MATHEMATICS OF OPERATIONS RESEARCH
Vol. 13, No. 1, February 1988
Printed in U.S.A.

DISCRETE EVENT DYNAMIC PROGRAMMING WITH SIMULTANEOUS EVENTS*†

PIERRE L'ECUYER‡ AND ALAIN HAURIE§

This paper deals with an infinite-horizon discrete-event dynamic programming model with discounting, and with Borel state and action spaces. Instead of the usual n-stage contraction assumption [4], uniform over all admissible state-action pairs, we propose milder conditions, sufficient for regularity, and allowing any number of simultaneous events. This model permits one to treat properly a number of problems typically associated with continuous-time maintenance models [5, 6, 11, 12].

The main results concern the uniform convergence of the dynamic programming (DP) procedure to the optimal cost-to-go function, the existence of an ϵ-optimal policy for any $\epsilon > 0$, and a set of sufficient conditions for the convergence of the DP procedure to an optimal policy.

1. Introduction. The aim of this paper is to extend the classical results of discounted infinite horizon dynamic programming to situations where a sequence of actions could be taken simultaneously. This class of problems stems from the modeling of continuous-time maintenance or replacement systems. For such models, there is no natural way to postulate the usual (strong) contraction hypothesis in the Denardo operator formalism as in [1, 2, 4, 7, 8, 14, 17, 18, 21]. The main point of this paper is to show that this class of models still admits of analysis via the contraction mapping approach, under a much weaker assumption called the *local contraction hypothesis*.

[6, 11, 12] give a motivation for the theory developed in the present paper. They deal with deteriorating systems which can be inspected, repaired, replaced, overhauled, etc. An event is defined as the undertaking of one of these maintenance actions. If the system is modeled in a continuous-time setting, there is no mathematical reason for eliminating the possibility of simultaneous events or for bounding away from zero the (expected) time delay between pairs of successive events. In such circumstance, the one-stage (expected) discount factor is not bounded away from one and the usual contraction assumption is obviously violated. However, for these systems, even if simultaneous events are allowed, it is economically unattractive to use a strategy which would generate too large a number of simultaneous (or almost simultaneous) events. Hence the idea of the local contraction hypothesis: formulate realistic assumptions on the one-stage cost function and on the class of admissible strategies such that, without preventing the occurrence of simultaneous events, it eliminates as candidates for optimality those strategies which would generate too many events in a short period of time.

*Received December 26, 1984, revised October 8, 1986.
AMS 1980 subject classification. Primary: 90C39.
IAOR 1973 subject classification. Main: Programming: Dynamic.
OR/MS Index 1978 subject classification. Primary: 111 Dynamic Programming.
Key words. Dynamic programming, discounting, local contraction.
†This research has been supported by "Action Structurante"-Québec, NSERC-Canada grant #A4952 and FCAR-Québec grant #EQ0428 to the second author, and by NSERC-Canada grant #A5468 and FCAR-Québec grant #EQ2831 to the first author.
‡Université Laval.
§Ecole des Hautes Etudes Commerciales.

0364-765X/88/1301/$01.25

Another difficulty in the mathematical analysis of continuous-time optimal maintenance problems stems from the fact that the state is continuous (e.g. includes the ages of the components), therefore the measurability issue has to be addressed. Bertsekas and Shreve [1, 19] have proposed a nice mathematical framework for the study of Discrete-Event Dynamic Programming (DEDP) models, with Borel state and action spaces, and a constant one-stage discount factor. Here, we extend their formalism to a larger class of models. This setting encompasses the classical models of *semi-Markov* and *Markov Renewal Decision Process* (MRDP) with discounting [4, 9, 10, 14]. Related models have been studied by Schäl [16, 18] and Whittle [22], but in a different mathematical framework and under different sets of assumptions.

The paper is organized as follows. In §2, we focus on the local contraction hypothesis and its relationship with the contraction mapping approach in Dynamic Programming. In order to eliminate unnecessary complexities in this section, we formulate the problem as if the state space were countable, putting aside temporarily the delicate measurability issues. In §3 we give the more general version of the model with Borel state and action spaces. In §4 we prove that an adaptation of the contraction mapping approach can be used to obtain the basic results of Dynamic Programming for our model. §5 deals with sufficient conditions for the existence of an optimal policy.

2. The local contraction hypothesis. Consider a *Discrete Event Dynamic Programming* (DEDP) model with state space X and action space A. Each state x in X has a nonempty set of admissible actions $A(x)$. At each of an infinite sequence of stages (events), the decision maker observes the state x and selects an action a from $A(x)$. A cost $g(x, a)$ is incurred for the current stage and the next state x' is generated randomly according to a probability measure $Q(\cdot|x, a)$ over X. A new action a' is selected from $A(x')$, and so on. All costs incurred in state x are discounted to a given point of reference by a state-dependent discount factor $\beta(x)$, $0 < \beta(x) \leqslant 1$. Each $Q(\cdot|x, a)$ is assumed to be concentrated on the set of states x' for which $\beta(x') \leqslant \beta(x)$. The expected one-stage discount factor associated with state x and action a in $A(x)$ is

$$\alpha(x, a) = \frac{1}{\beta(x)} \int_X \beta(x') Q(dx'|x, a)$$

and satisfies $0 \leqslant \alpha(x, a) \leqslant 1$.

Let us assume for the moment that X is countable. The *policy* space is then the set of all functions μ: $X \to A$ such that $\mu(x) \in A(x)$ for all $x \in X$.

