Contraction Mappings in the Theory Underlying Dynamic Programming
Author(s): Eric V. Denardo
Source: *SIAM Review*, Vol. 9, No. 2 (Apr., 1967), pp. 165-177
Published by: Society for Industrial and Applied Mathematics
Stable URL: https://www.jstor.org/stable/2027440
Accessed: 30-07-2019 15:27 UTC

REFERENCES
Linked references are available on JSTOR for this article:
https://www.jstor.org/stable/2027440?seq=1&cid=pdf-reference#references_tab_contents
You may need to log in to JSTOR to access the linked references.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at https://about.jstor.org/terms

Society for Industrial and Applied Mathematics is collaborating with JSTOR to digitize, preserve and extend access to *SIAM Review*

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

SIAM Review
Vol 9, No. 2, April, 1967

CONTRACTION MAPPINGS IN THE THEORY UNDERLYING DYNAMIC PROGRAMMING*

ERIC V. DENARDO†

1. Introduction. This article formulates and analyzes a broad class of optimization problems including many, but not all, dynamic programming problems. A key ingredient of the formulation is the abstraction of three widely shared properties of optimization problems. These properties are called the "contraction," "monotonicity," and "N-stage contraction" properties. The contraction property is satisfied by Shapley's [16] stochastic game and was used by him in a related manner. Other models which satisfy the contraction property include many of Bellman's [1] dynamic programming models, Karlin's [14], Howard's [12] and Blackwell's [2], [3] discounted dynamic programming models, and many of the sequential decision processes in [5]. The N-stage contraction property is a weakened form of the contraction property. It also encompasses some models of Derman [7], Derman and Klein [8], Eaton and Zadeh [9], and many N-stage dynamic programming problems. Each of the models cited above satisfies the monotonicity property. Hence, the formulation encompasses the models of several authors and provides further insight into the class of problems which satisfies Bellman's [1] Principle of Optimality. After completing the development, the author came across a paper of Zachrisson [17] which exploits order-preserving contractions in the analysis of a stochastic game.

"Policies" are introduced, and, for each policy δ, a return function, v_δ, is defined in a natural manner. A maximization operator A is introduced, and A is shown to inherit the contraction property. The fixed-point theorem for contraction mappings assures that the equation $Av = v$ has a unique solution, v^*. An optimal return function f is defined by $f = \sup_\delta v_\delta$. If the monotonicity and contraction assumptions are both satisfied, we conclude that $v^* = f$. Policies whose return functions approximate v^* are shown to exist, and a sufficient condition is provided for the attainment of v^* by some policy. Similar results are obtained for the case in which the monotonicity and N-stage contraction assumptions are both satisfied simultaneously.

Three techniques are provided for determining (or approximating) the fixed point v^* and for finding policies whose return functions attain or approximate v^*. The first is the method of successive approximations of mathematical analysis; it exploits only the contraction assumption. For the case in which both the N-stage contraction and monotonicity properties are satisfied, we provide equivalent mathematical programming formulations and a generalization of one of Howard's [12] policy improvement routines. Certain issues concerning history-remembering decision procedures and randomized policies are resolved.

The five examples in §8 serve as illustrations and applications of the develop-

* Received by the editors August 19, 1966.
† The RAND Corporation, 1700 Main Street, Santa Monica, California 90406.

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

ment. It might help explicate the notation to refer occasionally to the examples, especially Example 1, as one proceeds through the text. Apparent integrability issues in Blackwell's [3] and Jewell's [13] models are circumvented in Examples 3 and 4. Salient facts about metric spaces and contraction mappings comprise the Appendix.

2. A contraction assumption. Some terminology is now introduced. Examples are provided in §8. Let Ω be a set. An element of Ω is called a *point* and is often denoted by x. Associated with each point x is a *decision set* D_x. An element of D_x is called a *decision* and is often denoted by d_x. The *policy space* Δ is defined as the Cartesian product of the decision sets, i.e., $\Delta = \times_{x\in\Omega} D_x$. An element of Δ is called a *policy* and is often denoted by δ. Then, a policy can be interpreted as a decision procedure which specifies a decision for each point. Furthermore, any such combination of decisions constitutes a policy.

In order to introduce the return function, let V be the collection of all bounded functions from Ω to the reals, i.e., $v \in V$ if and only if $v:\Omega \to$ reals and $\sup_{x\in\Omega} | v(x)| < \infty$. A metric ρ on V is defined by $\rho(u, v) = \sup_{x\in\Omega} | u(x) - v(x)|$. The space V is complete in this metric.

Let h, the *return*, be a function ascribing a real number to each triplet (x, d_x, v) with $x \in \Omega$, $d_x \in D_x$ and $v \in V$. One might think of $h(x, d_x, v)$ as the total payoff for "starting" at point x and choosing d_x with the prospect of receiving $v(z)$ if the pair (x, d_x) causes a "transition" to point z. (Whether $v(z)$ could be realized by any policy is immaterial; $h(x, d_x, \cdot)$ describes what the pair (x, d_x) yields as a function of v.) The important contraction assumption is now introduced.

CONTRACTION ASSUMPTION. *For some* c *satisfying* $0 \leqq c < 1$,

$$| h(x, d_x, u) - h(x, d_x, v)| \leqq c\rho(u, v)$$

for each $u \in V$, $v \in V$, $x \in \Omega$ *and* $d_x \in D_x$.

The contraction assumption is satisfied by Shapley's terminating stochastic game [16], by Howard's [12], Blackwell's [3], and Jewell's [13] discounted dynamic programming models, as we verify in §8, and by Bellman's "type 2" equations in [1, Chap. IV]. A slightly weaker version of the assumption, introduced subsequently, also encompasses models of Derman [7], Eaton and Zadeh [9], and certain N-stage sequential decision processes [5], [8].

