APPROXIMATIONS OF DYNAMIC PROGRAMS, II

WARD WHITTI
Yale University and Bell Laboratories

This paper extends a procedure for approximating dynamic programs due to Fox (1971). Here, the monotone contraction operator model of Denardo (1967) is approximated by replacing the state space with a subset and defining two approximate local income functions so that the two associated approximate optimal return functions serve as lower and upper bounds for the original return function in the original model. Conditions are also given implying convergence of a sequence of approximate optimal return functions to the optimal return function in the original model.

1. Introduction and summary. In [8] we introduced a general framework for constructing and analyzing approximations of dynamic programming models. The purpose of this paper is to discuss a different approach, which is an extension of Fox (1971). As in [8], the model is the monotone contraction operator model of Denardo (1967) which includes, for example, Markov decision models with a criterion of discounted present value over an infinite horizon. The approximation is constructed by making the new state space a subset of the original state space. Associated with the smaller state space, we define two approximating local income functions in such a way that the two approximate optimal return functions \( f^- \) and \( f^+ \) are lower and upper bounds for the original optimal return function \( f \), i.e., \( f^- \leq f \leq f^+ \). Moreover, this can be done in such a way that if the approximating state space is enlarged, the new bounds are always at least as good as the old ones, if not better. Finally, we present conditions under which the sequences of optimal return functions associated with a sequence of approximating models converge to the optimal return function in the original model.

Our approximation scheme uses a fixed function to characterize future returns outside the designated subset. The goal is to obtain good decisions for many states inside the subset without examining the behavior outside the subset in detail. We actually discuss three approximation schemes which differ only in the function or functions we use outside the designated subset. Theorems 8-10 apply to an arbitrary function; Theorems 1 and 2 apply to two functions \( e \) and \( g \) with \( e \leq f \leq g \); and Theorems 3-7 apply to two functions \( e \) and \( g \) satisfying an extra monotonicity condition, namely, (2.3).

It turns out that the lower approximation is better behaved than the upper in two important ways. First, with our construction, any policy \( \pi^* \) for the original problem which is an extension of a policy attaining the supremum \( f^- \) over the designated subset in the lower approximate model has a return function \( e_{\pi^*} \) satisfying \( e_{\pi^*} < f^- \); for the upper approximate model we can say only that \( e_{\pi^*} < f^+ \), cf. Corollary to Theorem 3. Second, it is easier to conclude that a sequence of lower bounds \( \{f^-, k \geq 1\} \) converges to \( f \) than a sequence of upper bounds \( \{f^+, k \geq 1\} \), cf.


AMS 1970 subject classification. Primary 90.040.


Key words. Dynamic programming, Markov decision processes, bounds, approximations.

1 This research was initiated in the School of Organization and Management at Yale University and completed at Bell Laboratories.

The author now is associated with Bell Laboratories, Holmdel, New Jersey.

179
§4. However, both functions are important because together they provide a measure of
the error.

Just as with [8], the results here extend to N-stage contraction models, so that corresponding
results exist for finite-stage models with nonstationary strategies.

The paper is organized as follows. First, the monotone contraction operator model is
defined in §2. Two-sided bounds are established in §3. Limit theorems are proved in
§4; and a few concluding remarks are made in §5.

Finally, we mention that related research has recently been reported by White
(1977, 1978). He establishes a priori bounds for both Fox's (1971) approximation and
a successive approximation scheme similar to the one we suggest in Remark 5.1.