Let B be the set of all extended real-valued functions V: $X \to [-\infty, \infty]$, endowed with the supremum norm $\|V\| = \sup_{x \in X} |V(x)|$, and B_0 be the Banach space of all bounded functions in B. An operator ϕ mapping a closed subset of B_0 into itself is said to be *contracting* with factor α if $\alpha < 1$ and $\|\phi(V_2) - \phi(V_1)\| \leqslant \alpha \|V_2 - V_1\|$ for all V_1 and V_2 in that subset.

Defined below are three standard dynamic programming operators. For $V \in B$, $x \in X$ and $a \in A(x)$, let (when the integral exists):

$$H(V)(x, a) = g(x, a) + \frac{1}{\beta(x)} \int_X V(x') \beta(x') Q(dx'|x, a), \tag{1}$$

$$T(V)(x) = \inf_{a \in A(x)} H(V)(x, a). \tag{2}$$

For every policy μ, let

$$T_\mu(V)(x) = H(V)(x, \mu(x)). \tag{3}$$

It is easily seen that the operators T and T_μ are monotone; that is, if $V_1 \leqslant V_2$, then $T(V_1) \leqslant T(V_2)$ and $T_\mu(V_1) \leqslant T_\mu(V_2)$. Without further conditions, however, we cannot be certain that T or T_μ are contracting operators on B_0 (or on one of its proper subsets).

The usual assumptions on DEDP models that admits of analysis via contraction mappings [4, 9, 10, 13] include the uniform boundedness of the cost function g and the uniform boundedness, away from one, of the one-stage discounting function α. These assumptions imply that T and *each* T_μ are contracting operators.

In this paper, we propose a model for which these strong contraction assumptions are replaced by milder conditions, allowing any number of events during any time period. First, it calls for the existence of *one* policy $\tilde{\mu}$ for which $T_{\tilde{\mu}}$ is contracting. Other conditions make economically unattractive those strategies under which the n-stage discount factor does not converge to zero almost surely as n goes to infinity.

A policy μ is called *distinguished* if there exist three constants $\delta_1 < 1$, g_0 and g_1 such that for all x in X:

$$\alpha(x, \mu(x)) \leqslant \delta_1, \tag{4}$$

$$g_0 \leqslant g(x, \mu(x)) \leqslant g_1. \tag{5}$$

If μ is distinguished, then T_μ is contracting on B_0 with factor δ_1. The DEDP model is called *locally contracting* if there exist a distinguished policy $\tilde{\mu}$ and two constants K_1 and K_2 such that

$$K_1 + K_2 > 0 \tag{6}$$

and for all $x \in X$ and $a \in A(x)$,

$$K_1 + K_2\alpha(x, a) \leqslant g(x, a). \tag{7}$$

If $K_1 \geqslant 0$, then by (6) and (7), the cost is always nonnegative. The cost function can take positive and negative values if $K_1 < 0$ and $K_2 > -K_1 > 0$. Conditions (6) and (7) mean that a positive cost will be incurred if the expected discount factor is close to one. Due to these conditions, a high cost will be associated with any strategy that tends to generate too many simultaneous (or almost simultaneous) events. Condition (4) ensures that it is possible for the decision maker to use a policy under which the expected discount factor between any two successive stages is no larger than $\delta_1 < 1$. It does not imply, however, that an optimal strategy has this property.

In a locally contracting DEDP, it need not be the case that for every policy μ there exists an integer n such that T_μ^n, the n-fold composition of T_μ, is contracting. T is not necessarily contracting either, but the key to our analysis will be to show that there is a closed subset of B_0 and an integer n_0 such that T^n is contracting on that subset for all $n \geqslant n_0$. From this n-stage contraction property, the Dynamic Programming approach winds out. The mathematical proof of this property will be given in the forthcoming sections for a DEDP model defined in a more general setting.

3. The general DEDP model. Let X and A be two Borel spaces called the *state space* and *action space* respectively. The *constraint set* Γ is an analytic subset of $X \times A$ such that for each $x \in X$, the slice $A(x) = \{a \in A | (x, a) \in \Gamma\}$ is nonempty. $A(x)$ is the set of *admissible actions* in state x. The *cost function* g: $\Gamma \to (-\infty, \infty)$ is a lower semianalytic function; the *discount function* β: $X \to (0, 1]$ is Borel-measurable; and the *transition kernel* Q is a Borel measurable stochastic kernel on X given $X \times A$.

For a definition of the measurability concepts used in this and forthcoming sections, we refer to [1].

A *policy* μ is a universally measurable stochastic kernel on A given X such that $\mu(A(x)|x) = 1$ for all x in X. Let U denote the set of policies. When $\mu(\cdot|x)$ is degenerate for each x, the policy is called *nonrandomized* (NR) and can be viewed as a universally measurable function μ: $X \to A$.

A (Markov) *strategy* is a sequence $\pi = (\mu_0, \mu_1, \ldots, \mu_n, \ldots)$ such that each μ_n is a policy. It is NR if each μ_n is NR and it is called *stage-stationary* (SS) if all μ_n are identical ($\equiv \mu$). In the latter case, we also use the symbol μ to represent the SS strategy. In this paper, we consider only Markov strategies. It can be easily shown, by a direct adaptation of the proofs given in [21] and Proposition 8.1 of [1], that for any non-Markov strategy (where each μ_n may be conditioned on all the previous history of the process) and each initial state x, there exists a Markov strategy which is at least as good.