To verify that a contraction mapping is implicit in the contraction assumption, let δ_x denote the decision in δ which applies to point x. For each $\delta \in \Delta$, a function H_δ having domain V and range assumed to be contained in V is defined by

$$[H_\delta(v)](x) = h(x, \delta_x, v), \tag{1}$$

where $H_\delta(v)$ is the element of V which H_δ assigns v, and where $[H_\delta(v)](x)$ is the real number which the function $H_\delta(v)$ associates with the point x. The contraction assumption is equivalent to the following: for some c satisfying $0 \leqq c < 1$, $\rho[H_\delta u, H_\delta v] \leqq c\rho[u, v]$ for each $u \in V$, $v \in V$ and $\delta \in \Delta$. Hence, H_δ is a contraction mapping [10], and the fixed-point theorem for contraction mappings guarantees that H_δ has a unique fixed point v_δ. That is, for each policy δ there exists a unique element v_δ of V such that

$$v_\delta(x) = h(x, \delta_x, v_\delta) \quad \text{for each} \quad x \in \Omega. \tag{2}$$

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

The function v_δ is called the *return function* of the policy δ. Since we have not assumed that the process has a "starting point" from a definitional viewpoint, an argument like the above is required to preclude the possibility that no v_δ or several v_δ satisfy (2). The *optimum return function* f is defined by $f(x) = \sup_{\delta \in \Delta} v_\delta(x)$. The following simple inequality will prove useful.

THEOREM 1. *Suppose the contraction assumption is satisfied. For any* $\delta \in \Delta$ *and any* $v \in V$, *we have* $\rho(v_\delta, v) \leqq \rho(H_\delta v, v)/(1 - c)$.

Proof. The triangle inequality implies that

$$\rho(H_\delta^n v, v) \leqq \sum_{i=1}^{n} \rho(H_\delta^i v, H_\delta^{i-1} v)$$

$$\leqq \sum_{i=1}^{n} c^{i-1} \rho(H_\delta v, v) \leqq \rho(H_\delta v, v)/(1 - c).$$

Since $\rho(H_\delta^n v, v_\delta) \to 0$, one has $\rho(v_\delta, v) \leqq \rho(H_\delta v, v)/(1 - c)$, the desired result.

3. A maximization operator. Next, a map A having domain V is defined by

$$(Av)(x) = \sup_{d_x \in D_x} h(x, d_x, v) \tag{3}$$

for each $v \in V$ and $x \in \Omega$. We assume that the range of A is contained in V. Theorem 2 verifies that A is a contraction mapping.

THEOREM 2. *Suppose the contraction assumption is satisfied. For each* $u \in V$ *and* $v \in V$, *we have* $\rho(Au, Av) \leqq c\rho(u, v)$.

Proof. Consider arbitrary u, v and x, and write $(Au)(x) = (Av)(x) + k$. Consider the case $k > 0$. For each positive integer n, let d_x^n be an element of D_x such that $h(x, d_x^n, u) \geqq (Au)(x) - k/n$. Clearly, $(Au)(x) - k/n \geqq (Av)(x) \geqq h(x, d_x^n, v)$, the last by definition. Combining inequalities yields

$$0 \leqq (Au)(x) - (Av)(x) - k/n \leqq h(x, d_x^n, u) - h(x, d_x^n, v) \leqq c\rho(u, v).$$

Since the preceding is true for each n, $|(Au)(x) - (Av)(x)| \leqq c\rho(u, v)$; this inequality is trivial for $k = 0$ and similarly established for $k < 0$, completing the proof.

An early and important paper of Shapley [16] was apparently the first to use contraction mappings in a related setting. The fixed-point theorem guarantees that A has a unique fixed point, i.e., that there exists exactly one element v^* of V such that

$$v^*(x) = \sup_{d_x \in D_x} h(x, d_x, v^*) \quad \text{for each} \quad x \in \Omega. \tag{4}$$

Equation (4) is, in rather general notation, a "functional equation" of dynamic programming. Two relevant questions are whether v^* is approximated (or attained) by the return function v_δ of some policy and whether v^* is the optimal return function, i.e., whether $v^* = f$. Existence of policies satisfying $\rho(v_\delta, v^*) \leqq \epsilon$ is demonstrated next. An assumption sufficient for $v^* = f$ is introduced in §4. Without that assumption, one may have $v^* < f$, in which case interpretation of v^* is an open question. The choice of maximization in defining A was arbitrary; Theorem 2 holds with $(Av)(x) = \inf_{d_x} h(x, d_x, v)$.

COROLLARY 1. *For* $\epsilon > 0$, *there exists a policy* δ *such that* $\rho[H_\delta(v^*), v^*] \leqq \epsilon(1 - c)$, *and any such* δ *satisfies* $\rho(v_\delta, v^*) \leqq \epsilon$. *If* $\rho[H_\delta(v^*), v^*] = 0$, *then* $v_\delta = v^*$.

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

Proof. For $\epsilon > 0$, existence of a policy δ such that $\rho[H_\delta(v^*), v^*] \leqq \epsilon(1 - c)$ follows directly from (4). Substituting v^* for v in Theorem 1 yields $\rho(v_\delta, v^*) \leqq \rho[H_\delta(v^*), v^*]/(1 - c) \leqq \epsilon$ for $\epsilon \geqq 0$.

Corollary 2 is established by noting that extrema of continuous functions over compact sets are attained and then applying the last part of Corollary 1. We note that this approach to Corollary 2 circumvents the usual recourse to Tychonoff's theorem.

COROLLARY 2. *Suppose for each fixed x that $h(x, \cdot, v^*)$ is a continuous function of d_x in a topology for which D_x is compact. Then there exists a policy δ such that $v_\delta = v^*$.*

With B as an operator on V, define the modulus of B as the smallest number c such that $\rho(Bu, Bv) \leqq c\rho(u, v)$ for each u, v in V.

With I as an arbitrary nonempty set, suppose $\{B_\alpha : \alpha \in I\}$ is a collection of operators on V each of which has modulus c or less, define the function E having domain V by $(Ev)(x) = \sup_{\alpha \in I} (B_\alpha v)(x)$, and suppose that E has range contained in V. Then an argument similar to that for Theorem 2 establishes that E has modulus c or less.

4. A monotonicity assumption. The assumption given below suffices for v^* and f to be identical. For $u, v \in V$, we write $u \geqq v$ if $u(x) \geqq v(x)$ for each x, and $u > v$ if $u \geqq v$ and $u \neq v$.