2. Monotone contraction operators. We now introduce the monotone contraction
operator model of Denardo (1967), modified to allow for unbounded rewards. For
further discussion of the treatment of unbounded rewards, see van Nunen and
Wessels (1977) and references there. Let the state space $S$ be a nonempty set. For each
state $s \in S$, let the action space be a nonempty set $A_s$. Let the policy space $\Delta$ be the
Cartesian product of the action spaces. Let $\alpha : \Delta \to [0, \infty)$ and $\beta : S \to R$ be given
functions. Let the space $V$ of potential return functions be

$$
V = \left\{ u : S \to \mathbb{R} \mid \sup_{v \in S} |u(s)(\alpha(s) - \beta(s))| < \infty \right\}.
$$

(2.1)

Let $[v] = \sup_{s \in S} |u(s)(s) : s \in S \} for any $v \in R^S$, where $R^S$ is the set of all functions
mapping $S$ into $R$. Let

$$
d(v_0, v_2) = \sup_{v \in S} |u(s)(s) - v(s)(s)) : s \in S \}
$$

(2.2)

for $v_0, v_2 \in V$. This definition makes $(V, d)$ a complete metric space. Let a local
income function $h(t, s, v)$ be defined by assigning a real number to each triple $(s, a, v)$
with $s \in S, a \in A_s$, and $v \in V$. For each $s \in S$, let $H_s$ be a mapping of $V$ into $R^S$
defined by $H_s(x)(s) = h(s, a, v)$. Assume that the functions $H_s$ satisfy the basic
three properties:

(B) There exist constants $K_1$ and $K_2$ such that

$$
|u(s)(H_s - \beta)| \leq K_1 + K_2|u - \beta|,
$$

for all $s \in S$ and $v \in V$.

(M) If $v_1 < v_2$ in $V$, i.e., $v_1(s) < v_2(s)$ for all $s \in S$, $H_v v_1 < H_v v_2$ for all $\delta \in \Delta$.

(C) For a fixed $c, 0 < c < 1$,

$$
d(H_v v_1, H_v v_2) \leq c d(v_1, v_2)
$$

for all $s \in \Delta$ and $v_1, v_2 \in V$.

Property (B) implies that the range of $H_v$ and the ranges of associated extremal
operators are contained in $V$. Property (C) implies that $H_v$ has a unique fixed point,
say $v_\delta$, in $V$ for each $\delta \in \Delta$. The function $v_\delta$ is called the return function for policy $h$
. Let $f(\delta) = \sup_{v \in V} (v_\delta(s) : s \in S)$. The function $f$ is called the maximal return function. Let
the maximal operator $F$ be defined as $(Fv)(\delta) = \sup_{s \in S} (H_v(s)(s) : s \in S)$. As in [1], the
operator $F$ inherits properties B, M, C, and has $f$ as its unique fixed point. Parallelizing
Theorem 1 of [1], we have

$$
d(f, v) \leq (1 - c) d(Fv, v) \text{ for all } \delta \in \Delta.
$$

(2.3)

3. Two-sided bounds. Assume that we have found functions $e$ and $g$ in $V$ such
that $e < f < g$. We shall use the functions $e$ and $g$ together with subsets $S_e$ of
the state space $S$ in order to generate approximate models. Let $\gamma(v_0, v_2, \delta, \delta) \geq 0$ be a function
in $V$ defined by

$$
\gamma(v_0, v_2, \delta, \delta) = \begin{cases} 
\epsilon_v(\delta), & s \in S_e, \\
\epsilon_\delta(\delta), & s \in S - S_e.
\end{cases}
$$

(3.1)
for any \( \varepsilon, \delta \in V \). For any \( \varepsilon \in V \), let \( c = \gamma(\varepsilon, g, S_j) \) and \( c' = \gamma(\varepsilon, \delta, S_j) \). For each \( k \) and \( \delta \in \Delta \), define new return operators \( H_k^\alpha \) and \( H_k^\delta \) on \( V \) by

\[
H_k^\alpha = (H_k \circ c')^\alpha \quad \text{and} \quad H_k^\delta = (H_k \circ c)^\delta.
\]