To each initial state $x \in X$ and strategy $\pi = (\mu_0, \mu_1, \ldots, \mu_n, \ldots)$ there corresponds a probability measure $P_{\pi,x}$ on the set of all infinite sequences $h = (x_0, a_0, x_1, a_1, \ldots)$ in $H = \Gamma \times \Gamma \times \ldots$, with corresponding mathematical expectation $E_{\pi,x}$, and such that

$$P_{\pi,x}(x_0 = x) = 1, \tag{8}$$

$$P_{\pi,x}(\cdot|x_0, a_0, \ldots, x_n) = \mu_n(\cdot|x_n), \tag{9}$$

$$P_{\pi,x}(\cdot|x_0, a_0, \ldots, x_n, a_n) = Q(\cdot|x_n, a_n), \tag{10}$$

where the dots in (9) and (10) denote the appartenance of a_n to a Borel subset of A and of x_{n+1} to a Borel subset of X respectively (see [1, 13]). The variables x_n and a_n denote the state and action at stage n respectively. We also define

$$\beta_n = \frac{\beta(x_n)}{\beta(x_0)}, \tag{11}$$

which is the n-stage discount factor between stage 0 and stage n, and

$$c_n = \beta_n g(x_n, a_n), \tag{12}$$

the cost for stage n discounted to the initial stage.

The structure $(X, A, \Gamma, g, \beta, Q)$ is said to define a *regular DEDP model* if the following condition is satisfied:

Condition 1. For each state x and strategy π, $C = \sum_{n=0}^{\infty} c_n$ is well defined $P_{\pi,x}$-almost everywhere and $E_{\pi,x}(C)$ is well defined (i.e. takes a value in the interval $[-\infty, \infty]$).

Given a regular DEDP model, a strategy π and an initial state $x_0 = x$, define

$$V_\pi(x) = E_{\pi,x}(C) \tag{13}$$

and let

$$V_*(x) = \inf_{\pi \in \Pi} V_\pi(x) \tag{14}$$

where Π is the set of all strategies. Functions V_π and V_* represent respectively the total expected discounted cost associated with strategy π and the optimal total expected discounted cost. When π is SS, we write $P_{\mu,x}$, $E_{\mu,x}$ and V_μ instead of $P_{\pi,x}$, $E_{\pi,x}$ and V_π.

We say that strategy π is *optimal at* x if $V_\pi(x) = V_*(x)$, and ϵ-*optimal at* x, for $\epsilon > 0$, if $V_\pi(x) \leqslant V_*(x) + \epsilon$. If π is optimal (resp. ϵ-*optimal*) at x for every x in X, it

is called *optimal* (*ϵ-optimal*). Similar definitions hold for policies (regarded as SS strategies).

The operator T is defined over B_0 as in (1)–(2) and the operator T_μ is defined, for each policy μ viewed as a mapping that assigns to each state x a probability measure $\mu(\cdot|x)$ over $A(x)$, by

$$T_\mu(V)(x) = \int_{A(x)} H(V)(x, a)\mu(da|x).$$

Let B_u be the space of all universally measurable functions in B, and B_1 be the space of all functions in B_u which are lower semianalytic (B is defined in §2). As in [1, p. 26], we adopt the convention that $\infty - \infty = \infty$ and $0 \cdot \infty = 0$, so the sum and product of any two extended real numbers is well defined. In this way for each V in B_u, the expressions (1)–(3) are well defined. T_μ and T are the usual *dynamic programming operators*. From Proposition 7.46 in [1], T_μ maps B_u into B_u and can be composed. T does not map B_u into B_u but from Propositions 7.45 and 7.50 in [1], it maps B_1 into B_1. Let T_μ^n and T^n denote respectively the n-fold compositions of T_μ on B_u and of T on B_1.

The next lemma shows that if the structure $(X, A, \Gamma, g, \alpha, Q)$ corresponds to a *locally contracting* DEDP, then Condition 1 is satisfied and we have a regular DEDP model.

LEMMA 1. *If there exist a distinguished policy $\tilde{\mu} \in U$ and two constants K_1 and K_2 that satisfy* (6)–(7), *then Condition* 1 *is satisfied and for each π in* Π,

$$V_\pi(x) = E_{\pi,x}(C) = \sum_{n=0}^{\infty} E_{\pi,x}(c_n). \tag{15}$$

PROOF. For a given sequence h in H and a fixed positive integer n, we have from (7):

$$\sum_{i=0}^{n} c_i^- = \sum_{i=0}^{n} \beta_i g^-(x_i, a_i) \leqslant \sum_{i=0}^{n} \beta_i \max(0, -K_1 - K_2\alpha(x_i, a_i))$$

where $c_i^- = \max(-c_i, 0)$ and $g^-(\cdot) = \max(-g(\cdot), 0)$. Define the set of integers $\Phi = \{i | 0 \leqslant i \leqslant n \text{ and } -K_1 - K_2\alpha(x_i, a_i) > 0\}$. Let ν be the cardinality of Φ, let $\xi(1), \ldots, \xi(\nu)$ be the elements of Φ ranked by increasing order and define $\xi(\nu + 1) = n + 1$. If $K_1 \geqslant 0$ then, by (6), Φ is empty and $\sum_{i=0}^{\infty} c_i^- = 0$. If $K_1 < 0$ then $K_2 > 0$ and

$$\begin{aligned}
E_{\pi,x}\left[\sum_{i=0}^{n} c_i^-\right] &\leqslant E_{\pi,x}\left[\sum_{j=1}^{\nu} \beta_{\xi(j)}\left(-K_1 - K_2\alpha(x_{\xi(j)}, \alpha_{\xi(j)})\right)\right] \\
&= E_{\pi,x}\left[\sum_{j=1}^{\nu} \left(-K_1\beta_{\xi(j)} - K_2\beta_{\xi(j)+1}\right)\right] \\
&\leqslant E_{\pi,x}\left[\sum_{j=1}^{\nu} \left(-K_1\beta_{\xi(j)} - K_2\beta_{\xi(j+1)}\right)\right] \\
&\leqslant E_{\pi,x}\left[-K_1 - K_2\beta_{n+1} - (K_1 + K_2)\sum_{j=2}^{\nu} \beta_{\xi(j)}\right] \\
&\leqslant -K_1.
\end{aligned}$$