MONOTONICITY ASSUMPTION. *If $u \geqq v$, then $H_\delta(u) \geqq H_\delta(v)$ for each $\delta \in \Delta$.*

A monotonicity assumption was introduced by Mitten [15], and monotonicity assumptions were further developed in [5]. A host of return functions, including the examples of §8, satisfy the monotonicity assumption. The assumption is equivalent to $h(x, d_x, u) \geqq h(x, d_x, v)$ if $u \geqq v$.

THEOREM 3. *Suppose the monotonicity and contraction assumptions are satisfied. Then $v^* = f$.*

Proof. From Corollary 1 we know $v^* \leqq f$. Since $H_\delta v^* \leqq v^*$ for each δ, recursive application of the monotonicity assumption yields $H_\delta{}^n v^* \leqq v^*$ for each n. Since $\rho(H_\delta{}^n v^*, v_\delta) \to 0$, one has $v_\delta \leqq v^*$ for each δ. Since $f(x) = \sup_\delta v_\delta(x)$, this implies $f \leqq v^*$ and completes the proof.

Theorem 3 concludes that the solution to (4) is unique and is f, the optimal return function. A policy δ is called *ϵ-optimal* if $\rho(v_\delta, f) \leqq \epsilon$ and *optimal* if $v_\delta = f$. Corollary 1 and Theorem 3 demonstrate existence of an ϵ-optimal policy and Corollary 2 gives sufficient conditions for existence of an optimal policy.

A return function satisfying the contraction assumption and violating the monotonicity assumption is $h(x, d_x, v) = -v(x)/2$.

If the sequence $\{v_n\}$, $n = 0, 1, \cdots$, satisfies $v_n \geqq v_{n-1}$ for each n, we write $\{v_n\} \uparrow$. Lemma 1 contains useful consequences of the monotonicity assumption.

LEMMA 1. *Suppose the monotonicity assumption is satisfied. If $u \geqq v$, then $Au \geqq Av$. If $Av \geqq v$, then $\{A^n v\} \uparrow$. If $H_\delta v \geqq v$, then $\{H_\delta{}^n v\} \uparrow$.*

Proof. By definition, $Au \geqq H_\delta u$. Suppose $u \geqq v$. Then, $H_\delta u \geqq H_\delta v$, implying $Au \geqq H_\delta v$ for each δ; hence, $Au \geqq Av$. If $Av \geqq v$, then recursive application of the preceding statement yields $\{A^n v\} \uparrow$. If $H_\delta v \geqq v$, then recursive application of the monotonicity assumption yields $\{H_\delta{}^n v\} \uparrow$.

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

5. An N-stage contraction assumption. For a map B of V into itself to have a unique fixed point, it suffices that B^N be a contraction mapping for some positive integer N. This suggests a slightly weakened form of the contraction assumption, which we shall use only in conjunction with the monotonicity assumption.

N-STAGE CONTRACTION ASSUMPTION. *For each δ, the operator $H_\delta{}^N$ has modulus c or less, where c and N are independent of δ and $c < 1$. Furthermore, for each δ, H_δ has modulus 1 or less.*

The contraction assumption is satisfied by discounting future returns or by requiring a positive probability of termination from *every* point. For some problems, the probability of termination is nonzero from a proper subset S of Ω; however, every ergodic class of points contains at least one element of S. Most such models satisfy the N-stage contraction assumption. N-stage processes which evolve probabilistically also satisfy the N-stage contraction assumption.

As before, we assume that the range of A is contained in V. Define v_δ as the unique fixed point of the contraction mapping $H_\delta{}^N$; it follows that v_δ is the unique fixed point of H_δ. Since H_δ has modulus 1 or less, the triangle inequality implies $\rho(H_\delta{}^N v, v) \leqq N\rho(H_\delta v, v)$. Hence, as in Theorem 1, $\rho(v_\delta, v) \leqq \rho(H_\delta v, v)N/(1 - c)$.

Next, define f by $f(x) = \sup_\delta v_\delta(x)$. Toward showing $f \in V$, define the function E having domain V by $(Ev)(x) = \sup_\delta (H_\delta{}^N v)(x)$. Since $H_\delta{}^N v \leqq Ev \leqq A^N v$, the last by the monotonicity assumption, E has range contained in V. Then, the N-stage contraction assumption and observation at the end of §3 suffice for E to be a contraction mapping. Let v^* be the unique fixed point of E. Since $H_\delta{}^N v^* \leqq v^*$, one has $v_\delta \leqq v^*$ for each δ, implying $f \leqq v^*$. Hence, $f \in V$. Parts (a)–(c) of the following theorem have just been established. Proof of (d) and (e) is postponed until after Lemma 2.

THEOREM 4. *Suppose the monotonicity and N-stage contraction assumptions are satisfied. Then*:

(a) *v_δ is the unique fixed point of H_δ;*

(b) *$\rho(v_\delta, v) \leqq \rho(H_\delta v, v)N/(1 - c)$;*

(c) *E is a contraction mapping of modulus c or less;*

(d) *f is the unique fixed point of E and of A;*

(e) *if $v \leqq f$, then $\rho(A^N v, f) \leqq c\rho(v, f)$.*

Lemma 2 will prove useful both for Theorem 4 and for the optimization schemes in the next section. We shall prove only (a), since (b) and (c) are obvious. Let $\vec{1}$ be the unit function from Ω to the reals defined by $\vec{1}(x) = 1$ for each x in Ω.

LEMMA 2. *Suppose the monotonicity and N-stage contraction assumptions are satisfied. Then*:

(a) *if $Av \leqq v$, then $v \geqq f$; if $Av \geqq v$, then $v \leqq f$;*

(b) *$Av_\delta \geqq v_\delta$ for each δ in Δ;*

(c) *if $H_\delta v \geqq v$, then $v_\delta \geqq H_\delta v$.*

Proof of Lemma 2. First, suppose $Av \leqq v$. Then, $H_\delta v \leqq v$ for each δ, implying $H_\delta{}^n v \leqq v$ for each n. Hence, $v_\delta \leqq v$ for each δ, implying $v \geqq f$.