Also define associated maximal operators \( F_k^\alpha \), defined by \( F_k^\alpha = (f_k \circ c')^\alpha \) or, equivalently, by \( F_k^\alpha = \sup(H_k^\alpha, \delta \in \Delta) \). This is a minor modification of the scheme on page 666 of Fox (1971). While \( H_k^\alpha \) is defined on \( V \) with state space \( S \), it is obviously equivalent to a monotone contraction operator model with state space \( S_k \). It is easy to verify that \( H_k^\alpha \) and \( H_k^\delta \) are operators on \( V \) satisfying \( B, M, C \) for each \( k \) and \( \delta \in \Delta \). Let \( \tau_k^\alpha, \tau_k^\delta, \tau_k^* \) and \( \tau_k^* \) denote the fixed points of \( H_k^\alpha, H_k^\delta, F_k^\alpha \) and \( F_k^\delta \), respectively. Just as with \( f_k^\delta = \sup(\tau_k^\delta, \delta \in \Delta) \). To avoid confusion, let \( f_k^\alpha = \sup(\tau_k^\alpha, \alpha \in A) \).

**Theorem 1.** \( f_k^\alpha < f < f_k^\delta \).

**Proof.** Consider only \( \alpha \). By property (M), \( F_k^\alpha f = (F_k^\alpha)^\alpha \) \( \Rightarrow (F_k^\alpha)^\alpha = (f_k^\alpha) \geq f \). By property (M) plus induction, \( F_k^\alpha f \geq f \) for all \( n \geq 1 \). Since \( d(F_k^\alpha)^n, f_k^\alpha \to 0 \) as \( n \to \infty \), \( f_k^\alpha > f \).

The quality of the bounds obviously depends on the functions \( e \) and \( g \). Bounds for Theorem 1 are always available using \( K_i \) from condition (B).

**Theorem 2.** For all \( \delta \in \Delta \), \( \beta = (1 - c)^{-1}K_i \alpha^{-1} \leq \tau_k^\alpha \leq \beta = (1 - c)^{-1}K_i \alpha^{-1} \).

**Proof.** Using the triangle inequality, (C) and (B) with \( v = \beta \), we obtain

\[
\|a(H_k^\alpha \beta - \beta)\| \leq \sum_{k=1}^{\infty} \|a(H_k^\alpha \beta - H_k^{\alpha-1} \beta)\| \leq \sum_{k=1}^{\infty} e^{k-1} \|a(H_k^\beta \beta - \beta)\| \leq (1 - c)^{-1}K_i \|
\]

so that \( \|a(\tau_k^\alpha - \beta)\| \leq (1 - c)^{-1}K_i \). ■

**Remark.** The standard Markov decision model with discounting in which all one-step rewards are bounded by \( K_1 \) is covered by Theorem 2 with \( \alpha(t) = 1 \) and \( \beta(t) = 0 \) for all \( t \in S \).

To get monotone approximations, we now assume the functions \( e \) and \( g \) satisfy

\[
H_e g \geq \varepsilon \quad \text{for some } \delta \in \Delta \quad \text{and} \quad H_g \varepsilon \leq g \quad \text{for all } \delta \in \Delta.
\]

Such functions \( e \) and \( g \) exist because \( \varepsilon_k \) and \( f \) are possible assignments, but we are not providing a procedure for finding them. Let \( \Delta = \{ \delta \in \Delta : H_e g \geq \varepsilon \} \). Let \( S_1 \) and \( S_2 \) be subsets of \( S \) with \( S_1 \subseteq S_2 \).

**Lemma 1.** Let \( e \) and \( g \) be given satisfying (3.3).

(a) For any \( \varepsilon \in V \) with \( v < g \) and \( \delta \in \Delta, H_e \varepsilon \leq H_{\varepsilon} \varepsilon < \varepsilon \).

(b) For any \( \varepsilon \in V \) with \( v > e \) and \( \delta \in \Delta, e \leq H_{\varepsilon} \varepsilon < H_g \varepsilon \).