In either case, $\sum_{i=0}^{\infty} c_i^-$ converges almost surely and (15) follows from the monotone convergence theorem. ■

4. Contraction properties. In this section, we consider a locally contracting DEDP model$(X, A, \Gamma, g, \beta, Q)$ where $\tilde{\mu} \in U$, $\delta_1 < 1$, $g_0, g_1 \geqslant 0$, K_1 and K_2 satisfy (4)–(7). Such a model allows for situations where for no n, the operator T_μ^n is contracting for every μ. However, we will find a closed subset B_2 of $B_0 \cap B_1$ (B_0 and B_1 are defined in §§2 and 3 respectively) and a real number η such that for every integer $n \geqslant \eta$, T is an n-stage contraction mapping on B_2 (i.e. B_2 is closed under T and T^n is contracting on B_2).

Define

$$B_2 = \left\{ V \in B_1 | K_1 + \min(0, K_2) \leqslant V \leqslant \frac{g_1}{1 - \delta_1} \right\} \tag{16}$$

which is a closed subset of the Banach space B_0.

LEMMA 2. *B_2 is closed under T.*

PROOF. Let $V \in B_2$. From Lemma 7.30 and Propositions 7.47 and 7.48 in [1], $T(V)$ is in B_1. For every x in X,

$$T(V)(x) \leqslant T_{\tilde{\mu}}(V)(x) \leqslant g_1 + \alpha(x, \tilde{\mu}(x))\left(\frac{g_1}{1 - \delta_1}\right) \leqslant \frac{g_1}{1 - \delta_1}.$$

On the other hand, since $K_1 + K_2 > 0$,

$$\begin{aligned} T(V)(x) &\geqslant \inf_{a \in A(x)} \left[K_1 + K_2\alpha(x, a) + \alpha(x, a)(K_1 + \min(0, K_2))\right] \\ &\geqslant K_1 + \min(0, K_2). \end{aligned}$$

Therefore $T(V)$ is in B_2. ■

The next lemma states that T^n is contracting on B_2 if n is large enough. T_μ^n might also be contracting for some μ, but only under additional assumptions. Part (b) will be used in the proof of proposition 4.

LEMMA 3. *Let $\alpha_1 \in (0, 1)$ and*

$$\eta = \frac{g_1/(1 - \delta_1) - K_1 - \min(0, K_2)}{(K_1 + K_2)\alpha_1}. \tag{17}$$

(a) *For any integer $n \geqslant \eta$, T^n is contracting with factor α_1.*

(b) *Let $\mu \in U$, $\epsilon > 0$ and n_1 an integer such that $n_1 \geqslant \eta + \epsilon/((K_1 + K_2)\alpha_1)$ and $T_\mu^{n_1}(K_1 + \min(0, K_2)) \leqslant g_1/(1 - \delta_1) + \epsilon$. Then, for any $n \geqslant n_1$, T_μ^n is contracting on $B_0 \cap B_1$ with factor α_1.*

PROOF. Let V_1 and V_2 in B_2, $\epsilon_1 > 0$ and $n \geqslant \eta$. By Proposition 7.50 in [1], there is a sequence $\mu_0, \mu_1, \ldots, \mu_{n-1}$ of nonrandomized policies such that

$$T_{\mu_i}\left(T^{n-i-1}(V_1)\right) \leqslant T\left(T^{n-i-1}(V_1)\right) + \frac{\epsilon_1}{n}$$

for $i = 0, \ldots, n - 1$. Using induction on n, one easily sees that

$$T_{\mu_0} \cdots T_{\mu_{n-1}}(V_1) \leqslant T^n(V_1) + \epsilon_1 \leqslant \frac{g_1}{1 - \delta_1} + \epsilon_1. \tag{18}$$

Setting $\mu_i = \mu_{n-1}$ for $i \geqslant n$ and $\pi = (\mu_0, \ldots, \mu_{n-1}, \mu_n, \ldots)$, we have

$$\begin{aligned} & T_{\mu_0} \cdots T_{\mu_{n-1}}(V_1)(x) \\ & \quad = E_{\pi, x}\left[\sum_{i=0}^{n-1} c_i + \beta_n V_1(x_n)\right] \\ & \quad \geqslant E_{\pi, x}\left[K_1\beta_0 + K_2\beta_n + (K_1 + K_2)\sum_{i=1}^{n-1}\beta_i + (K_1 + \min(0, K_2))\beta_n\right] \\ & \quad \geqslant K_1 + \min(0, K_2) + n(K_1 + K_2)E_{\pi, x}(\beta_n). \end{aligned} \tag{19}$$

From (17–19), we obtain

$$E_{\pi, x}(\beta_n) \leqslant \frac{g_1/(1 - \delta_1) + \epsilon_1 - K_1 - \min(0, K_2)}{n(K_1 + K_2)} \leqslant \alpha_1 + \frac{\epsilon_1}{n(K_1 + K_2)}$$

and

$$\begin{aligned} T^n(V_2) - T^n(V_1) &\leqslant T_{\mu_0} \cdots T_{\mu_{n-1}}(V_2)(x) - T_{\mu_0} \cdots T_{\mu_{n-1}}(V_1)(x) + \epsilon_1 \\ &\leqslant \|V_2 - V_1\| E_{\pi, x}(\beta_n) + \epsilon_1 \\ &\leqslant \|V_2 - V_1\|\alpha_1 + \left(1 + \frac{1}{n(K_1 + K_2)}\right)\epsilon_1. \end{aligned}$$

Letting $\epsilon_1 \to 0$ and since V_1 and V_2 can be interchanged, (a) follows.