The other half of (a) is more difficult. First, since $Af \geqq H_\delta f \geqq H_\delta v_\delta = v_\delta$ for each δ, one has $Af \geqq f$. Suppose $Av \geqq v$. Define u by $u(x) = \max\{v(x), f(x)\}$. Then, $Au \geqq Av$ and $Au \geqq Af$, implying $Au \geqq u$. For arbitrary positive ϵ, pick δ such

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

that $H_\delta u \geqq Au - \epsilon\vec{1}$. Claim: $H_\delta^n(Au) \geqq Au - n\epsilon\vec{1}$. Since $Au \geqq u$, the claim is true for $n = 1$. Suppose it is true for n. Then $H_\delta^{n+1}(Au) \geqq H_\delta(Au - n\epsilon\vec{1}) \geqq H_\delta(Au) - n\epsilon\vec{1} \geqq Au - (n+1)\epsilon\vec{1}$, completing an inductive argument. Since $v_\delta \leqq f \leqq u \leqq Au$, the N-stage contraction assumption assures $H_\delta^N(Au) \leqq v_\delta + c\rho(Au, v_\delta)\vec{1}$. Combining inequalities, $0 \leqq Au - v_\delta \leqq [N\epsilon + c\rho(Au, v_\delta)]\vec{1}$. Suppose $Au > f$. Then take $\epsilon = \rho(Au, v_\delta)(1 - c)/2N$. By substitution, $\rho(Au, v_\delta) \leqq \rho(Au, v_\delta)(1 + c)/2 < \rho(Au, v_\delta)$, a contradiction. Hence, $Au = f$ and $v \leqq f$ as desired.

Proof of Theorem 4. Parts (a)–(c) are established. As noted previously, $Af \geqq f$. Then $A(Af) \geqq Af$ and the first part of Lemma 2 implies $Af \leqq f$. Hence, $Af = f$. If $Ag = g$, then Lemma 2 implies $g = f$; hence f is the unique fixed point of A.

We have established $f \leqq v^*$. Toward showing $v^* = f$, define u by $u(x) = \max_{0 \leqq n < N} (A^n v^*)(x)$. Then, $Au \geqq A^n v^*$ for $1 \leqq n \leqq N$. Since $A^N v^* \geqq Ev^* = v^*$, one has $Au \geqq u$. Hence, by Lemma 2, $u \leqq f$, implying $v^* \leqq f$. Hence $v^* = f$. For (e), note that $A^N v \geqq Ev$. If $v \leqq f$, then $Ev \leqq A^N v \leqq f = Ef$, implying $\rho(A^N v, f) \leqq \rho(Ev, Ef) \leqq c\rho(v, f)$.

Part of the N-stage contraction assumption is that H_δ has modulus 1 or less. Similar results hold if for fixed finite m each H_δ has modulus m or less, the only difference being that $\rho(v_\delta, f) \leqq \rho(H_\delta v, v)(1 + m + \cdots + m^{N-1})/(1 - c)$. An example in which the modulus of H_δ is 2, but the modulus of H_δ^2 is $\frac{1}{2}$, is given by $\Omega = \{1, 2\}$, $D_1 = D_2 = \{0\}$, $u = (u_1, u_2)$, $H_0(u_1, u_2) = (2u_2, u_1/4)$.

We now strengthen the hypothesis of Theorem 4 in a manner which suffices for A^N to be a contraction mapping. First, let π denote a sequence of N policies, i.e., $\pi = (\delta_1, \delta_2, \cdots, \delta_N)$, and define the operator B_π on V as the composition of the N operators H_{δ_i}, i.e., $B_\pi = H_{\delta_N} H_{\delta_{N-1}} \cdots H_{\delta_1}$. We now strengthen the N-stage contraction assumption by replacing the condition that, for each δ, H_δ^N has modulus c or less by the more restrictive condition that for each π the operator B_π has modulus c or less. Note that $(A^N v)(x) = \sup_\pi (B_\pi v)(x)$; hence, by the observation in §3, A^N is a contraction mapping of modulus c or less.

For an example satisfying the N-stage contraction assumption but not its strengthened form, let $\Omega = \{a\}$, $D_a = \{1, 2, \cdots\}$ and $h(a, n, v) = \max\{v/2, \min[n + 1, v - n]\}$. The monotonicity and N-stage contraction assumptions are satisfied with $N = 2$ and $c = \frac{1}{2}$. Also, $Av = \max\{v/2, \max_{1 \leqq n < \infty}[\min(n + 1, v - n)]\}$. Note that $A^n(2^n + 1) - A^n(2^n) = 1$ for each n; in fact, A^n has modulus 1 for every n. R. Strauch of The RAND Corporation first indicated to the author some of the difficulties attendant on the N-stage contraction assumption.

6. Optimization schemes. This section contains three techniques for approximating (or determining) v^* and for finding policies whose returns approximate or attain v^*. One technique is an application of the method of "successive approxitions" [10] of mathematical analysis, the second technique provides general mathematical programming equivalents, and the third generalizes one of Howard's [12] policy improvement routines.

The first technique exploits the contraction assumption, but not the monotonicity assumption. Suppose the contraction property is satisfied. Theorem 2

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

implies that $\rho(A^n v, v^*) \to 0$ for any v; this means that v^* can be approximated by successive applications of A to any initial vector v_0. Let $v_n = A^n v_0$ for $n \geqq 1$. We now compute a bound on $\rho(v_n, v^*)$. Since $\rho(v_m, v^*) \to 0$, one has $\rho(v_n, v^*) \leqq \sum_{i=1}^{\infty} \rho(v_{n+i}, v_{n+i-1}) \leqq \sum_{i=0}^{\infty} c^i \rho(v_{n+1}, v_n) = \rho(v_{n+1}, v_n)/(1 - c)$. Since $\rho(v_{n+k}, v^*) \leqq c^k \rho(v_n, v^*)$, one has $\rho(v_n, v^*) \leqq \min_{1 \leqq i \leqq n} \{c^{n-i+1} \rho(v_{i-1}, v^*)\}$. Define E_n recursively by $E_1 = \rho(v_1, v_0)/(1 - c)$ and $E_n = \min \{cE_{n-1}, \rho(v_n, v_{n-1})c/(1 - c)\}$. Combining the above inequalities yields $\rho(v_n, v^*) \leqq E_n$ for $n \geqq 1$. Suppose $v_n = H_\delta v_{n-1}$. Then, $\rho(v_\delta, v^*) \leqq \rho(v_\delta, v_{n-1}) + \rho(v_{n-1}, v^*) \leqq 2\rho(v_n, v_{n-1})c/(1 - c)$. Charnes and Schroeder [4] suggest bounds along these lines for a stochastic game. If the N-stage contraction (for $N > 1$) and monotonicity assumption are satisfied, Theorem 4 assures $\rho(v_n, f) \to 0$, providing $v_0 \leqq f$. Methods similar to the above yield bounds for $\rho(v_n, f)$ and $\rho(v_\delta, f)$ in this case as well.