**Proof.** (a) Since \( v < g \), \( v' \leq v \leq g \) by (M), \( H_e \varepsilon \leq H_{\varepsilon} \varepsilon < H_g \varepsilon < g \).

Therefore, \( H_e \varepsilon = (H_e \circ \varepsilon)' \leq (H_{\varepsilon} \circ \varepsilon)' = H_{\varepsilon} \varepsilon \) and

\[
H_{\varepsilon} \varepsilon = (H_{\varepsilon} \circ \varepsilon)' \leq (H_{\varepsilon} \circ \varepsilon)' = (H_{\varepsilon})' < g' = g.
\]

(b) The reasoning is similar. ■

Let \( (f_k)^* = \gamma(0, g, S_j), f_k^\alpha = \sup(\tau_k^\alpha, \delta \in \Delta) \) and \( f_\alpha = \sup(\tau_\alpha, \delta \in \Delta) \).
THEOREM 3. Let \( e \) and \( g \) satisfy (3.3).
(a) For each \( \delta \in \Delta, \epsilon_0 < \epsilon_2 < \epsilon_1 < g \).
(b) For each \( \delta \in \Delta, \epsilon_0 < \epsilon_2 < \epsilon_1 < \epsilon_3 \).
(c) \( e < s_0 < s_{a^*} < s_0 < f_1 < s_2 < f_2 < s_{a^*} < g \).

PROOF. (a) By Lemma 1, \( H_{2n} \rho \epsilon_0 < H_{2n+1} \rho \epsilon_0 \). By Property (M) and induction, \( H_{2n} \rho \epsilon_0 < \epsilon_0 \) for all \( n \). Since \( d(H_{2n} \rho \epsilon_0, s_{a^*}) \longrightarrow 0 \) as \( n \rightarrow \infty \), \( \epsilon_0 < \epsilon_1 \). Similarly, \( H_{2n} \rho \epsilon_1 = (H_{2n} \rho \epsilon_0)^{1/2} (H_{2n} \rho \epsilon_0)^{1/2} \) so that \( \epsilon_2 > \epsilon_1 \). Finally, \( H_{2n} \rho \epsilon_2 < g \) by Lemma 1, so that \( \epsilon_2 < g \) by the same reasoning.

REMARK. If \( e < (1 - \delta)^{-1} K_1 \), then for all \( \delta \not\in \Delta \), \( K_1 \not\in \Delta \) in the standard Markovian decision process with discounting in which all one-step rewards are bounded by \( K_1 \) and \( 1 < \delta < 0 \), then \( H_{2n} \rho \epsilon_0 < g \) for all \( \delta \not\in \Delta \). Suppose \( \delta_a^* \) and \( \delta_a^* \) are policies in \( \Delta \) which are extensions of \( \epsilon \) and \( g \) satisfy (3.3) with \( \Delta \not\in \Delta \).

Corollary. Assume (3.3). If \( \delta_a^* \not\in \Delta \), then

\[
s_a^* - e < \epsilon(a) < \epsilon(a) < \epsilon(a) < s_a^* - e, \quad s \in S_a.
\]

4. Convergence. Suppose that we have functions \( e \) and \( g \) satisfying (3.3) and a sequence \( (\epsilon_n) \) (for \( \delta \not\in \Delta \)) and \( (\epsilon_n) \) converge pointwise on \( S \) monotonically to limits \( e^* \) and \( g^* \) such that \( e < e^* < \epsilon_1 < \epsilon_2 < g \). However, we need not have \( e^* = e^* = e^* \); see the first example on page 669 of Fox [2].

THEOREM 4. If \( h(s, \delta(t), \epsilon_a) \rightarrow h(s, \delta(t), \epsilon_a) \) whenever \( \epsilon_a \rightarrow \epsilon \) pointwise monotonically in \( V \), then \( \epsilon_a^* \rightarrow \epsilon^* \) and, if \( \delta \not\in \Delta \), \( \epsilon_a^* \rightarrow \epsilon_a^* \).