Under the assumptions of (b), we have

$$T_\mu^{n_1}(K_1 + \min(0, K_2)) \geqslant K_1 + \min(0, K_2) + n(K_1 + K_2)E_{\mu, x}(\beta_{n_1})$$

and then

$$E_{\mu, x}(\beta_{n_1}) \leqslant \frac{\alpha_1\eta}{n_1} + \frac{\epsilon}{n_1(K_1 + K_2)} \leqslant \alpha_1$$

from which we obtain that for any V_1 and V_2 in $B_0 \cap B_1$,

$$T_\mu^{n_1}(V_2) - T_\mu^{n_1}(V_1) \leqslant \|V_2 - V_1\|\alpha_1.$$

Since we can interchange V_1 and V_2 and since

$$T_\mu^n(V_2) - T_\mu^n(V_1) = T_\mu^{n-n_1}\left(T_\mu^{n_1}(V_2) - T_\mu^{n_1}(V_1)\right) \leqslant \|T_\mu^{n_1}(V_2) - T_\mu^{n_1}(V_1)\|$$

for all $n > n_1$, this completes the proof. ■

From the two previous lemmas and the fixed point theorem for contraction mappings [1, 4], there is a unique fixed point $\tilde{V}_*$ in B_2 such that $T(\tilde{V}_*) = \tilde{V}_*$ and $\lim_{n\to\infty}\|T^n(V) - \tilde{V}_*\| = 0$ for all V in B_2. Thus, the DP algorithm converges to this $\tilde{V}_*$ and it remains to prove that $\tilde{V}_* = V_*$. This is done in the following proposition. Related results appear in [1, 2, 4, 8, 16, 18, 19, 22] for other DP models.

PROPOSITION 4. *For any V in B_2, we have*:
(a) $T(V) = V$ *if and only if* $V = V_*$;
(b) $T(V) \leqslant V$ *implies* $V_* \leqslant V$;
(c) $T(V) \geqslant V$ *implies* $V_* \geqslant V$;
(d) $\lim_{n\to\infty} \|T^n(V) - V_*\| = 0$;
(e) V_* *is in* B_2;
(f) *A policy* μ *is optimal if and only if* $T_\mu(V_*) = V_*$;
(g) *A policy* μ *is optimal if and only if* $T(V_\mu) = V_\mu$ *and* $V_\mu \in B_2$.

The proof needs the two following lemmas.

LEMMA 5. *If there exists* $\tilde{\beta} > 0$ *such that*

$$p \overset{\text{def}}{=} P_{\pi,x}\left[\limsup_{n\to\infty} \beta_n \geqslant \tilde{\beta}\right] > 0 \tag{20}$$

then $V_\pi(x) = \infty$.

PROOF. We have in that case

$$E_{\pi,x}\left[\sum_{i=0}^{n} c_i\right] \geqslant E_{\pi,x}\left[K_1\beta_0 + K_2\alpha(x_n, a_n) + (K_1 + K_2)\sum_{i=1}^{n}\beta_i\right]$$

$$\geqslant (K_1 + \min(0, K_2)) + n(K_1 + K_2)\tilde{\beta}p$$

and from (15), $V_\pi(x) = \infty$. ∎

LEMMA 6. *Let* μ, α_1, ϵ *and* n_1 *satisfy the assumptions of Lemma* 3(b). *Then*
(a) $T_\mu(V_\mu) = V_\mu$;
(b) $\lim_{n\to\infty} \|T_\mu^n(V) - V_\mu\| = 0$ *for all* V *in* B_2.

PROOF. The idea of the proof is to find a closed subset of B_1 on which the operators $T_\mu^{n_1}, T_\mu^{n_1+1}, \ldots, T_\mu^{2n_1}$ are closed, and then to apply the fixed point theorem.

Let $K_0 = K_1 + \min(0, K_2)$. From (7), $g \geqslant K_0$. From the assumptions and from (1), (3), we have

$$g(x, \mu(x)) - n_1|K_0| \leqslant T_\mu^{n_1}(K_0)(x) \leqslant \frac{g_1}{1 - \delta_1} + \epsilon$$

and then

$$g(x, \mu(x)) \leqslant \frac{g_1}{1 - \delta_1} + \epsilon + n_1|K_0| \tag{21}$$

for all x in X. Call g_2 the r.h.s. of (21). Let $K_3 = (2n_1g_2 + (1 + \alpha_1)|K_0|)/(1 - \alpha_1)$ and $B_3 = \{V \in B_1 | K_0 \leqslant V \leqslant K_3\}$, which includes B_2. Let V be in B_3 and n an integer such that $n_1 \leqslant n \leqslant 2n_1$. We have $T_\mu^n(V) \geqslant T_\mu^n(K_0) \geqslant T^n(K_0) \geqslant K_0$. From (1), (3), (21) and Lemma 3(b), we also have

$$\begin{aligned} T_\mu^n(V) &= T_\mu^n(K_0) + T_\mu^n(V) - T_\mu^n(K_0) \\ &\leqslant 2n_1g_2 + |K_0| + \alpha_1\|V - K_0\| \\ &\leqslant 2n_1g_2 + (1 + \alpha_1)|K_0| + \alpha_1|K_3| \\ &= K_3. \end{aligned}$$