If the monotonicity and N-stage contraction assumptions are both satisfied, policy improvement schemes and mathematical programming formulations are available. In the ensuing, interpret "min v" ("max v") as the function whose value at x is the smallest (largest) value of $v(x)$ over those v's satisfying the constraint. Consider the following two mathematical programs:

Program I	*Program II*
min v	max v
subject to	subject to
$Av \leqq v$.	$Av \geqq v$.

Since $Af = f$, f is feasible for both programs. By part (a) of Lemma 2, f is optimal for both programs. Of course, if Ω contains finitely many points, "min v" is equivalent to minimizing $\sum_{x \in \Omega} v(x)$ or any other positive combination of the $v(x)$'s. A linear programming formulation for Example 1 (§8) was first obtained by D'Epenoux [6]. Derman and Klein [8] obtained a programming formulation for an N-stage Markovian process, also a special case of the above.

Though the two programs look similar, Program I has an inherent advantage. Its constraint is satisfied if and only if $h(x, d_x, v) \leqq v(x)$ for each x and d_x, while Program II's constraint is satisfied if for every x one has $h(x, d_x, v) \geqq v(x)$ for at least one d_x. Hence, Program I can be written equivalently as "min v subject to $h(x, d_x, v) \leqq v(x)$ for each x and d_x."

Next, Howard's policy improvement algorithm is generalized. Suppose the monotonicity and N-stage contraction assumptions are satisfied and that Av is attained for each v; i.e., $Av = H_\gamma v$ for some policy γ which may depend on v. The "n" in the policy improvement routine given below may be any positive integer.

(1) Pick any initial δ.
(2) Calculate v_δ.
(3) Calculate $u = A^{n-1} v_\delta$, then $v = Au = H_\gamma u$.
(4) If $\rho(u, v) > \epsilon$, replace δ by γ and go to Step 2. If $\rho(u, v) \leqq \epsilon$, calculate v_γ and stop.

Lemmas 1 and 2 imply $v_\delta \leqq A^{n-1} v_\delta = u \leqq Au = H_\gamma u \leqq v_\gamma$. If $N = 1$, i.e., if the contraction property is satisfied, then $\rho(v_\gamma, f) \leqq c^n \rho(v_\delta, f)$ by Theorem 2.

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

If $N > 1$, one starts with δ and iterates Steps 2 through 4 enough times to assure N applications of A; the resulting policy, ξ, satisfies $\rho(v_\xi, f) \leqq c\rho(v_\delta, f)$. Hence, $\rho(v_\delta, f) \to 0$. Then, the process terminates in a finite number of iterations. The final policy γ satisfies $v_\gamma \geqq u$, implying $\rho(v_\gamma, f) \leqq \rho(u, f) \leqq \epsilon N/(1 - c)$. Howard's policy improvement routine is the above with $n = 1, \epsilon = 0$ and applied to Example 1. Calculating v_δ is called "policy evaluation", and finding γ is called "policy improvement." If calculating Av is far quicker than determining v_δ, then setting $n > 1$ might improve the speed of the algorithm; n might also vary (adapt) during the progress of the algorithm. The amount of computation required to compute v_δ may depend on the value of n chosen; for example, in Howard's problem, more pivoting may be required for $n > 1$. In Example 2, calculating v_δ amounts to solving Example 1 and determining Av requires solving M zero-sum, two-person games.

7. Symmetries. Somewhat loosely, a "history" of a point x is a sequence of prior points, decisions and transitions which ended with a transition to point x. A "history-remembering" decision procedure is one in which the decision d_x selected for x can depend on the history of x. Can history-remembering decision procedures increase the optimal return function? Theorem 5 answers "No" to this question in rather a general setting. We comment that while this result is intuitively clear for Example 1, one's intuition may be less certain for stochastic games (Example 2) in which the ergodic classes depend on the strategies chosen. History-remembering decision procedures were considered in [3], [5], [7], [8], and [16].

The history-remembering decision procedure may be thought of as resulting in a separate problem which contains considerable internal symmetry. Certain processes in which history is not remembered still contain considerable internal symmetry. For instance, Example 4 "looks the same" from points (i, t) and (i, t') differing only in time. Theorem 5 also demonstrates that Example 4 is essentially the same as Example 1.

Consider two optimization problems—problem U and problem P—of the type we are discussing. Problem U is described using unprimed notation—e.g., Ω, D_x, h, etc., and problem P is described using primed notation—e.g., Ω', D_x', h', etc. Suppose $\Omega' = \bigcup_{x\in\Omega} E_x$, where $\{E_x\}_{x\in\Omega}$ is a collection of nonempty pairwise disjoint subsets of Ω'. Roughly, E_x is the subset of Ω' containing those points which are "equivalent" to x; it may correspond to the various histories of x. Let e be the map of V into V' defined by $[e(v)](z) = v(x)$ for each $z \in E_x$ and each x. Roughly, the function e maps v into that function whose value is $v(x)$ at each point which is equivalent to x. Problem P is said to be *generated* from problem U if, in addition to the above, (i) $D_z' = D_x$ for each $z \in E_x$ and each x, and (ii) $h'[z, d_x, e(v)] = h(x, d_x, v)$ for each $z \in E_x$, each x, d_x, and v. It is convenient to introduce the map ω of Δ into Δ' defined by $[\omega(\delta)]_z = \delta_x$ for each $z \in E_x$, each x and each δ. The proof of Theorem 5 is routine and is omitted.