PROOF. Consider only \( \delta \). It suffices to show that the unique fixed point of \( H_{0}^* \) is \( \epsilon_a^* \). Since \( \epsilon_a^* = H_{0}^* \epsilon_a^* \), \( \epsilon_a^* \) is \( h(s, \epsilon_a^*), \epsilon_a^* \) for \( s \in S_a \). For any fixed \( s \), \( \epsilon_a^* \rightarrow \epsilon_a^* \) monotonically. By the continuity condition, \( h(s, \epsilon_a^*), \epsilon_a^* \) \rightarrow h(s, \delta(t), \epsilon_a^*). Hence, \( \epsilon_a^* = \epsilon_a^* \) as desired.

REMARKS. The first example on page 669 of Fox [2] illustrates how the continuity is important for treating unbounded rewards.

Let \( f^* \) and \( f^* \) be the pointwise-convergent limits of \( (f_a^*) \) and \( (f_a^*) \).

THEOREM 5. Under the conditions of Theorem 4, \( f^* \).

PROOF. For any \( s \in S_a \) and \( \delta \not\in \Delta \), there exists a \( \delta_a^* \not\in \Delta \) such that \( \epsilon_a(t) \rightarrow f_a(t) < \epsilon \).

By Theorem 3,

\[
f_a(t) > f_a(t) > \epsilon_a(t) > \delta_a(t) - \epsilon_a(t) > \epsilon_a(t) > \delta_a(t) - \epsilon_a(t).
\]

By Theorem 4, \( \epsilon_a(t) - \epsilon_a(t) \rightarrow 0 \) as \( k \rightarrow \infty \), so that \( \lim_{k \rightarrow \infty} \epsilon_a(t) - \epsilon_a(t) < \epsilon \).

Since \( \epsilon \) and \( \epsilon \) were arbitrary, the proof is complete.

REMARK. Obviously Theorem 5 is most useful when \( f_a = f_a \), which certainly occurs when \( \Delta \not\in \Delta \). See the remark following Theorem 3.

The second example on page 669 of Fox [2] shows that \( f^* = f^* \) need not hold without extra conditions. Extra conditions are contained in the easily verified.
Theorem 6. If \( v_{\alpha}(x) \to v_{\Gamma}(x) \) as \( k \to \infty \) uniformly in \( \delta \) for \( \delta \in \Delta \), then \( f^*(x) = f(x) \).

Theorem 7. If, in addition to the condition of Theorem 4, \( \Delta \) is a compact topological space and \( h(s, \delta, x) \) is a continuous function of \( \alpha \), then \( v_{\alpha}(x) \to v_{\Gamma}(x) \) uniformly in \( \delta \), \( \delta \in \Delta \), and \( v_{\alpha}(x) \to v_{\Gamma}(x) \) uniformly in \( \delta \), \( \delta \in \Delta \).

Proof. By Theorem 4 and its proof, for all \( k \) sufficiently large, \( v_{\alpha}(s) - v_{\Gamma}(s) = h(s, \delta, x) - h(s, \delta, x) = 0 \) monotonically. By Tychonoff's theorem, page 166 of Royden (1968), \( \Delta \) is compact. By Dini's theorem, page 162 of Royden (1968), and the new conditions, the convergence is uniform in \( \delta \).

A stronger mode of convergence than pointwise convergence follows from stronger conditions. What is more important, it is no longer necessary for the functions \( e \) and \( g \) to satisfy (3.3) or even \( e \leq f \leq g \). Any function \( w \) in \( V \) can be used outside \( S_k \). We use \( v_{\alpha}(x) \), etc., without \( + \) or \( - \) to indicate that \( w \) is used outside \( S_k \). Let

\[
\omega(\alpha, \delta, k) = \sup_{r \in \mathbb{R}} \{ r(s)(h(s, \delta, x), v_{\alpha}(x)) - h(s, \delta, x, v) \},
\]

(4.1)

Theorem 8. If \( \omega(\alpha, \delta, k) \to 0 \) as \( k \to \infty \) for each \( v \in V \), then \( d(v_{\alpha}(x), v_{\Gamma}(x)) \to 0 \).