Thus, B_3 is closed under T_μ^n. From the fixed point theorem, there is a unique $\tilde{V}_\mu^n$ in B_3 such that $T_\mu(\tilde{V}_\mu^n) = \tilde{V}_\mu^n$ and $\lim_{k\to\infty}\|T_\mu^{kn}(V) - \tilde{V}_\mu^n\| = 0$ for all V in B_3. These $\tilde{V}_\mu^n$ are clearly identical for all n between n_1 and $2n_1$, since for $n_1 \leqslant n_i < n_j \leqslant 2n_1$, we have

$$\lim_{k\to\infty}\left\|T_\mu^{kn_in_j}(V) - \tilde{V}_\mu^n\right\| = 0$$

for both $n = n_i$ and $n = n_j$. Let $\tilde{V}_\mu \overset{\text{def}}{=} \tilde{V}_\mu^{n_1}$. We thus have

$$\lim_{n\to\infty}\|T_\mu^n(V) - \tilde{V}_\mu\| = 0$$

for all V in B_3. Furthermore the following holds:

$$\tilde{V}_\mu = T_\mu^{n_1+1}\left(\tilde{V}_\mu\right) = T_\mu\left(T_\mu^{n_1}\left(\tilde{V}_\mu\right)\right) = T_\mu\left(\tilde{V}_\mu\right).$$

Now it only remains to show that $\tilde{V}_\mu = V_\mu$.

For any V in B_3 and x in X, we have

$$\begin{aligned}\tilde{V}_\mu(x) &= \lim_{n\to\infty} T_\mu^n(V)(x) = \lim_{n\to\infty} E_{\mu,x}\left[\sum_{i=0}^{n} c_i + \beta_n V(x_n)\right] \\ &= V_\mu(x) + \lim_{n\to\infty} E_{\mu,x}\left[\beta_n V(x_n)\right].\end{aligned}$$

Since V and $\tilde{V}_\mu$ are bounded, V_μ is bounded and from Lemma 5 the latter limit equals zero. This completes the proof. ■

PROOF OF PROPOSITION 4. To prove (a), (d) and (e), it suffices to show that $\tilde{V}_* = V_*$. For any strategy $\pi = (\mu_0, \mu_1, \ldots) \in \Pi$, $V \in B_2$ and $x \in X$, we have from Lemmas 1 and 5:

$$\begin{aligned}V_\pi(x) &= \lim_{n\to\infty}\left[T_{\mu_0}\cdots T_{\mu_{n-1}}(V)(x) - E_{\pi,x}(\beta_n)V(x_n)\right] \\ &\geqslant \lim_{n\to\infty}\left[T_{\mu_0}\cdots T_{\mu_{n-1}}(V)(x)\right] \\ &\geqslant \lim_{n\to\infty} T^n(V)(x) = \tilde{V}_*(x) \qquad (22)\end{aligned}$$

and therefore

$$V_*(x) = \inf_{\pi\in\Pi} V_\pi(x) \geqslant \tilde{V}_*(x).$$

To verify the reverse inequality, choose $\alpha_1 \in (0,1)$, $\epsilon > 0$ and an integer $n_1 \geqslant \eta + \epsilon/((K_1 + K_2)\alpha_1)$. For any $\epsilon_1 \in (0, \epsilon)$, from Proposition 7.50 in [1], there exists a nonrandomized policy μ such that $T_\mu(\tilde{V}_*) \leqslant T(\tilde{V}_*) + \epsilon_1/n_1$. Since $T(\tilde{V}_*) = \tilde{V}_*$, we obtain

$$\begin{aligned}T_\mu\left(\tilde{V}_*\right) &\leqslant \tilde{V}_* + \frac{\epsilon_1}{n_1}, \\ T_\mu^2\left(\tilde{V}_*\right) &\leqslant T_\mu\left(\tilde{V}_*\right) + \frac{\epsilon_1}{n_1} \leqslant \tilde{V}_* + \frac{2\epsilon_1}{n_1} \\ &\vdots \\ T_\mu^{n_1}\left(\tilde{V}_*\right) &\leqslant T_\mu\left(\tilde{V}_*\right) + \frac{(n_1-1)\epsilon_1}{n_1} \leqslant \tilde{V}_* + \epsilon_1\end{aligned}$$

and by the monotonicity of T_μ,

$$T_\mu^{n_1}(K_1 + \min(0, K_2)) \leqslant T_\mu^{n_1}(\tilde{V}_*) \leqslant \tilde{V}_* + \epsilon_1 \leqslant \frac{g_1}{1 - \delta_1} + \epsilon_1.$$

From Lemma 3(b), we then obtain:

$$T_\mu^{2n_1}(\tilde{V}_*) \leqslant T_\mu^{n_1}(\tilde{V}_* + \epsilon_1) \leqslant T_\mu^{n_1}(\tilde{V}_*) + \alpha_1\epsilon_1 \leqslant \tilde{V}_* + \epsilon_1 + \alpha_1\epsilon_1$$

$$\vdots$$

$$T_\mu^{in_1}(\tilde{V}_*) \leqslant T_\mu^{(i-1)n_1}(\tilde{V}_* + \epsilon_1) \leqslant \cdots \leqslant \tilde{V}_* + \sum_{j=1}^{i} \alpha_1^{j-1}\epsilon_1$$

and from Lemma 6,

$$\tilde{V}_* \leqslant V_\mu = \lim_{i \to \infty} T_\mu^{in_1}(\tilde{V}_*) \leqslant \tilde{V}_* + \frac{\epsilon_1}{1 - \alpha_1}.$$

Letting $\epsilon_1 \to 0$, we obtain the proof of (a), (d) and (e). Properties (b) and (c) are direct consequences of (d) and the monotonicity of T. It remains to verify (f) and (g).