THEOREM 5. *Let problem P be generated from problem U and suppose problem P satisfies either* (a) *the contraction assumption or* (b) *both the N-stage contraction and the monotonicity assumptions. Then, problem U satisfies the same assumption*(*s*)

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

and has a unique fixed point v^*. *Furthermore,* $e(v^*)$ *is the unique fixed point of problem* P, *and* $\rho[v_{\omega(\delta)}, e(v^*)] = \rho(v_\delta, v^*)$ *for each* $\delta \in \Delta$.

In other words, the fixed points for problems U and P are essentially the same, and a policy δ whose return approximates or attains v^* has an equivalent policy $\omega(\delta)$ whose return approximates or attains $e(v^*)$. Thus, for the purpose of determining fixed points and ϵ-optimal policies, one can study problem U rather than the more complex problem P.

A pair of related problems also results from considering randomized (unprimed) and nonrandomized (primed) decisions. In this case, one has $\Omega = \Omega'$, $V = V'$ and $D_x' \subset D_x$, since the nonrandomized decisions are a subset of randomized ones. Suppose, as is the case in [3], [5], [7], and [8], that for each d_x in D_x and each v in V there exist $d_x^- \in D_x'$ and $d_x^+ \in D_x'$ such that $h(x, d_x^-, v) \leqq h(x, d_x, v) \leqq h(x, d_x^+, v)$. We note specifically that this condition does not hold for zero-sum, two-person games, e.g., Example 2. If the condition does hold, then $A = A'$, implying that the fixed points of the primed and unprimed problems are identical.

Suppose the N-stage contraction and monotonicity properties are satisfied. Let Δ^* be the collection (perhaps empty) of all optimal policies, i.e., $\delta \in \Delta^*$ if and only if $v_\delta = f$. Let $D_x^* = \{d_x \in D_x : f(x) = h(x, d_x, f)\}$ and let $\Delta^+ = \times_{x\in\Omega} D_x^*$. We now show that $\Delta^* = \Delta^+$. If $\delta \in \Delta^+$, then $H_\delta f = f$, implying $v_\delta = f$; hence, $\Delta^+ \subset \Delta^*$. If $\delta \notin \Delta^+$, then $H_\delta f < f$, implying $v_\delta < f$; hence, $\Delta^+ = \Delta^*$. Should Δ^* contain several policies, one can use this result to optimize on a secondary criterion with D_x replaced by D_x^*.

8. Examples. As illustrations of various aspects of the development, models of Howard [12], Shapley [16], Blackwell [3], Jewell [13], and Fox [11] are now reviewed. These examples by no means exhaust the possibilities; the theory covers many finite-stage processes, nonstationary processes, variations and combinations of the above models, and a diverse array of other return functions.

Example 1 (Howard's [12] Infinite-Horizon Discounted Model). The state space, Ω, consists of the first n integers. The decisions available at state (point) i constitute a finite set and are numbered consecutively, $1, 2, \cdots, M_i$. A policy is then an n-tuple $(k_1, k_2, \cdots, k_n)$, where k_i is the decision relating to state i and $1 \leqq k_i \leqq M_i$. The process has an immediate reward $r(i, k)$ depending on the state and the decision, a discount factor c $(0 \leqq c < 1)$ affecting future returns, and a set of transition probabilities governing the evolution of the process. Let $P[j:i, k]$ be the probability of a transition to state j given state i and decision k. Hence, $\Omega = \{1, 2, \cdots, n\}$, $D_i = \{1, 2, \cdots, M_i\}$, and

$$h(i, k, v) = r(i, k) + c \sum_{j=1}^{n} P[j:i, k]v(j).$$

The contraction assumption is satisfied, since

$$\begin{aligned} |h(i, k, u) - h(i, k, v)| &= c \left| \sum_{j=1}^{n} P[j:i, k]\cdot[u(j) - v(j)] \right| \\ &\leqq c \sum_{j=1}^{n} P[j:i, k]\cdot|u(j) - v(j)| \leqq c\rho(u, v). \end{aligned}$$

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

If $u \geqq v$, then $c \sum_{j=1}^{n} P[j:i, k]u(j) \geqq c \sum_{j=1}^{n} P[j:i, k]v(j)$, implying $h(i, k, u) \geqq h(i, k, v)$, or that the monotonicity assumption is satisfied.

Example 2 (Shapley's [16] Stochastic Game). Let $\Omega = \{1, 2, \cdots, M\}$, where point n is thought of as the nth of a collection of M zero-sum two-person rectangular games. Play moves from game to game, eventually terminating, with a payoff from Player II to Player I at each play. Given that game n is being played, the strategies chosen by Players I and II determine (i) the expected value of the immediate payoff from Player II to Player I, and (ii) the probability law determining whether another game will be played and, if so, which game will be played next. Player I is presumed to maximize his minimum gain; Player II minimizes his maximum loss.

Let r_{ij}^{n} be the immediate payoff from Player II to Player I if they play game n and choose pure strategies j and i, respectively, with $1 \leqq i \leqq M_n$ and $1 \leqq j \leqq m_n$. Let $p[m:i, j, n]$ be the probability that the next game played is game m, given that game n is now played, and that pure strategies i and j are chosen by Players I and II, respectively. Assume $\sum_{m=1}^{M} p[m:i, j, n] \leqq c < 1$ for each i, j, and n. Let p^n (q^n) be a randomized strategy for Player I (II) for game n, where $p_i{}^n$ is, for instance, the probability that Player I chooses pure strategy i for game n and where $\sum_{i=1}^{M_n} p_i{}^n = 1$. With strategies p^n and q^n for game n, and with v as the terminating reward function, the one-stage expected return function h is given by

$$h(n, p^n, q^n, v) = \sum_{i,j} p_i{}^n q_j{}^n \{r_{ij}^{n} + \sum_{m} p[m:i, j, n] \cdot v(m)\}.$$