Proof. Consider only \( + \). By (2.3),

\[
d(v_{\alpha}(x), v_{\Gamma}(x)) < (1 - c)^{-1} d(H_{\alpha}(x), v_{\alpha}(x)),
\]

\[
< (1 - c)^{-1} \omega(\alpha, \delta, k).
\]

since \( H_{\alpha}(x) = (H_{\alpha}(x), h(x), v) \).

Remark. Obviously \( d(v_{\alpha}(x), v_{\Gamma}(x)) \to 0 \) is not possible.

To obtain corresponding results for the optimal return function, let

\[
\omega(\alpha, \delta, k) = \sup_{r \in \mathbb{R}} \{ r(s)(h(s, \delta, x), v_{\alpha}(x)) - h(s, \delta, x, v) \},
\]

(4.2)

Theorem 9. If \( \omega(\alpha, \delta, k) \to 0 \) as \( k \to \infty \), then \( d(f_{\alpha}(x), f_{\Gamma}(x)) \to 0 \).

Proof. As in the proof of Theorem 8,

\[
d(f_{\alpha}(x), f_{\Gamma}(x)) = \sup_{r \in \mathbb{R}} v_{\alpha}(x), r_{\alpha}(x)) \to 0
\]

With some extra conditions, there is pointwise convergence of \( v_{\alpha} \) to \( v_{\Gamma} \) for each \( \delta \) and pointwise convergence of \( f_{\alpha} \) to \( f_{\Gamma} \), using the arbitrary fixed function \( w \) outside \( S_k \) for each \( k \).

Theorem 10. (a) If

(i) \( S \) is countable,

(ii) \( H_{\alpha} \) maps \( V_{\alpha} \) into itself, where

\[
V_{\alpha} = \{ \phi \in V : \gamma_{\alpha}(x) < \epsilon(x) < \gamma_{\Gamma}(x) \}
\]

and \( \gamma_{\alpha}(x), \gamma_{\Gamma}(x) \) are real-valued functions,

(iii) \( h(s, \delta, x), v_{\alpha} \to v_{\alpha}(s, \delta, x, v) \) as \( n \to \infty \) for each \( s \in S \), \( v \in V \), and sequence \( \{ v_{\alpha}, n \to \infty \} \) in \( V_{\alpha} \) converging pointwise to \( v_{\alpha} \),

then \( v_{\alpha} \) converges pointwise to \( v_{\Gamma} \) as \( k \to \infty \).

(b) If, in addition, (ii) holds for all \( \delta \in \Delta \) and the convergence in (iii) is uniform in \( \delta \), then \( f_{\alpha} \) converges pointwise to \( f \) as \( k \to \infty \).

Proof. (a) By (i) and (ii), \( V_{\alpha} \) with the product topology in a compact metric space.

By (ii), \( v_{\alpha} \in V_{\alpha} \) for all \( k \). Hence, every subsequence of \( \{ v_{\alpha} \} \) has a convergent
subsequence. Let $v_0$ be the limit of some convergent subsequence. By the continuity condition (iii), for $k$ sufficiently large (so that $s \in S_k$), $v_{2k}(s) = h(s, \delta(s), v_{2k}) \to h(s, \delta(s), v_0)$ as $k \to \infty$, while $v_{2k-1}(s) \to v(s)$ as $k \to \infty$, so that $H_{2k} = v_{2k}$. Since $v_0$ is the unique fixed point of $H_2$ in $[v_0, v_0]$, since all convergent subsequences of $(v_{2k})$ have the same limit, the sequence $(v_{2k})$ itself converges to this limit.