If μ is an optimal policy, then $V_\mu = V_* \in B_2$ and the conditions of Lemma 6 are satisfied for μ with $\epsilon = 0$. Therefore, $T_\mu(V_\mu) = V_\mu$ or, equivalently, $T_\mu(V_*) = V_*$. If $T_\mu(V_*) = V_*$, then using again Lemma 6, we have

$$V_\mu = \lim_{n \to \infty} T_\mu^n(V_*) = V_* = T(V_*) = T(V_\mu) \in B_2.$$

If $T(V_\mu) = V_\mu \in B_2$ then from (a), $V_\mu = V_*$ and μ is optimal. ∎

5. Existence of optimal policies. We now have optimality conditions, but no guarantee that an optimal policy or strategy exists (see [2]). Sufficient conditions for the existence of an optimal nonrandomized policy are given below. Related results appear in [1, 19].

PROPOSITION 7. (a) *For any $\epsilon > 0$, there exists a nonrandomized ϵ-optimal policy.*

(b) *There exists an optimal nonrandomized policy if and only if for any x in X, the infimum $\inf_{a \in A(x)} H(V_*)(x, a)$ is attained.*

(c) *If for any x in X there exists a strategy optimal at x, then there exists an optimal nonrandomized policy.*

(d) *Let $V \in B_2$. If there is a nonnegative integer n_0 such that for every integer $n \geqslant n_0$, real number λ and state $x \in X$, the set $U_n(V)(x, \lambda) = \{a \in A(x) | H(T^n(V))(x, a) \leqslant \lambda\}$ is compact, then there is a sequence $\mu_0, \mu_1, \mu_2, \ldots$ of nonrandomized policies such that $T_{\mu_n}(T^n(V)) = T^{n+1}(V)$ for all $n \geqslant n_0$, and a nonrandomized policy μ such that for all x in X, $\mu(x)$ is an accumulation point of the sequence $\mu_0(x), \mu_1(x), \mu_2(x), \ldots$. This policy is optimal.*

PROOF. (a) Such a policy has been constructed in the proof of Proposition 4(a).

(b) From Proposition 7.50 in [1], the infimum is attained if and only if there is a nonrandom $\mu \in U$ such that $T_\mu(V_*) = T(V_*) = V_*$, and from Proposition 4(f), this is true if and only if μ is optimal.

(c) Let $x \in X$ and strategy $\pi = (\mu_0, \mu_1, \dots)$ optimal at x. From equation (22) and the monotone convergence theorem, we have

$$V_*(x) = V_\pi(x) \geqslant \lim_{n\to\infty} T_{\mu_0} \cdots T_{\mu_n}(V_*)(x) = T_{\mu_0}\Big(\lim_{n\to\infty} T_{\mu_1} \cdots T_{\mu_n}(V_*)\Big)(x)$$

$$\geqslant T_{\mu_0}\Big(\lim_{n\to\infty} T^n(V_*)\Big)(x) = T_{\mu_0}(V_*)(x) \geqslant T(V_*)(x) = V_*(x).$$

Therefore, $V_*(x) = T_{\mu_0}(V_*)(x)$, the infimum $\inf_{a\in A(x)} H(V_*)(x, a)$ is attained and the conclusion follows from (a).

(d) Let $n \geqslant n_0$ and $x \in X$. Let $\lambda_0 \geqslant \lambda_1 \geqslant \dots$ be a real nonincreasing sequence that converges to $T^{n+1}(V)(x)$. For each integer $i \geqslant 0$, $U_n(V)(x, \lambda_i)$ is compact, nonempty and $U_n(V)(x, \lambda_{i+1}) \subseteq U_n(V)(x, \lambda_i)$. Then $\bigcap_{i=0}^{\infty} U_n(V)(x, \lambda_i)$ is a compact nonempty set and $\inf_{a\in A(x)} H(T^n(V))(x, a)$ is attained for every point in that set. By Proposition 7.50 in [1], for every $n \geqslant n_0$, there exists a nonrandomized policy μ_n such that $T_{\mu_n}(T^n(V)) = T^{n+1}(V)$. From Proposition 4(d), for every $\epsilon > 0$, there is an $n_1 \geqslant n_0$ such that $\|T^n(V) - V_*\| \leqslant \epsilon$ for all $n \geqslant n_1$. Let $n_2 \geqslant n_1$. For $n \geqslant n_2$, we have

$$T_{\mu_n}(T^{n_2}(V)) \leqslant T_{\mu_n}(T^n(V) + 2\epsilon) \leqslant T_{\mu_n}(T^n(V)) + 2\epsilon \leqslant T^{n+1}(V) + 2\epsilon \leqslant V_* + 3\epsilon$$

and then

$$\mu_n(x) \in U_{n_2}(V)(x, V_*(x) + 3\epsilon) \quad \text{for all } x \in X. \tag{23}$$

As in the proof of Lemma 4 in [17], one can construct a policy μ such that for any $x \in X$, $\mu(x)$ is an accumulation point of the sequence $\{\mu_1(x), \mu_2(x), \dots\}$. From (23) and since $U_{n_2}(V)(x, V_*(x) + 3\epsilon)$ is compact, $\mu(x)$ is in $U_{n_2}(V)(x, V_*(x) + 3\epsilon)$ which is a subset of $A(x)$, i.e. $H(T^{n_2}(V))(x, \mu(x)) \leqslant V_*(x) + 3\epsilon$.