Note that with fixed n and v, the above is the payoff function of a rectangular game whose i,jth entry is the term in the brackets, "{ }". Hence, by the minimax theorem for rectangular games, an operator B on V is defined by

$$(Bv)(n) = \max_{p} \min_{q} h(n, p, q, v) = \min_{q} \max_{p} h(n, p, q, v).$$

Let D_n and O_n be the sets of all randomizations over $\{1, 2, \cdots, M_n\}$ and $\{1, 2, \cdots, m_n\}$, respectively. Let $\Delta = \times_{n=1}^{M} D_n$ and $\Pi = \times_{n=1}^{M} O_n$, with $\delta \in \Delta$ and $\pi \in \Pi$. (Then, δ is an M-tuple of probability distributions.) Define the operator $H_{\delta,\pi}$ on V by $[H_{\delta,\pi}(v)](n) = h(n, \delta_n, \pi_n, v)$. For each fixed δ and π, $H_{\delta,\pi}$ obeys the contraction assumption and, hence, is a contraction mapping; let $v_{\delta,\pi}$ be its fixed point. Define H_δ by $(H_\delta v)(n) = \min_\pi [H_{\delta,\pi}(v)](n)$. Theorem 2, though minimizing instead of maximizing, guarantees that H_δ satisfies the contraction assumption for each δ. Then, noting that $(Bv)(n) = \max_\delta (H_\delta v)(n)$ and reapplying Theorem 2 assures that B has a unique fixed point. Clearly, $H_{\delta,\pi}$ satisfies the monotonicity assumption. Then, f is the unique solution of the following:

(a) $v(n) = \max_p \min_q h(n, p, q, v)$ for each n,
(b) $v(n) = \min_q \max_p h(n, p, q, v)$ for each n,
(c) $f = \max_\delta \min_\pi v_{\delta,\pi}$, and
(d) $f = \min_\pi \max_\delta v_{\delta,\pi}$.

Two applications of Theorem 5 settle issues concerning history-remembering policies. In the policy improvement routine in §6, calculating v_δ amounts to

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

solving Example 1, and calculating Au amounts to finding the values of each of M rectangular games.

Example 3 (Blackwell's [3] Discounted Dynamic Programming Model). This example differs from Shapley's model in that Player II is a dummy (plays a fixed strategy) and in that Ω need not be finite or even countable. Blackwell [3] deals directly with certain troublesome integrability issues; we circumvent these by introducing an operator ψ which is the same as the integral if the integral exists and which obeys the contraction and monotonicity assumptions whether or not the integral exists. Let $r(x, d_x)$ be the expectation of the immediate return associated with making decision d_x at point x and let $\mu[\cdot \mid x, d_x]$ be the probability measure over Ω determined by the pair (x, d_x), with $\mu[\Omega \mid x, d_x] = 1$. With $0 \leqq c < 1$ and with $K(v) = \{u \in V : u \geqq v, u \text{ integrable}\}$, let

$$h(x, d_x, v) = r(x, d_x) + c\psi(v : x, d_x),$$

$$\psi(v : x, d_x) = \inf_{u \in K(v)} \int_{z \in \Omega} u(z)\, d\mu[z \mid x, d_x].$$

The property of ψ which we shall verify and then exploit is that for each fixed x and d_x, ψ is subadditive, i.e., $\psi(u + v : x, d_x) \leqq \psi(u : x, d_x) + \psi(v : x, d_x)$. Fixing x and d_x for the remainder of the discussion, we abbreviate $\psi(u : x, d_x)$ by the symbol $\psi(u)$. As defined, $K(u + v) \supset \{u' + v' : u' \geqq u, v' \geqq v, u' \text{ integrable}, v' \text{ integrable}\}$. Hence, by set inclusion,

$$\psi(u + v) \leqq \inf_{u' \in K(u), v' \in K(v)} \int_{z \in \Omega} [u'(z) + v'(z)]\, d\mu[z \mid x, d_x] = \psi(u) + \psi(v).$$

Then, $\psi[(u - v) + v] \leqq \psi(u - v) + \psi(v)$, or $\psi(u) - \psi(v) \leqq \psi(u - v)$. As ψ is defined, it may be that $\psi(u) \neq -\psi(-u)$. However, $|\psi(u)| \leqq \sup_{x \in \Omega} |u(x)|$. Then,

$$|\psi(u) - \psi(v)| \leqq \max\{|\psi(u - v)|, |\psi(v - u)|\} \leqq \rho(u, v),$$

$$|h(x, d_x, u) - h(x, d_x, v)| = c\,|\psi(u) - \psi(v)| \leqq c\rho(u, v),$$

verifying that the contraction assumption is satisfied. The monotonicity assumption is routinely verified. Note that f might not be measurable and that the same definition of ψ works for both maximizing and minimizing problems.

Example 4 (Jewell's [13] Continuous-Time Infinite-Horizon Discounted Model). This example differs from the preceding three in that the process evolves in continuous time; other than that, it is similar to Example 1. The function "ψ" is again used to circumvent some integrability problems, and the model is shown to be generated from Example 1, which has no integrability problems.

Let $\Omega' = \{1, 2, \cdots, n\} \times \text{reals}$ and $x' = (i, t)$, with $1 \leqq i \leqq n$ and $t \in \text{reals}$. Let $D_i' = D'_{(i,t)} = \{1, 2, \cdots, m_i\}$, independent of t. Selection of decision k from D_i' at point (i, t) causes a transition to some state at some time greater than t. It is convenient to represent this phenomenon by two random variables. Let $P[X_{i,k} = j]$ be the probability that transition occurs to state j given that decision k is made at state i. Let $P[X_{i,j,k} < u]$ be the probability that the interval of time

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

to transition is less than u, given that the decision k is made at state i and given the condition that transition will occur to state j. Both $X_{i,k}$ and $X_{i,j,k}$ are independent of t; the present value at time t (not time 0) of the return from time t on is

$$h'[(i, t), k, v'] = r(i, k) + \sum_{j=1}^{n} P[X_{i,k} = j] \int_{u=0}^{\infty} v'(j, t + u)e^{-\alpha u}\, dP[X_{i,j,k} < u].$$

The above expression is technically correct only if v' is integrable. In general, we replace "$\int$" by "ψ" and assume $\int dP[X_{i,j,k} < u]\cdot e^{-\alpha u} \leqq c < 1$ for each i, j, and k. The contraction assumption is then satisfied, and the monotonicity assumption is again routinely verified.

One can readily check that Example 4 is generated from Example 1 with $\Omega = \{1, 2, \cdots, n\}, D_i = \{1, 2, \cdots, m_i\}, P[j : i, k] = P[X_{i,k} = j]\cdot \int dP[X_{i,j,k} < u] \cdot e^{-\alpha u}$, $E_i = \{(i, t) : t \in \text{reals}\}$, and $r(i, k)$ unchanged. Thus, Theorem 5 allows us to investigate the optimization problem in a far simpler environment having a finite number of points and decisions and no integrability difficulties. Then, Corollary 2 guarantees existence of a stationary optimal policy for Example 4.