(b) Note that
\[
|f_k(s) - f(s)| = \sup_{\delta \in \Delta} |v_{2k}(s) - \sup_{\delta \in \Delta} v_{2k}(s)|
\leq \sup_{\delta \in \Delta} |v_{2k}(s) - v(s)|
= \sup_{\delta \in \Delta} |h(s, \delta(s), v_{2k}) - h(s, \delta(s), v_0)| \to 0.
\]

Remarks. (1) Condition (ii) of Theorem 10 holds for the discounted Markov decision problem with
\[
\gamma = \{ v \in \mathcal{V} : |a(v - \beta)| \leq (1 - c)^{-1} (M_1 + M_2) \}
\]
f if
\[
|a(\pi_s - (1 - c)\beta)| < M_s,
\]
and
\[
|\mathbb{E}_{q_\delta} a^{-}e| = \sup_{s \in \mathcal{S}} \alpha(s) \sum_{j=1}^{\infty} (1/a(j)) q_j(j | s) < c < 1,
\]
where $\pi_s$ is the one-step reward function and $q_\delta$ is the Markov transition kernel, cf. Lemma 3.2.2 of [4] or Theorem 6.1 of [9].

(2) For the discounted Markov decision problem above with $\alpha(s) = 1$ and $\beta(s) = 0$ for all $s$, condition (ii) of Theorem 10 always holds. Convergence uniformly in $\delta$ for Theorem 10(b) obviously holds if
\[
\lim_{m \to \infty} \sup_{s \in \mathcal{S}} \sum_{j=1}^{m} q_j(j | s) = 0. \tag{4.3}
\]
In this case, pointwise convergence of $f_k$ to $f$ has also been established under stronger conditions by D. J. White (1977, 1978) by different methods.

5. Closing remarks. (1) A promising approach is to combine the technique here with successive approximations. For example, we could calculate $F_1^{s} = \ldots = F_1^{s'} = g$, where $F_2^{s} = \sup_{\delta \in \Delta} (H_{2s} \delta) = 1$ and $r_\delta$ is a positive integer or a positive integer-valued stopping time, as described in [4]. Obviously, $f_1^{s} < F_1^{s} = \ldots = F_1^{s'} = g$ and $f_2^{s} < f_2^{s'} < \ldots < f_1^{s} < \ldots < f_1^{s'} < g$.

(2) We could also work with subsets of the action-saturn stochastic games, just as indicated in [8] of Fox (1971). Let the local income function $h(s, \delta(s), \xi(s), e) = \sum_{s \in S} h(s, \delta(s), \xi(s), e(s))$, where $h(s)$ and $\xi(s)$, are probability measures on action spaces $A_1(s)$ and $A_2(s)$. The associated return operator is $[H_{2s} e](s) = h(s, \delta(s), \xi(s), e)$. If the initial bounds $e$ and $g$ satisfy $H_{2g} < g$ and $H_{2e} > e$ for all $\delta$ and $\xi$, then $e < f < f < g$ as in Theorem 3, where $f$ is the value. In order to get $f = f$, it suffices to have $A_1(s)$ and $A_2(s)$ be compact metric spaces and $h(s, a_1, a_2) = \infty$ and $e(s, a_3, a_4, a_5)$ is continuous in $(a_3, a_4, a_5)$. Then, by basic weak convergence theory, the spaces of probability measures on $A_1(s)$ and $A_2(s)$ with the topology of weak
convergence are metrizable as compact metric spaces and \( h(s, \delta_t(x), \xi_t(x), v) \)
\( \rightarrow h(s, \delta(x), \xi(x), v) \) whenever \( \delta_t(x) \rightarrow \delta(x) \) and \( \xi_t(x) \rightarrow \xi(x) \).

Acknowledgments. I am grateful to the referees for their helpful comments and D. J. White for sending me his unpublished papers. Theorem 10 here was added after seeing these papers.

References


BELL LABORATORIES, HOLMDEL, NEW JERSEY 07733