Since this is true for any $n_2 \geqslant n_1$, we obtain from Lebesgue's dominated convergence theorem:

$$T_\mu(V_*)(x) = H(V_*)(x, \mu(x)) = \lim_{n_2\to\infty} H(T^{n_2}(V))(x, \mu(x)) \leqslant V_*(x) + 3\epsilon.$$

This holds true for all x in X and $\epsilon > 0$. Therefore $T_\mu(V_*) = V_*$ and from Proposition 4(f), μ is optimal. ■

When the condition in Proposition 7(d) is verified, an optimal policy can theoretically be obtained via the DP procedure. It is verified, in particular, if each $A(x)$ is finite, or if each $A(x)$ is compact, g and V are lower semicontinuous, and α and Q are continuous on $X \times A$ (see Propositions 7.31–7.33 in [1] and Theorem 11.11 in [3]).

Acknowledgement. We wish to thank Michèle Breton, who helped us improve the derivation of Lemma 6, and an anonymous referee who helped recast the paper in a more readable way.

References

[1] Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic Optimal Control: The Discrete Time Case*. Academic Press, New York.

[2] Blackwell, D. (1965). Discounted Dynamic Programming. *Ann. Math. Statist.* **36** 226–235.

[3] Choquet, G. (1969). *Cours d'Analyse, tome* II: *Topologie*. Masson et cie, Paris (in French).

[4] Denardo, E. V. (1967). Contraction Mappings in the Theory Underlying Dynamic Programming. *SIAM Rev.* **9** 165–177.

[5] Haurie, A. and L'Ecuyer, P. (1982). A Stochastic Control Approach to Group Preventive Replacement in a Multicomponent System. *IEEE Trans. Automat. Control* **AC-27** 387–393.

[6] _____ and _____. (1986). Approximation and Bounds in Discrete Event Dynamic Programming. *IEEE Trans. Automat. Control* **AC-31** 227–235.
[7] van Hee, D. M. and Wessels, J. (1978). Markov Decision Processes and Strongly Excessive Functions. *Stochastic Process. Appl.* **8** 59–76.
[8] Hinderer, K. (1970). *Foundations of Nonstationary Dynamic Programming with Discrete Time Parameter*. Lectures Notes in Oper. Res. and Math. Systems, **33**, Springer-Verlag, Berlin and New York.
[9] Howard, R. A. (1964). Research in Semi-Markovian Decision Structures. *J. Oper. Res. Soc. Japan* **6**, **4** 163–199.
[10] Jewell, W. S. (1963). Markov Renewal Programming. I and II. *Oper. Res.* **11** 939–971.
[11] L'Ecuyer, P. (1983). Processus de décision markoviens à étapes discrètes: Application à des problèmes de remplacement d'équipement. Ph.D. thesis (in French), published in *Les cahiers du GERAD*, report no. G-83-06, Ecole des H. E. C., Montréal.
[12] _____ and Haurie, A. (1987). The Repair vs Replacement Problem: A Stochastic Control Approach. *Optim. Control Appl. Meth.* **8** 219–230.
[13] _____ and _____. (1983). Discrete Event Dynamic Programming in Borel Spaces with State Dependent Discounting. Report no. DIUL-RR-8309, Département d'informatique, Univ. Laval, Ste-Foy, Canada.
[14] Ross, S. M. (1970). *Applied Probability Models with Optimization Applications*. Holden-Day, San Francisco.
[15] Rudin, W. (1974). *Real and Complex Analysis*, second ed., McGraw-Hill, New York.
[16] Schäl, M. (1972). On Continuous Dynamic Programming with Discrete Time Parameter. *Z. Wahrsch. Verw. Gebiete* **21** 279–288.
[17] _____. (1974). A Selection Theorem for Optimization Problems. *Arch. Math.* **25** 219–224.
[18] _____. (1975). Conditions for Optimality in Dynamic Programming and for the Limit of n-Stage Optimal Policies to be Optimal. *Z. Wahrsch. Verw. Gebiete* **32** 179–196.
[19] Shreve, S. E. and Bertsekas, D. P. (1979). Universally Measurable Policies in Dynamic Programming. *Math. Oper. Res.* **4**, **1** 15–30.
[20] Stone, L. D. (1973). Necessary and Sufficient Conditions for Optimal Control of Semi-Markov Jump Processes. *SIAM J. Control* **11** 187–201.
[21] Strauch, R. E. (1966). Negative Dynamic Programming. *Ann. Math. Statist.* **37** 871–890.
[22] Whittle, P. (1979). A Simple Condition for Regularity in Negative Programming. *J. Appl. Probab.* **16** 305–318.

L'ECUYER: DEPARTEMENT D'INFORMATIQUE, UNIVERSITÉ LAVAL, STE-FOY, QUÉBEC, CANADA G1K 7P4

HAURIE: GERAD, ECOLE DES HAUTES ETUDES COMMERCIALES, MONTRÉAL, QUÉBEC, CANADA H3T 1V6