_\beta(f_0) \leqq V_\beta(f_0) + [\tau(\beta)/(1 - \beta)]\delta$ for all $n$, so that, for all $g \varepsilon F$

$$V_\beta(g) = \lim_{n\to\infty} L^n V_\beta(f_0) \leqq V_\beta(f_0) + [\tau(\beta)/(1 - \beta)]\delta \quad \text{for } \beta \text{ near } 1. \tag{9}$$

But

$$V_\beta(g) - V_\beta(f_0) = \frac{x(g) - x(f_0)}{1 - \beta} + y(g) - y(f_0) + \epsilon(\beta, g) - \epsilon(\beta, f_0). \tag{10}$$

(9) and (10) imply $x(g) \leqq x(f_0)$.

Take any $f^*$ which is $\beta$-optimal for a set of $\beta$'s having 1 as a limit point. From (10), with $g = f^*$ we obtain $x(f^*) \geqq x(f_0)$, so that $x(f^*) = x(f_0)$. For any $g \varepsilon F^*$, we have $V_\beta(f^*) - V_\beta(g) = y(f^*) - y(g) + \epsilon(\beta, f^*) - \epsilon(\beta, g)$, so that, letting $\beta \to 1$ through a sequence for which $f^*$ is $\beta$-optimal, we obtain $y(f^*) \geqq y(g)$ for all $g \varepsilon F^*$. The last assertion of (e) is now immediate.

Theorem 4 does not describe an algorithm which is guaranteed to lead to optimal or even near optimal policies, and which is comparable in simplicity to the algorithm described by Theorem 3 for $\beta < 1$. The algorithm is simple until we reach an $f$ for which $G(s, f)$ is empty. At this point, if $E(s, f)$ contains for each $s$ only the single element $f(s)$, $f$ is optimal. If not, we know only that $x(g) \leqq x(f)$ for all $g$, so that we have a policy which maximizes our average return. In one case the verification of (d) is immediate. This is the case in which there is a single terminal state $s^*$ which is certain to be reached eventually, no matter where we start or which policy we use, and which can never be left once reached. In this case for every $g$, $Q^*(g)$ is the matrix with every row the $s^*$ unit vector, so that $f$ will satisfy the hypothesis of (d) and be nearly optimal. In general, the checking of (d) is tedious and, if it fails, we are reduced to determining the set $F^*$, calculating $y(g)$ for each $g \varepsilon F^*$, and selecting a $g$ for which $y(g)$ is maximal.

THEOREM 5. *There is an optimal policy which is stationary.*

PROOF. For each $s$ and $f$, the $s$th coordinate of $V_\beta(f)$ is a rational function of $\beta$, as the representation $V = (I - \beta Q)^{-1}r$ shows. Let $f^*$ be $\beta$-optimal for a set of $\beta$'s having 1 as a limit point. Then, for every $g$, $V_\beta(f^*) \geqq V_\beta(g)$ for a set of $\beta$'s having 1 as a limit point. Since all coordinates of $V_\beta(f^*)$ and $V_\beta(g)$ are rational functions of $\beta$,

$$V_\beta(f^*) \geqq V_\beta(g) \qquad \text{for all } \beta \text{ near } 1.$$

Since this holds for every $g \varepsilon F$, $f^*$ is optimal.

We close with two examples.

EXAMPLE 1. *An $f$ which satisfies the hypotheses of* (d) *of Theorem* 4, *but is not optimal.* There are two states, 1 and 2, and two actions, 1 and 2. In state 1 action 1 yields \$1, and the system remains in state 1 with probability .5 and moves to state 2 with probability .5 while action 2 yields \$2 and the system moves to state 2 with certainty. In state 2, either action yields 0 and the system remains in state 2. There are clearly only two effectively different elements of $F$: $f{:}f(1) = 1$ and $g{:}g(1) = 2$. We have, starting in state 1,

$$V_\beta(f^\infty) = 1 + \tfrac{1}{2}\beta + \tfrac{1}{4}\beta^2 + \cdots = 2/(2 - \beta),$$

$$V_\beta(g^\infty) = 2.$$

Thus, $U(\beta) = 2$ and $f^{(\infty)}$ is nearly optimal but not optimal. The verification that $f$ satisfies the hypotheses of (d) of Theorem 2 is straightforward.

EXAMPLE 2. *An $f$ for which $G(s, f)$ is empty for all $s$, but which is not nearly optimal.* Again there are two states, 1 and 2, and two actions, 1 and 2. In state 1, action 1 yields \$3 and the system remains in state 1 with probability .5. Action 2 yields \$6, and the system moves to state 2. In state 2, either action loses \$3 and the system remains in state 2 with probability .5 and moves to state 1 with probability .5. Again, there are only two effectively different elements of $F$: $f{:}f(1) = 1$ and $g{:}g(1) = 2$. Straightforward calculations yield

$$x(f) = x(g) = \begin{pmatrix}0\\0\end{pmatrix}, \qquad y(f) = \begin{pmatrix}3\\-3\end{pmatrix}, \qquad y(g) = \begin{pmatrix}4\\-2\end{pmatrix},$$

so that

$$V_\beta(g) - V_\beta(f) \to \begin{pmatrix}1\\1\end{pmatrix} \qquad \text{as } \beta \to 1$$

and $f$ is not nearly optimal. The verification that $G(s, f)$ is empty for each $s$ is straightforward.

## REFERENCES


[1] BELLMAN, RICHARD (1957). *Dynamic Programming*. Princeton Univ. Press.

[2] DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1957). The inventory problem, I and II. *Econometrica* **20** 187–222 and 450–466.

[3] HOWARD, RONALD A. (1960). *Dynamic Programming and Markov Processes*. Technology Press and Wiley, New York.

[4] KARLIN, S. (1955). The structure of dynamic programming models. *Naval Research Logistics Quart.* **2** 285–294.

[5] KEMENY, J. G. and SNELL, J. L. *Finite Markov Chains*. Van Nostrand, New York.