Example 5 (Fox's [11] Age Replacement with Discounting). For a particularly simple example in which D_x is nonfinite, we consider a model of "age replacement" due to Fox [11]. An item (e.g., a light bulb) is replaced at the earlier of (a) a planned replacement interval, d, after installation (at cost c_1) and (b) the time at which it fails (at cost c_2). The costs are incurred at the replacement times. Restricting ourselves to planned replacement intervals of at least ϵ, the problem is generated from one with $\Omega = \{1\}, D_1 = [\epsilon, \infty], F$ as the failure-time distribution, h as the present value of the income stream, α as the discount rate, v as a real number, and

$$h(1, d, v) = (c_1 + v)e^{-\alpha d}(1 - F(d)) + (c_2 + v) \int_0^d e^{-\alpha u}\, dF(u)$$

for $\epsilon \leqq d \leqq +\infty$. The monotonicity assumption is obviously satisfied, and the contraction assumption is readily verified if $F(0^+) < 1$ and $\epsilon > 0$. Substituting $v(d)$ both for $h(1, d, v)$ and for v in the above yields a function of one variable, d, which can be minimized by the method of successive approximations; for details, see Fox [11]. Subsequently, we shall remove the restriction $\epsilon > 0$.

Acknowledgment. Part of the material contained in this article was developed at Northwestern University under the very stimulating direction of Dr. L. G. Mitten, the author's Ph.D. thesis adviser.

Appendix. Salient facts about contraction mappings are now listed. Proofs can be found in Ẽlsgol'c [10] and many standard texts on analysis.

First, some terms are defined. Consider a set V. A function ρ mapping $V \times V$ to the reals is called a *metric* if (i) $\rho(u, v) \geqq 0$ for all $u, v \in V$; (ii) $\rho(u, v) = 0$ if and only if $u = v$; and (iii) $\rho(u, v) \leqq \rho(u, w) + \rho(w, v)$ for u, v and w in V. If

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms

ρ is a metric for V, then V is called a *metric space*. Let $A: V \to V$. The function A is called a *contraction mapping* if for some c satisfying $0 \leqq c < 1$ one has $\rho(Au, Au) \leqq c\rho(u, v)$ for every $u, v \in V$. The element v^* of V is called a *fixed point* of A if $Av^* = v^*$. A sequence $\{v_n\}$, $n = 1, 2, \cdots$, of elements of V is called a *Cauchy sequence* if for every $\epsilon > 0$ there exists an M such that $\rho(v_m, v_n) < \epsilon$ for every $m, n > M$. A metric space is said to be *complete* if for every Cauchy sequence $\{v_n\}$, $n = 1, 2, \cdots$, there exists an element v of V such that $\lim_{n\to\infty} \rho(v_n, v) = 0$. For a map A of V into itself the function A^n is defined recursively by $A^1 = A$ and $A^{n+1} = A(A^n)$.

FIXED-POINT THEOREM. *Let V be a complete metric space. Suppose for some integer N that A^N is a contraction mapping. Then A has a unique fixed point, v^*. Furthermore, $\lim_{n\to\infty} \rho(A^n v, v^*) = 0$ for any v.*

REFERENCES

[1] R. BELLMAN, *Dynamic Programming*, Princeton University Press, Princeton, 1957.

[2] D. BLACKWELL, *Discrete dynamic programming*, Ann. Math. Statist., 33 (1962), pp. 719–726.

[3] ———, *Discounted dynamic programming*, Ibid., 36 (1965), pp. 226–235.

[4] A. CHARNES AND R. G. SCHROEDER, *On some tactical antisubmarine games*, Systems Research Memorandum No. 131, The Technological Institute, Northwestern University, Evanston, Illinois, 1965.

[5] E. V. DENARDO, *Sequential decision processes*, Doctoral thesis, Northwestern University, Evanston, Illinois, 1965.

[6] F. D'EPENOUX, *Sur un problème de production de stockage dans l'aleatoire*, Rev. Française Recherche Operationelle, 14 (1960), pp. 3–16.

[7] C. DERMAN, *On sequential decisions and Markov chains*, Management Sci., 9 (1962), pp. 16–24.

[8] C. DERMAN AND M. KLEIN, *Some remarks on finite horizon Markovian decision models*, Operations Res., 13 (1965), pp. 272–278.

[9] J. H. EATON AND L. A. ZADEH, *Optimal pursuit strategies in discrete-state probabilistic systems*, Trans. ASME Ser. D. J. Basic Engrg., 82 (1962), pp. 23–29.

[10] L. È. ÈLSGOL'C, *Qualitative Methods in Mathematical Analysis*, Trans. by A. A. Brown and J. M. Danskin, American Mathematical Society, Providence, 1964.

[11] B. FOX, *Age replacement with discounting*, Operations Res., to appear.

[12] R. A. HOWARD, *Dynamic Programming and Markov Processes*, Technology Press of M.I.T., Cambridge, 1960.

[13] W. S. JEWELL, *Markov-renewal programming. I: Formulation, finite return models. II: Infinite return models, example*, Operations Res., 11 (1963), pp. 938–948, 949–971.

[14] S. KARLIN, *The structure of dynamic programming models*, Naval Res. Logist. Quart., 2 (1955), pp. 285–294.

[15] L. G. MITTEN, *Composition principles for synthesis of optimal multi-stage processes*, Operations Res., 12 (1964), pp. 610–619.

[16] L. S. SHAPLEY, *Stochastic games*, Proc. Nat. Acad. Sci. U.S.A., 39 (1953), pp. 1095–1100.

[17] L. E. ZACHRISSON, *Markov games*, Advances in Game Theory, M. Dresher, L. S. Shapley and A. W. Tucker, eds., Princeton University Press, Princeton, 1964, pp. 211–253.

This content downloaded from 142.244.5.18 on Tue, 30 Jul 2019 15:27:18 UTC
All use subject to https://about.jstor.org/terms